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

Authored By:

Georgios Georgiadis∗, Geert Poels
Ghent University,
Faculty of Economics and Business Administration,
Tweekerkenstraat 2, B-9000 Gent, Belgium

Presented By:

Sonu Adhikari,st124409,
Paper Reading as part of course requirement for
Business Analytics and Intelligence.

Presentation Outline

01.

INTRODUCTION

02.

BACKGROUND

03.

RESEARCH METHODOLOGY

04.

DISCUSSION AND FUTURE WORK

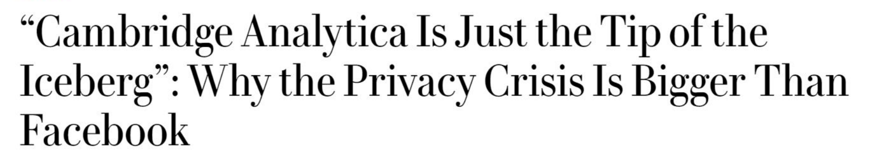
05.

CONCLUSION

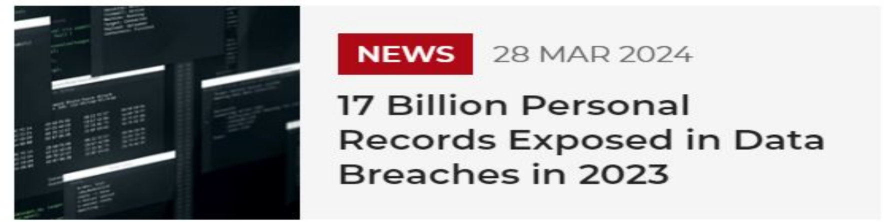
06.

Q & A

Let's set the context first...

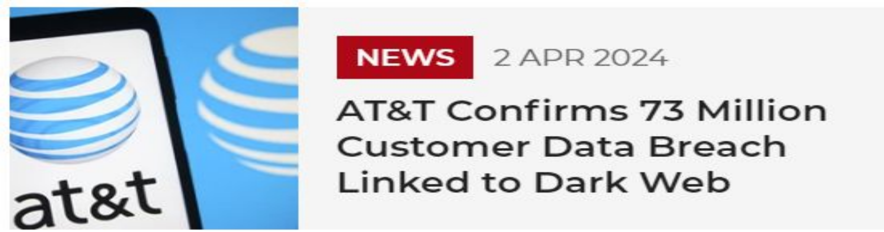


"Cambridge Analytica Is Just the Tip of the Iceberg": Why the Privacy Crisis Is Bigger Than Facebook



