

Association for Information Systems

**AIS Electronic Library (AISeL)**

---

International Conference on Information  
Systems 2023 Special Interest Group on Big  
Data Proceedings

Special Interest Group on Big Data Proceedings

---

12-12-2023

## **Towards Establishing a Comprehensive Privacy Impact Assessment Methodology for Big Data Analytics in Compliance with the General Data Protection Regulation**

Georgios Georgiadis

Geert Poels

Follow this and additional works at: <https://aisel.aisnet.org/sigbd2023>

---

This material is brought to you by the Special Interest Group on Big Data Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in International Conference on Information Systems 2023 Special Interest Group on Big Data Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# **Towards Establishing a Comprehensive Privacy Impact Assessment Methodology for Big Data Analytics in Compliance with the General Data Protection Regulation**

*Short Paper*

Georgios Georgiadis, Geert Poels

Ghent University, Faculty of Economics and Business Administration,

Tweekerkenstraat 2, B-9000 Gent, Belgium

{firstname, lastname}@UGent.be

## **Purpose**

The emergence of Big Data Analytics (BDA) technologies represents a significant shift towards data-centric decision-making, fundamentally altering approaches in industry and individual prosperity. These innovations have greatly improved the ability to collect, store, process, and analyse large amounts of structured and unstructured data, allowing for the discovery of insights that were previously inaccessible (Hordri et al., 2017). At its core, BDA aims to uncover correlations, influential factors, market patterns, consumer preferences, and other valuable information by utilising advanced algorithms, computational methods, and statistical analysis. This knowledge empowers organisations to improve their operations, develop targeted marketing strategies, enhance product creation, and ultimately enhance customer satisfaction and experience.

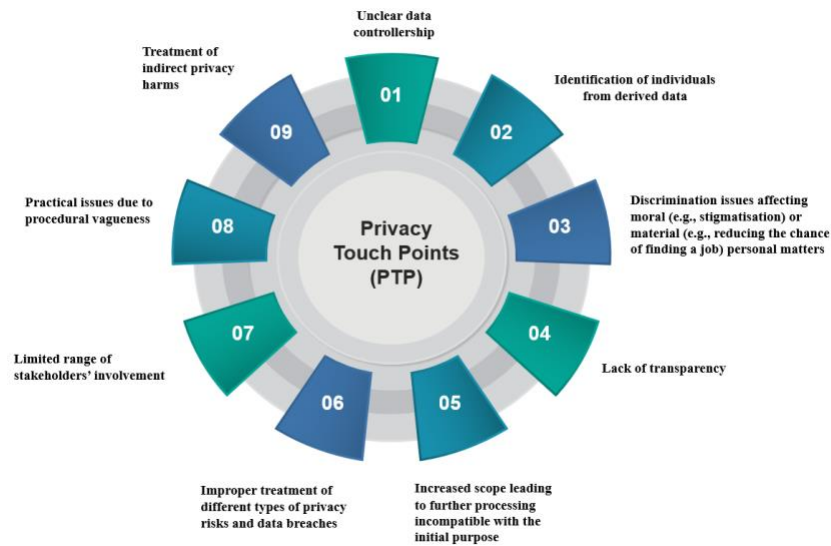
However, the adoption of BDA has sparked debates about privacy and personal data protection, requiring a balanced approach between data utility and individual rights, as outlined in legislations such as the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the US (EU, 2016; Pardau, 2018; Tene & Polonetsky, 2011). As the world becomes increasingly data-driven, it is crucial to address these issues in order to fully leverage the potential of BDA. The GDPR, similar to the CCPA, sets strict guidelines for handling personal data. Concerns such as accidental data exposure and the processing of sensitive information have prompted the implementation of Data Protection Impact Assessments (DPIAs), which are Privacy Impact Assessments (PIAs) designed to proactively evaluate and mitigate risks in data processing (WP29, 2017).

Based on our systematic literature review (Georgiadis & Poels, 2022), we have identified nine Privacy Touch Points (PTPs) which we define as risks to personal data protection and privacy specific to BDA. These PTPs may be overlooked in a DPIA if the specific context of personal data processing in BDA is not taken into account. Our goal is to enhance the DPIA process by examining these PTPs and identifying gaps in current DPIA practices.

