Research Context for Delphi Interviewees

Purpose of the research

The purpose of our research is to investigate and improve the Data Protection Impact Assessment (DPIA) methodology in the context of Big Data Analytics (BDA). As BDA technologies revolutionise the handling of large volumes of both structured and unstructured data, they bring forward both opportunities and challenges. Our research focuses on the intersection BDA and privacy and personal data protection through a systematic literature review (SLR) and a Delphi study. We specifically examine nine identified Privacy Touch Points (PTPs) that pose risks to personal data and privacy in BDA applications. These were discovered during the SLR. The Delphi study and the subsequent individual interviews aim to determine how the DPIA can be adapted and improved to effectively address the PTPs identified as important and relevant by the Delphi experts.

The nine PTPs

The nine PTPs were originally identified through the SLR. Below is an extract without the supporting references of the PTPs from our journal article published in the CLSR (Georgiadis & Poels, 2022).

PTP (1): **Unclear data controllership.** The transfer of big data from one place or recipient to another result in a diverse controllership. Consequently, it is difficult to understand and ensure compliance with privacy and data protection requirements among data controllers. Existing mechanisms that rely on privacy and data protection by policy – such as (informed and explicit) consent and privacy notices or 'notice-and-consent' models fail to provide transparency and control sufficient for individuals to understand how their data have been collected and used.

PTP (2): **Identification of individuals from derived data.** By exploiting computer power and vast quantities of data, Big Data Analytics techniques enable inferences from undisclosed information that can be used to identify an individual, whether alone or combined with other information. Such features have made re-identification easier, arguably to the point of undermining the value of anonymisation.

PTP (3): **Discrimination issues affecting moral (e.g. stigmatisation) or material (e.g. reducing the chance of finding a job) personal matters.** The increasing risk of statistical discrimination or 'data determinism' a by-product of Big Data Analytics, may cause both material and non-material damage to data subjects in accordance with Recital 75 GDPR..

PTP (4): **Lack of transparency.** The opacity of contemporary data processing activities, along with how big data are eventually used, may have nothing to do with the purpose set or anticipated at the time of initial data collection. Apart from transparency infringement, this could lead to intervenability issues, as it increases the risk that data subjects will be unable to assess and exercise their rights. The loss of data subjects' trust in the processing of their data could result in diminished data quality, a critical factor in producing reliable analytics.

PTP (5): **Increased scope leading to further processing incompatible with the initial purpose (function creep).** New types of privacy threats could result from the ability of big data to re-contextualise data. Perhaps the most compelling is the forecast for the global genetic testing market with its massive consumer genomics databases, which is expected to achieve a substantial growth rate of 11.3% between 2018 and 2025.[1] Unlike other types of personal data, an individual's genetic or human genomic sequence data are immutable. Even with today's computing power, the processing of these data may reveal abundant sensitive[2] information about an individual that bears no relation to what was originally intended. Worse still, attempts to

[1] https://www.alliedmarketresearch.com/genetic-testing-market

[2] The GDPR refers to this as a 'special category' of data.

protect these data using techniques like anonymisation are not invincible to re-identification attacks.

PTP (6): Improper treatment of different types of privacy and data protection risks and data breaches. With the growing integration of government programmes resulting from transformations of public administration around ICTs such as e-government and open data, big data–sharing activities are becoming commonplace. The same applies for the new BDA approaches that have gradually appeared in the market following the as-a-Service model. These could lead to new types of data breaches and, in turn, inadequate handing of personal information. In addition, as field researchers and practitioners have remarked, existing security schemes and technologies strive to satisfy pertinent requirements and expectations. However, current research confirms that the security and data protection of big data have received relatively less focus despite being strategically vital for organisations. As a rich organisational asset with distinct features, BDA are also subject to another type of breach connected to the unauthorised disclosure or loss of knowledge when information is extracted from raw data. Big data may not be personal per se, but when combined and analysed using sophisticated algorithms, they can furnish additional facts about an individual. This poses a new type of privacy and security threat in the context of the GDPR. Consequently, we expect that a wide array of complex and costly data breach incidents will occur in the years to come.