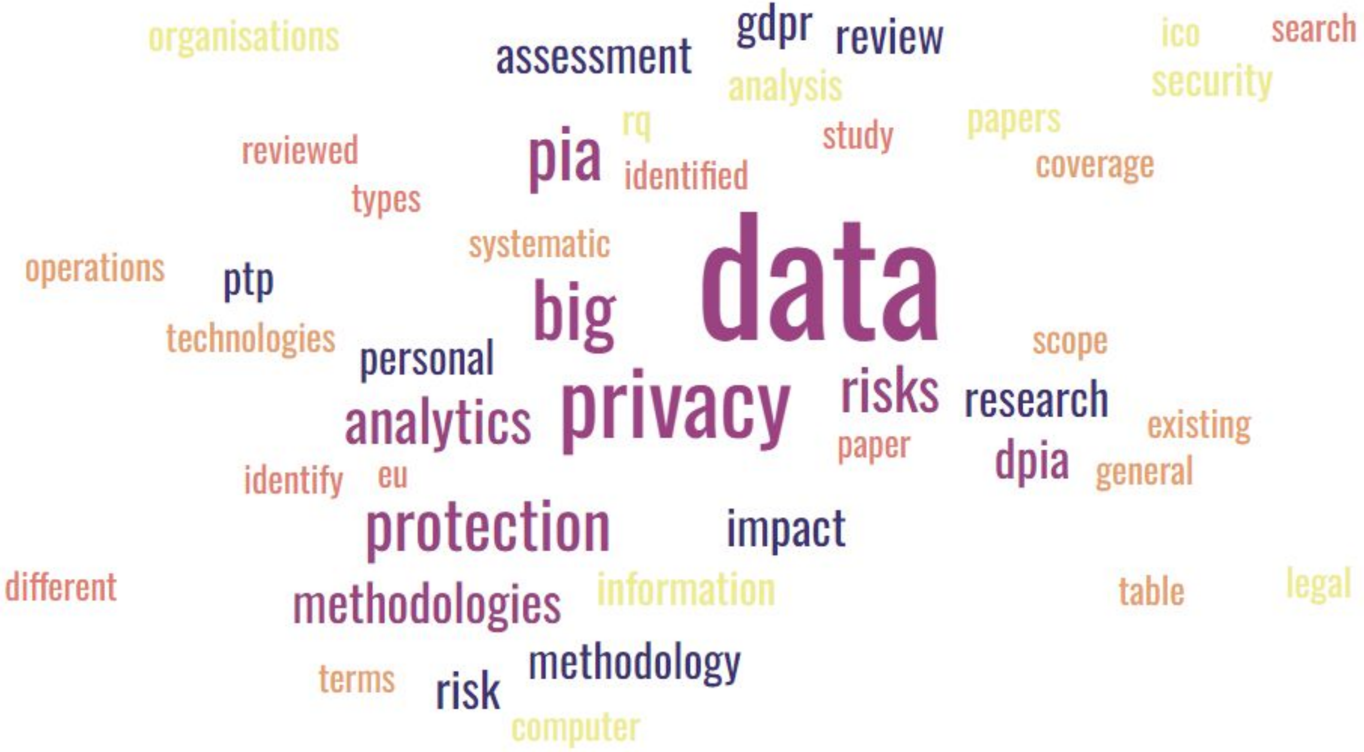
NEWS 28 MAR 2024

17 Billion Personal Records Exposed in Data Breaches in 2023



NEWS 2 APR 2024

AT&T Confirms 73 Million Customer Data Breach Linked to Dark Web



gdpr
search
review
ico
organisations
assessment
analysis
security
rq
study
papers
reviewed
pia
identified
coverage
types
systematic
data
operations
ptp
big
technologies
scope
personal
risks
research
privacy
analytics
existing
paper
dpia
identify
eu
general
protection
impact
different
information
table
legal
methodologies
methodology
terms
risk
computer

Keywords:

- Big data analytics
- Data protection
- Data protection directive
- General data protection regulation
- Governance
- Information security
- Privacy
- Privacy impact assessment
- Systematic literature review

ABSTRACT

Big Data Analytics has enabled businesses and organizations to process and utilize vast amounts of data, leading to improved efficiency and new opportunities.

However, the benefits of Big Data Analytics are restricted by the unintentionally heightened risk of personal data being compromised or breached.

The GDPR mandates Data Protection Impact Assessments to identify appropriate controls and mitigate risks associated with the protection of personal data.

The research aims to develop a DPIA methodology that addresses the unique privacy and data protection challenges in Big Data Analytics environments.

01.

Introduction

INTRODUCTION

- Rapid advancements in the technological areas and social media have facilitated the collection of vast amounts of data, both structured and unstructured.

- This ‘big data’ is often too complex to handle with existing data processing frameworks, which has led to the emergence of Big Data Analytics.

- While big data analytics offers efficiency and opportunities, it also increases the risk of data breaches, potentially exposing sensitive information.

- Therefore, GDPR had been introduced to impose Data Protection Impact Assessment (DPIA) in cases where processing of personal data could impose high risks related to data protection and privacy, that could negatively impact the rights and freedoms of individuals.

- Although the concept of the DPIA became known through the GDPR and is regarded as a relatively new instrument in most European member states as well as private and public organisations, it has existed in the form of the Privacy Impact Assessment (PIA)2 since the mid-1990s

02.