Our research aims to explore how a DPIA methodology can be tailored to better suit BDA environments where personal data is processed. The paper presents a comprehensive Delphi study that begins with the nine PTPs to gather expert opinions on their relevance, importance for inclusion in a DPIA, and suggestions on how to address them in a DPIA. The Delphi study is currently ongoing, with individual, semi-structured interviews being conducted to develop practical guidance on conducting a DPIA in a BDA environment.

## Research Design

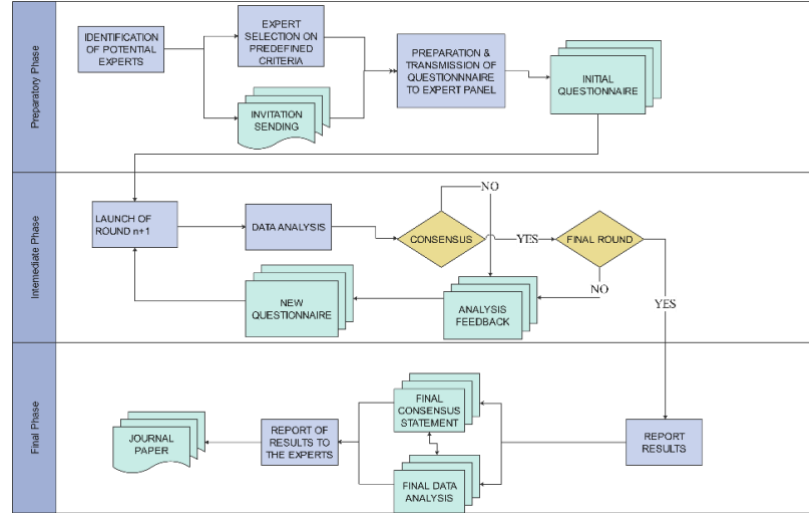
Our research design incorporates a variety of methods. We started with a Systematic Literature Review (SLR) based on the method presented in Okoli and Schabram (2011). This review led us to discover nine PTPs (see Figure 1) and gain insights into the level of coverage of these PTPs by existing PIA methodologies.



**Figure 1: The Nine Privacy Touch Points (PTPs)**

Building on the findings of our SLR (Georgiadis & Poels, 2022), we conducted a Delphi study. The study was divided into three phases: preparatory, intermediate, and final (Figure 2). The Delphi study is a widely used research method that gathers expert opinions through a series of feedback rounds to search for consensus while ensuring controlled and anonymous input (Okoli & Pawlowski, 2004).

In the *preparatory phase*, we established criteria for selecting participants from diverse sectors such as the private sector, academia, and public organisations. We aimed to include individuals with the appropriate knowledge and expertise on the subject matter. We also paid attention to achieving a balanced representation and diversity in terms of size and relevant disciplines, including law, security, and computer science.



**Figure 2: Delphi Study Phases**

In the *intermediate phase*, we administered three separate questionnaires to the experts for each round of the Delphi study. We reused questions until a consensus was reached. The first two rounds focused on affirming and possibly expanding the PTPs. We also asked the experts about their insights on the fundamental elements of a DPIA framework. Specifically, we asked for their views on the building blocks or components of a future DPIA framework that need to be prioritised to adapt the DPIA for use in a BDA context. These views were based on a conceptual model for such a framework that we developed using the research of (Kloza et al., 2021). The conceptual model provides a graphical view of the various components of a DPIA framework and their interconnections.

In the third round, the main focus was on developing a DPIA method suitable for the BDA environment and capable of assessing the PTPs identified as BDA-specific risks to personal data protection and privacy. After each round, we provided experts with a comprehensive report summarising the anonymised survey responses, consensus scores, and our analytical insights. Experts also had the opportunity to offer their reflections beyond the predefined answer choices for each question through the open-ended sections in our surveys.

In the *final phase*, we formulated the final consensus statements and thoroughly examined the survey results from both a statistical and qualitative perspective. As a spontaneous research design decision, we decided to include individual, semi-structured follow-up interviews with experts, some of whom were not part of the Delphi study panel. These interviews were conducted to gain a deeper understanding of our analysis results and to define a list of practical ideas for each PTP to consider when designing a revised DPIA methodology for BDA environments, online discussions were organised to discuss specific findings of the Delphi study. These interviews are currently in progress and will be completed in early 2024.