PTP (7): Limited range of stakeholders' involvement. For practical reasons – namely, to avoid increasing the workload and time required to conduct a DPIA – Article 35(9) GDPR considers optional the participation of external (to the organisation) stakeholders whose data are predominantly processed in a BDA context. However, to ensure fairness and transparency, the perspectives and concerns of a broad range of stakeholders are vital in risk assessment , as they allow for the building of less privacy-invasive systems and processing operations.

PTP (8): Practical issues due to procedural vagueness. Existing DPIA and PIA methodologies have been criticised for being rather theoretical. As they do not provide a practical guide for assessing (for instance) the design of complex systems or technologies involving different data privacy concerns, including an objective way to assess and present privacy impacts and monitor the efficiency of countermeasures. To properly assess privacy and data protection risks that are often poorly understood and resource intensive to explore, it is necessary to have a common language and a clear process along with an understanding of the different roles involved and patterns or templates to use. For example, big data operations that are not similar in terms of the risks presented and that involve multiple or joint data controllers are inherently complex and require a more accurate and detailed guide than three clauses in Article 26 GDPR.

PTP (9): Treatment of indirect privacy harms.[3] Many domain researchers argued that societal and ethical concerns should also be addressed when dealing with big data. As an example, they mentioned abuse practices in the solidarity principle that results from the accumulation of information power whenever health insurance organisations require unrestricted access to knowledge. The exploitation of big data in healthcare – especially during the COVID-19 outbreak, which caused an unprecedented pressure on services to prevent needless deaths – has generated significant ethical challenges that, apart from unlawful practices, can affect societal norms and values. Similar to ordinary surveillance, big data introduce a type of pervasive social surveillance that can be either voluntary or mandatory, as when issued by courts. The difficulty of assessing algorithmic processing via artificial intelligence techniques poses manifold new threats due to the wide-ranging ethical dimensions of big data with regard to their interactions with less entrenched information flow norms that may differ across socio-cultural contexts. In this context, the European Data Protection Supervisor and the EDPB have strongly endorsed ethics in the processing of big data to emphasise (as have many scholars) that the real problem arises in use. The rapid emergence of BDA has made it practically impossible to explain all the possible uses of data at the time they are initially gathered.

[3] Privacy harm differs from privacy risk in that it refers to the negative impact on a data subject, a group of data subjects or the entire society caused by loss of control or privacy violations against external factors such as societal. Privacy harms are often used to assess privacy risks based on risk sources and weaknesses.