BACKGROUND

Privacy in the era of big data

Current Scenario in the world of Big Data

- Big Data Analytics create value for a wide range of actors in the global economy
- there are some growing concerns, the most pressing of which are related to the secure and lawful processing of personal data
- State-of-the-art tools and measures are required to properly manage big data's so-called '5Vs' (volume, variety, velocity, value and veracity)

GDPR(General Data Protection Regulation)

- Data protection regulatory framework across all European member states since May 2018
- Fundamental role in shaping digital markets in the European Union (EU).
- ensure the fair processing of personal data by setting an array of binding data protection principles
- some authors have assumed that Big Data Analytics cannot adhere to all data protection principles under the GDPR

Privacy Impact Assessment (PIA)

- PIA is a management tool that helps organisations identify, analyse and minimise privacy risks, whether organisational or technical
- It emerged and fully developed between 1995 and 2005
- Helps in informed decision-making by revealing internal communication gaps and hidden assumptions.
- Enable organisations to reduce costs in terms of management time, legal expenses, and potential negative media coverage
- Opponents of PIAs consider them a costly bureaucratic hassle or forced exercise to appease the legal team
- In reality, a properly conducted PIA starts at the beginning of a project, where the cost of change is relatively minimal and influence on project direction is much higher.

Data Privacy Impact Assessment (DPIA) under GDPR

- DPIA mandated by the GDPR as a form of PIA whose primary focus is on compliance with the seven fundamental principles of personal data protection at the core of the GDPR

- DPIA aims to identify and address issues related to the handling of personal data early in a project's life cycle.

- DPIA and PIA are different concepts though they seem similar. The difference between the two lies in scope and basic value drivers.

- PIAs goes beyond the scope of legal frameworks by addressing, for example, general ethical considerations about privacy and personal data handling

- DPIA becomes compulsory when the likelihood of data processing activities results in high risk that could negatively impact the rights and freedoms of the individuals whose data are concerned

03.

RESEARCH METHODOLOGY

STEPS OF THE RESEARCH

1. Systematically review all privacy and data protection risks identified in the literature that are relevant to Big Data Analytics and summarise them into 9 Privacy Touch Points(risks).

2. Analyse the coverage of these Privacy Touch Points in 10 PIA methodologies that received attention in the publications sample.

3. Reveal strengths and weaknesses of existing PIA methodologies when used in Big Data Analytics environments and help in the development of an enhanced methodology for conducting a DPIA