## Findings

### Consensus Criteria

Measuring consensus among members of the expert panel involves using specific criteria. These criteria serve as a benchmark to determine when the study can be concluded. Evaluating the results of consecutive rounds of the Delphi study is crucial as it allows for an unbiased evaluation of the experts' consensus. This evaluation helps researchers ensure the accuracy and reliability of their results. For our research, we used four different consensus criteria (see Table 1) based on and modified from the study by (Van Looy et al., 2017). In our framework, consensus can be categorised as either positive or negative. A positive consensus indicates agreement on the relevance or importance of a PTP, while a negative consensus signifies

disagreement. We consider the consensus to be ‘strong’ when all criteria are met, and ‘almost strong’ when three out of four criteria are in agreement (see Table 2). Additionally, we examined the experts’ reasoning from open-ended questions to identify any potential discrepancies in the scores.

Condition	Definition (all scores are indicated on a 5-point Likert scale)
# 1	35% of the experts strongly agree (i.e., score 5) or strongly disagree (i.e., score 1)
# 2	70% of the experts agree (i.e., score 4 or 5) or disagree (i.e., score 1 or 2)
# 3	The interquartile range (i.e., the difference between the highest and lowest of the scores of the 50% of the experts that are in the middle when scores are ranked) is less or equal to 1.25
# 4	No expert strongly disagrees/agrees if conditions (1) and (2) based on the frequencies indicate a tendency towards either positive or negative consensus

**Table 1: Consensus Criteria**

Consensus Type	Condition
Strong consensus	All four conditions are met: #1–#4
Almost strong consensus	At least three criteria are met
No consensus	Less than three criteria are met

**Table 2: Consensus Types and Conditions**

### **Round 1: Establishing Consensus on PTPs**

Consensus was reached regarding the relevance and importance of several PTPs, as shown in Table 3. Specifically, PTP (2) ‘identification of individuals from derived data’ and PTP (3) ‘discrimination issues affecting moral or material personal matters’ were recognised as being relevant. It was also emphasised that PTPs such as PTP (1) ‘unclear data controllership’ are crucial for a DPIA. These findings highlight the critical need to address risks related to identity and discriminatory practices within BDA. Open-ended responses provided further insight into these findings, with experts highlighting challenges such as anonymisation and the complexities of data controllership. These insights indicate the intricate nature of managing privacy and protecting personal data in data processing contexts involving BDA.

Measurements	Level of Consensus		
	Strong Positive	Almost Strong Positive	No Consensus
<b>Relevance</b>	PTP (2), PTP (3)	PTP (1), PTP (4), PTP (7)	PTP (5), PTP (6), PTP (8), PTP (9),
<b>Importance</b>	NA	PTP (1), PTP (3)	PTP (2), PTP (4), PTP (5), PTP (6), PTP (7), PTP (8), PTP (9)

**Table 3: Delphi Study Results on Relevance and Importance of PTPs after Round 1**

## Round 2: Refining Consensus and Exploring DPIA Improvements

During the second round, the goal was to reach a consensus on the PTPs and initiate discussions about enhancing the DPIA process. This round resulted in a higher consensus rate, with experts in agreement about the relevance and importance of the majority of the PTPs (see Table 4). However, PTP (8) ‘Practical issues due to procedural vagueness’ remained a point of contention. Additionally, experts provided valuable insights on how to improve the DPIA process by assessing the PTPs. The responses highlighted the need to consider a wide range of stakeholder perspectives and the challenges involved in achieving transparency in BDA processes. These findings indicated the necessity for a more inclusive and transparent DPIA framework that can effectively address the complex dynamics of BDA.

Measurements	Level of Consensus		
	Strong Positive	Almost Strong Positive	No Consensus
Relevance	PTP (2), PTP (3), PTP (4), PTP (5)	PTP (1), PTP (7), PTP (9)	PTP (6), PTP (8)
Importance	PTP (1), PTP (2), PTP (3), PTP (5)	PTP (4), PTP (6), PTP (9)	PTP (7), PTP (8)