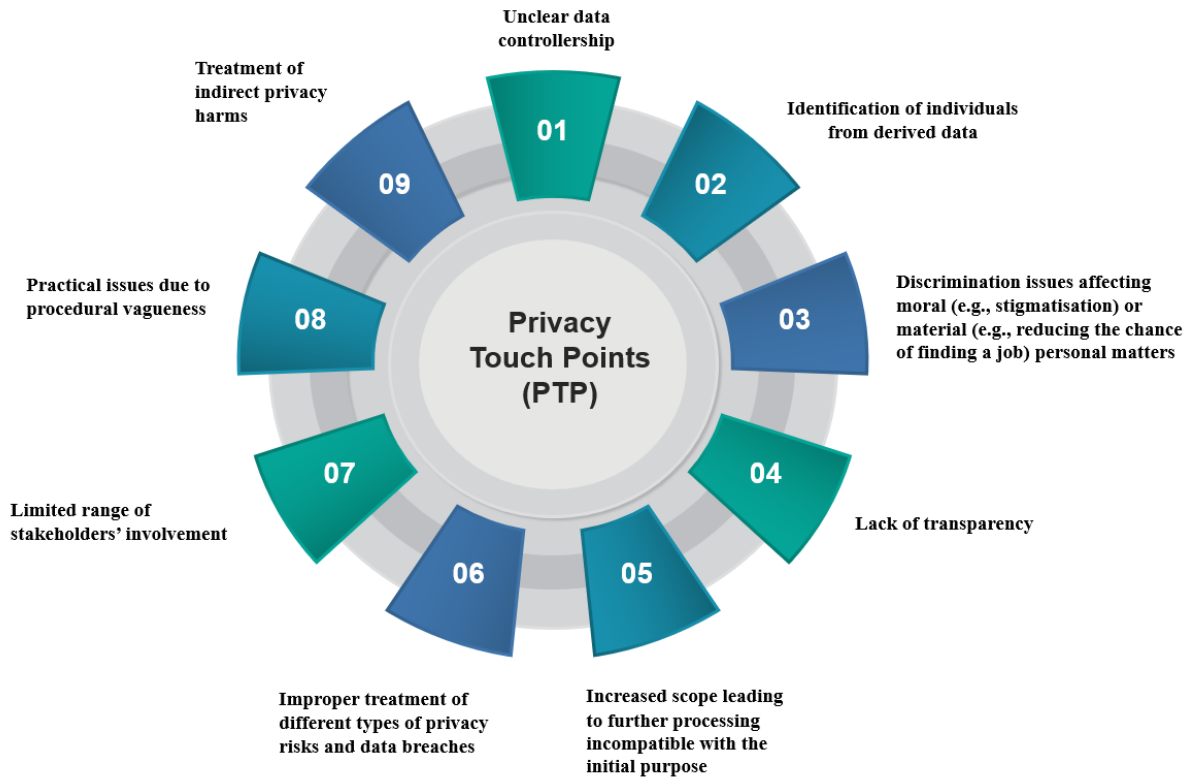

Figure 1: The nine PTPs

The outcome of the Delphi Study

We conducted a Delphi study consisting of three rounds. In each round, experts were presented with a questionnaire containing Likert scale and open-ended questions on PTPs, current impact assessment methodologies and the conceptual framework of the DPIA that we had developed based on our literature findings (see Figure 2). The input from each round was assessed against four clearly defined consensus criteria, which we used to measure the degree of relevance and importance of each PTP.

The conclusion from the third round, presented in Table 1, shows that the experts could not reach an agreement on the procedural indeterminacy of the current DPIA methodology[4]. Some of them considered that it is sufficient in its current form, but many of them agreed that the DPIA needs to be improved or extended for use in BDA, prioritising the 'guidelines', the 'knowledge base' and the 'method' based on the conceptual framework.

.

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

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)