Research Questions

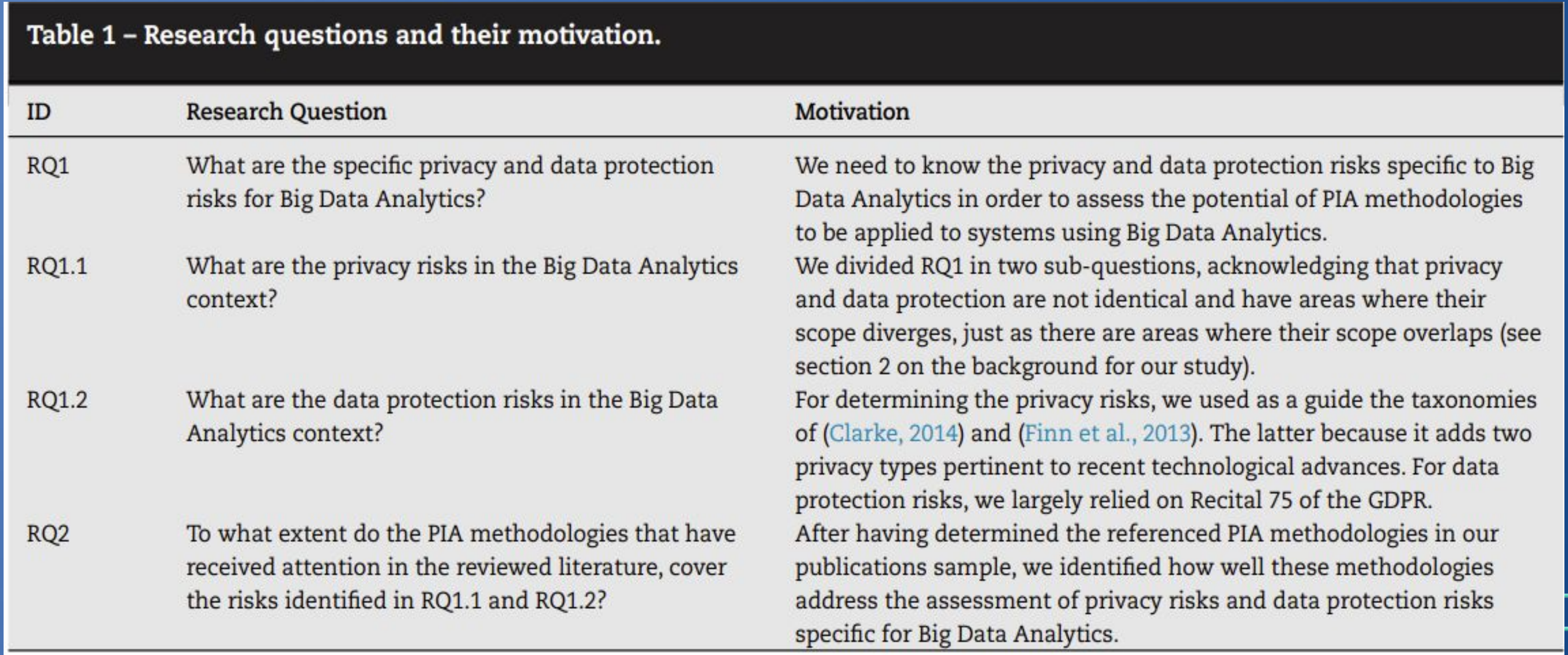


Table 1 – Research questions and their motivation.

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.

RQ1: What are the specific privacy and data protection risks for Big Data Analytics?

1. **Nine** key risks or harms associated with the storage of Big Data and the application of Big Data Analytics.

2. **Four** of these risks relate more to privacy and **five** to data protection.

3. To avoid the abundant use of the term 'risk' in the paper, we named these themes '**Privacy Touch Points' (PTPs)**

4. PRIVACY RISKS: (PTPs (4), (6), (8) and (9))

5. DATA PROTECTION RISKS: (PTPs (1), (2), (3), (5) and (7))

Risks = Privacy Touch Points (PTP)

Data Privacy Risks

- **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

- **Treatment of indirect privacy harms:** societal and ethical concerns aren’t addressed when dealing with big data

- Improper treatment of different types of privacy risks and data breaches

- **Practical issues due to procedural vagueness:** Existing PI methodologies are rather vague and theoretical instead of practical

Data Protection Risks

- **Unclear data controllership:** transfer of big data from one place or recipient to another results in a diverse controllership. Consequently, it is difficult to understand and ensure compliance with privacy requirements amongst data controllers
- **Identification of individuals from derived data:** Big Data Analytics techniques enable inferences from undisclosed information that can be used to identify an individual
- Discrimination issues affecting moral (e.g. stigmatisation) 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