**Table 4: Delphi Study Results on Relevance and Importance of PTPs after Round 2**

## Round 3: Final Consensus and Suggesting DPIA Enhancements

The third round revisited PTPs that had not reached a consensus in the previous rounds and sought expert opinions on specific DPIA improvements (see Table 5). PTP (6), ‘Improper treatment of different types of privacy risks and data breaches’, finally achieved a strong positive consensus on relevance, while PTP (7), ‘limited stakeholder involvement’, reached an almost strong positive consensus on importance. This outcome highlighted the growing recognition of the need to effectively manage privacy risks and involve a diverse range of stakeholders in DPIA processes. The suggestions for DPIA enhancements were diverse and reflective of the complex nature of BDA. For example, for PTP (1), ‘unclear data controllership’, a more sophisticated consent model was proposed, underscoring the need for clarity in data controllership roles. For PTP (2), ‘identification of individuals from derived data’, experts called for more robust guidelines on anonymisation and de-identification, acknowledging the challenges in ensuring complete anonymity in BDA contexts. Experts also emphasised the importance of addressing ethical and social impacts, particularly for PTP (3), ‘discrimination issues affecting moral or material personal matters’. This perspective resonated with the growing awareness of the societal implications of BDA, highlighting the need for DPIAs to encompass broader ethical considerations. It is important to note that of all the PTPs, only PTP (8) was excluded from subsequent evaluation as the panel of experts failed to reach a consensus on the existence of operational challenges arising from procedural ambiguities within the DPIA methodology. However, the other eight PTPs are considered validated by the experts.

Measurements	Level of Consensus		
	Strong Positive	Almost Strong Positive	No Consensus
Relevance	PTP (2), PTP (3), PTP (4), PTP (5), PTP (6)	PTP (1), PTP (7), PTP (9)	PTP (8)
Importance	PTP (1), PTP (2), PTP (3), PTP (5)	PTP (4), PTP (6), PTP (7), PTP (9)	PTP (8)

**Table 5: Delphi Study Results on Relevance and Importance of PTPs after Round 3**

## Originality/Value

The value of our research is multifaceted. Firstly, it addresses a crucial gap in the current understanding of personal data protection and privacy in the BDA context. The identification of nine specific PTPs based on a comprehensive literature review and the validation of eight of them by experts provides a focused framework for assessing personal data protection and privacy risks when data processing involves BDA. This finding is particularly relevant given the increasing reliance on BDA across various sectors and the escalating concerns about privacy and data protection in this context. Secondly, the paper's methodological rigour, exemplified using the Delphi study method, adds to its credibility. Engaging with a diverse panel of experts from various fields, the study validates and refines the identified PTPs. This collaborative approach not only enriches the research findings but also ensures that they are grounded in a broad spectrum of professional insights. Another significant aspect of our research is its practical applicability. The findings from the Delphi study and the subsequent interviews will provide concrete suggestions for improving the DPIA in the BDA context. These recommendations will not only be theoretically sound but also actionable, offering clear guidance for practitioners and policymakers in enhancing data protection strategies.

In conclusion, we firmly believe that our research will make a substantial contribution to both academic and practical discourses in personal data protection and privacy. It bridges a critical knowledge gap in the DPIA process for data processing involving BDA and proposes actionable solutions, thus paving the way for more robust privacy protections in an increasingly data-driven world.

## References

- EU. (2016). Regulation (EU) 2016/679 of the European parliament and of the council. *Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Georgiadis, G., & Poels, G. (2022). 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. *Computer Law & Security Review*, 44. <https://doi.org/10.1016/j.clsr.2021.105640>
- Hordri, N., Samar, A., Yuhani, S., & Shamsuddin, S. (2017). A systematic literature review on features of deep learning in big data analytics. *International Journal of Advances in Soft Computing & Its Applications*, 9(1).

- Kloza, D., Van Dijk, Niels, Casiraghi, Simone, Vazquez Maymir, Sergi, & Tana, Alessia. (2021). The concept of impact assessment. In *Border Control and New Technologies*. Academic & Scientific Publishers.  
<https://doi.org/10.46944/9789461171375>
- Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information and Management*, 42(1), 15–29.  
<https://doi.org/10.1016/j.im.2003.11.002>
- Okoli, C., & Schabram, K. (2011). A guide to conducting a systematic literature review of information systems research. <http://dx.doi.org/10.2139/ssrn.1954824>
- Pardau, S. (2018). The California Consumer Privacy Act: Towards a European-style privacy regime in the United States? *Journal of Technology Law & Policy*, 23(1), 68–114.
- Tene, O., & Polonetsky, J. (2011). Privacy in the age of big data: A time for big decisions. *Stanford Law Review Online*, 64, 63–69.
- Van Looy, A., Poels, G., & Snoeck, M. (2017). Evaluating business process maturity models. *Journal of the Association for Information Systems*, 18(6), 461–486.
- WP29. (2017). *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679*.  
[https://ec.europa.eu/newsroom/article29/item-detail.cfm?item\\_id=611236](https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236)