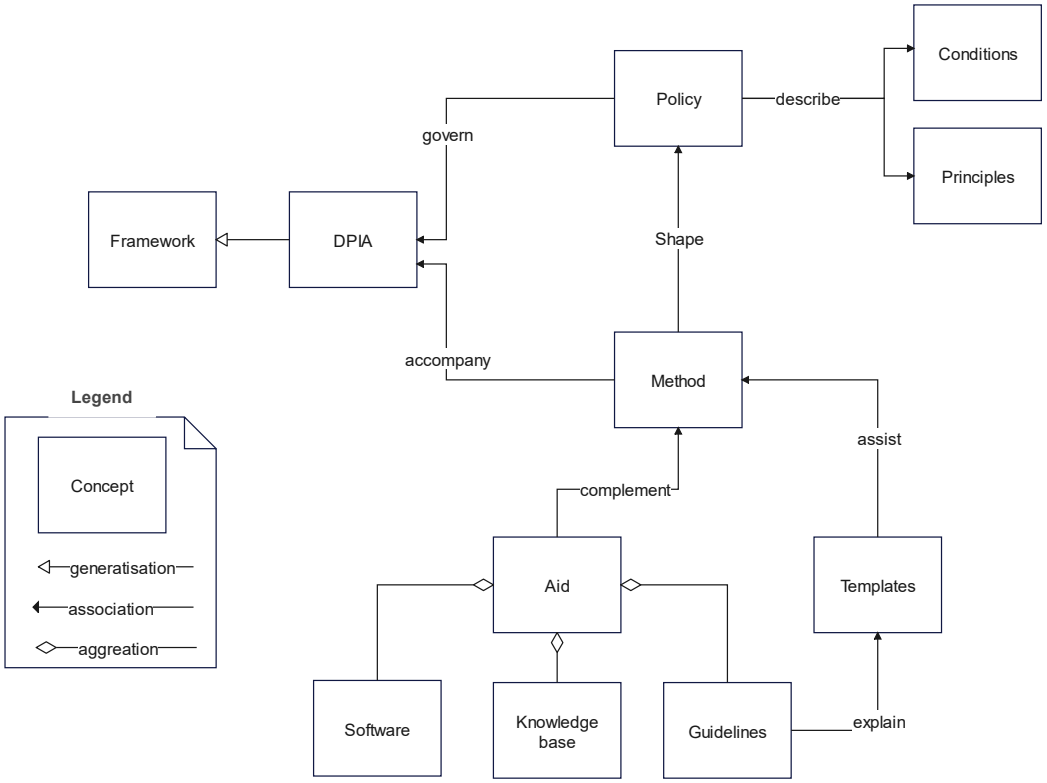

Figure 2: DPIA Conceptual Model

Brief Explanation of the Conceptual Model Terms

'Policy' refers to the set of rules that govern the conduct of a DPIA. 'Method' provides the structure approach or set of procedures (methods) for conducting a DPIA. 'Aid' and 'Templates' complement and assist the method by providing the necessary tools or resources that support the method. 'Knowledge base' refer to the collection of best practices, historical assessments, case studies and expert insights that support the DPIA whereas 'Guidelines' represent a set of detailed instructions or directions on how to utilise the tools and document templates within the DPIA. Finally, 'Principles' refer to the fundamental concepts or beliefs that guide the DPIA process. These are the bedrock on which the DPIA's policy, framework, and methods are built, ensuring that the process adheres to core values and objectives such as fairness, transparency, accountability, and risk minimisation whereas 'Conditions' refer to the specific scenarios, requirements, or triggers that necessitate the initiation of a DPIA.

Aim of the individual interviews

The aim of the individual, semi-structured interviews is to obtain additional views on the PTPs, focussing on their practical perspective, which we firmly believe will be very beneficial for a future "version" of the DPIA or any other type of impact assessment with a particular focus on BDA processes.

Questions for the interviewees

Considering the outcome of the Delphi study, where experts could not reach a consensus on the procedural indeterminacy of the current DPIA methodology, and the suggestions to prioritise guidelines, knowledge base, and methods for BDA, here are the list of questions for the expert interviews:

1. **Procedural Indeterminacy**[5] **and BDA Complexity**: In view of the Delphi study's findings on procedural indeterminacy, what specific improvements do you suggest for the DPIA to address the complexities inherent in BDA?

2. **Enhancing Guidelines for BDA**: The Delphi study indicated a need for better guidelines specific to BDA. What guidelines would you propose to ensure the DPIA is more effectively tailored to BDA applications?

3. **Expanding the Knowledge Base for DPIA in BDA**: How can we expand or adapt the DPIA knowledge base to better encompass the unique challenges and evolving aspects of Big Data Analytics?

4. **Methodological Adaptations for BDA**: Based on your experience and the Delphi study insights, what methodological changes or additions (Art 35 GDPR) would you recommend to make the DPIA more suitable for assessing BDA processes?

5. **Balancing Rigor and Flexibility in DPIA for BDA**: Considering the varied opinions on the current DPIA's sufficiency, how can we balance the need for rigorous assessment with the required flexibility for diverse BDA contexts?

6. **Incorporating Diverse Stakeholder Perspectives in BDA DPIA**: How should the DPIA process be adapted to effectively incorporate diverse stakeholder perspectives, especially in the intricate nature of Big Data environments?

[5] 'Procedural Indeterminacy' refers to a state or condition where the processes or procedures to be followed are not clearly defined, lack specificity, or are open to various interpretations. In legal, organisational, or systematic contexts, this concept implies that there are no strict guidelines or established sequences of actions to dictate how a particular operation or process should be carried out. This could sometimes lead to inconsistency in implementation, as different individuals or groups might choose different ways to approach the same process.

7. **Addressing Ethical and Societal Implications in BDA DPIA**: What ethical and societal considerations should be integrated into the DPIA process, particularly for BDA?

8. **Enhancing Transparency and Accountability in BDA DPIA**: In your opinion, what steps should be taken to enhance transparency and accountability in BDA as part of the DPIA?

9. **Future-Proofing DPIA for BDA**: Considering the rapid technological evolution in the field of Big Data, how can the DPIA be designed to remain relevant and effective for future BDA advancements?

If time permits, I would like to hear your opinion on how the remaining PTPs should be treated in the context of a DPIA.