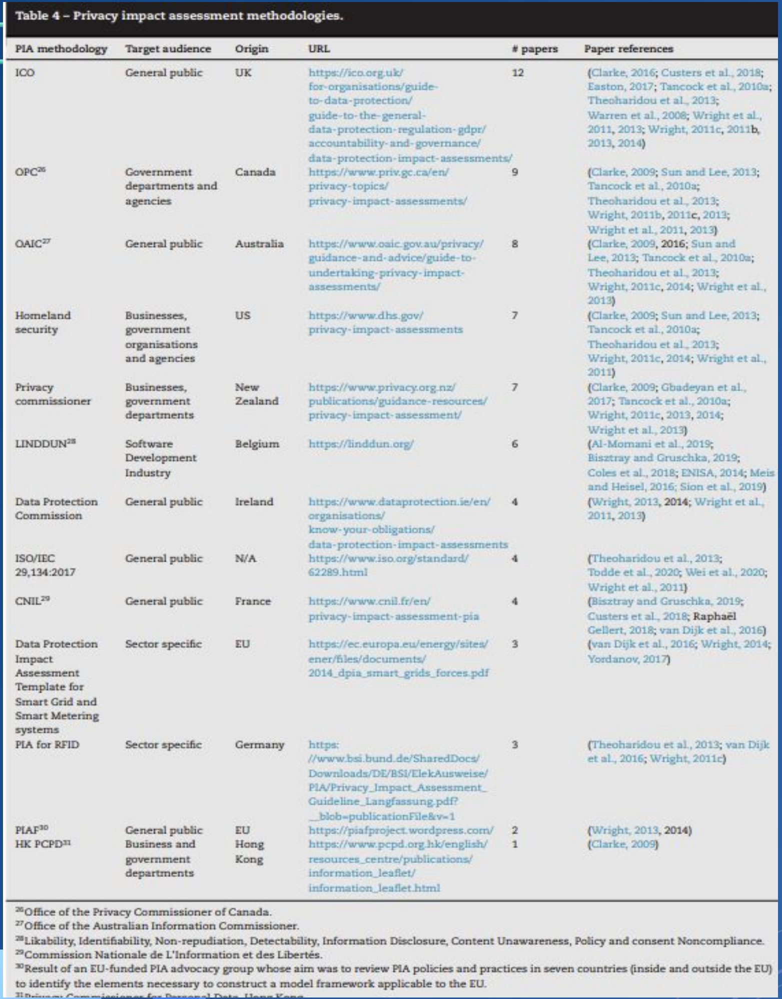


Table 4 – Privacy impact assessment methodologies.

PIA methodology	Target audience	Origin	URL	# papers	Paper references
ICO	General public	UK	https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/	12	(Clarke, 2016; Custers et al., 2018; Easton, 2017; Tancock et al., 2010a; Theoharidou et al., 2013; Warren et al., 2008; Wright et al., 2011, 2013; Wright, 2011c, 2011b, 2013, 2014)
OPC[26]	Government departments and agencies	Canada	https://www.priv.gc.ca/en/privacy-topics/privacy-impact-assessments/	9	(Clarke, 2009; Sun and Lee, 2013; Tancock et al., 2010a; Theoharidou et al., 2013; Wright, 2011b, 2011c, 2013; Wright et al., 2011, 2013)
OAIC[27]	General public	Australia	https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-undertaking-privacy-impact-assessments/	8	(Clarke, 2009, 2016; Sun and Lee, 2013; Tancock et al., 2010a; Theoharidou et al., 2013; Wright, 2011c, 2014; Wright et al., 2013)
Homeland security	Businesses, government organisations and agencies	US	https://www.dhs.gov/privacy-impact-assessments	7	(Clarke, 2009; Sun and Lee, 2013; Tancock et al., 2010a; Theoharidou et al., 2013; Wright, 2011c, 2014; Wright et al., 2011)
Privacy commissioner	Businesses, government departments	New Zealand	https://www.privacy.org.nz/publications/guidance-resources/privacy-impact-assessment/	7	(Clarke, 2009; Gbadeyan et al., 2017; Tancock et al., 2010a; Wright, 2011c, 2013, 2014; Wright et al., 2013)
LINDDUN[28]	Software Development Industry	Belgium	https://linddun.org/	6	(Al-Momani et al., 2019; Bisztray and Gruschka, 2019; Coles et al., 2018; ENISA, 2014; Meis and Heisel, 2016; Sion et al., 2019)
Data Protection Commission	General public	Ireland	https://www.dataprotection.ie/en/organisations/know-your-obligations/data-protection-impact-assessments	4	(Wright, 2013, 2014; Wright et al., 2011, 2013)
ISO/IEC 29,134:2017	General public	N/A	https://www.iso.org/standard/62289.html	4	(Theoharidou et al., 2013; Todde et al., 2020; Wei et al., 2020; Wright et al., 2011)
CNIL[29]	General public	France	https://www.cnil.fr/en/privacy-impact-assessment-pia	4	(Bisztray and Gruschka, 2019; Custers et al., 2018; Raphaël Gellert, 2018; van Dijk et al., 2016)
Data Protection Impact Assessment Template for Smart Grid and Smart Metering systems	Sector specific	EU	https://ec.europa.eu/energy/sites/ener/files/documents/2014_dpia_smart_grids_forces.pdf	3	(van Dijk et al., 2016; Wright, 2014; Yordanov, 2017)
PIA for RFID	Sector specific	Germany	https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/ElekAusweise/PIA/Privacy_Impact_Assessment_Guideline_Langfassung.pdf?__blob=publicationFile&v=1	3	(Theoharidou et al., 2013; van Dijk et al., 2016; Wright, 2011c)
PIAF[30]	General public	EU	https://piafproject.wordpress.com/	2	(Wright, 2013, 2014)
HK PCPD[31]	Business and government departments	Hong Kong	https://www.pcpd.org.hk/english/resources_centre/publications/information_leaflet/information_leaflet.html	1	(Clarke, 2009)

[26] Office of the Privacy Commissioner of Canada.

[27] Office of the Australian Information Commissioner.

[28] Likability, Identifiability, Non-repudiation, Detectability, Information Disclosure, Content Unawareness, Policy and consent Noncompliance.

[29] Commission Nationale de L'Information et des Libertés.

[30] Result of an EU-funded PIA advocacy group whose aim was to review PIA policies and practices in seven countries (inside and outside the EU) to identify the elements necessary to construct a model framework applicable to the EU.

[31] Privacy Commissioner for Personal Data, Hong Kong.

PIA Methodology	Target Audience	Origin
ICO	General Public	UK
OPC	Government Departments	Canada
OAIC	General Public	Australia
Homeland Security	Businesses,government organisations and agencies	US
Privacy commissioner	Businesses,government departments	New Zealand
LINDDUN	Software Development Industry	Belgium
Data Protection Commission	General public	Ireland
ISO/IEC 29,134:2017	General public	N/A
CNIL	General public	France
Data Protection Impact Assessment Template for Smart Grid and Smart Metering systems	Sector specific	EU
PIA for RFID	Sector Specific	Germany
PIAF	General Public	EU
HK PCPD	Business and government departments	Hong Kong

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?

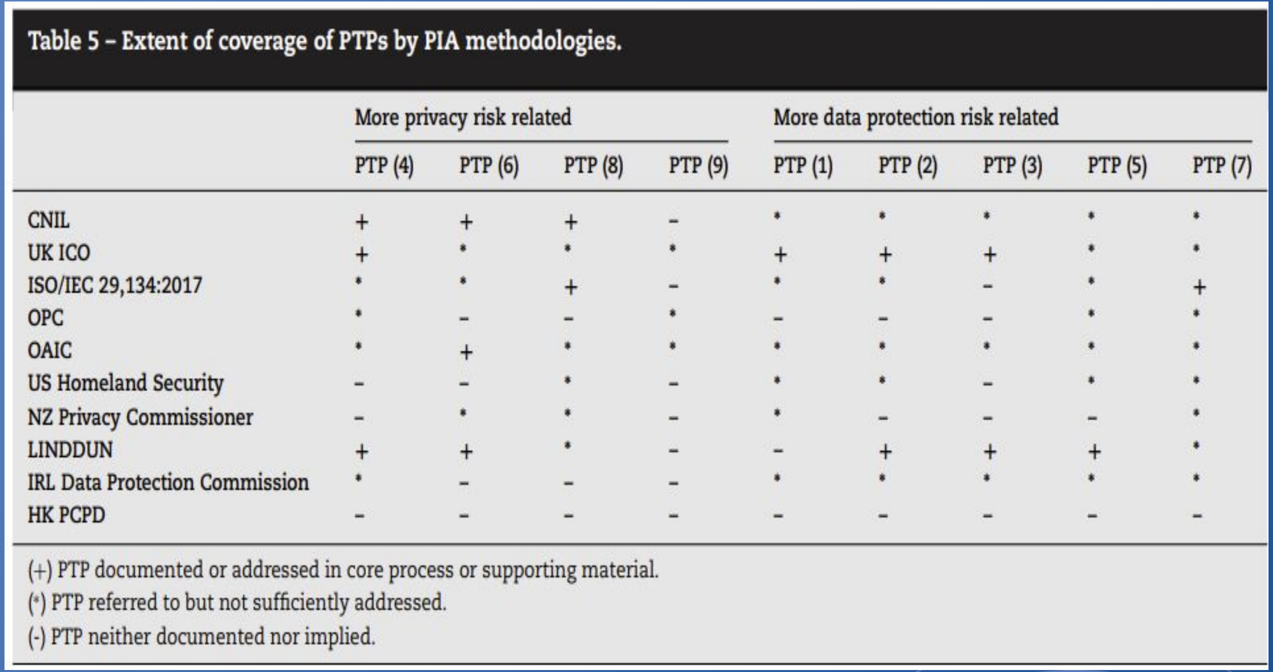


Table 5 – Extent of coverage of PTPs by PIA methodologies.

	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.

- **CNIL:** Good coverage of most privacy related PTPs, but not of the more data protection related PTPs.
- **UK ICO PIA:** Exactly the opposite of CNIL
- **LINDDUN** : Broadest coverage with two out of four privacy related PTPs and three out of five data protection related PTPs sufficiently covered.

Overall, No single PIA methodology appears capable of addressing all Big Data Analytics-specific
privacy and data protection risks !

04.

DISCUSSION AND FUTURE WORK

DISCUSSION

- This literature review provides a qualitative analysis of existing PIA methodologies that could be used to assess privacy or data protection risks for applications relying on Big Data Analytics technologies

- The study also revealed existing PIA methodologies’ weaknesses and strengths with respect to privacy and data protection impact assessment in Big Data Analytics environments

- It also helps obtain a broad understanding of how to better address Big Data Analytics-specific privacy and data protection risks in DPIAs imposed by the GDPR

FUTURE WORK

- To design specific DPIA guidance for the Big Data Analytics context bottom up, but by considering both the nine categories of privacy and data protection risks defined in this paper as Privacy Touch Points and how these risks were addressed in PIA methodologies that cover them well.
- Use Delphi technique by designing questionnaires and sending them to a panel composed of both privacy and big data experts.

05.

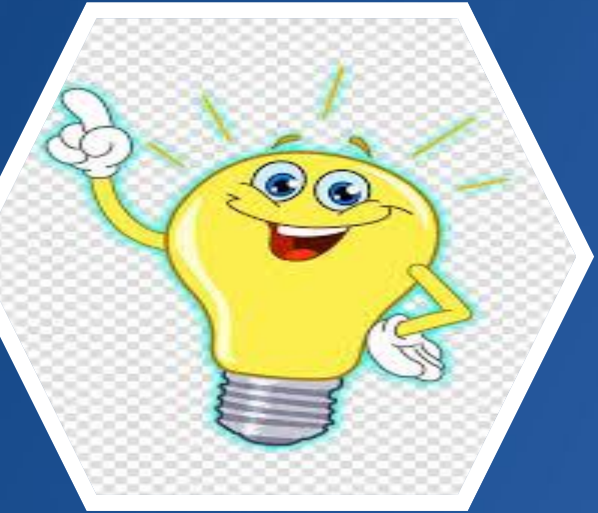
CONCLUSION

CONCLUSION

01. Despite the large number of papers discussing the general use of privacy impact assessments, there is still a need for a methodology more pertinent to privacy and data protection risks in environments storing big data and applying Big Data Analytics algorithms.

02. The research goal is to develop a more comprehensive DPIA methodology to suggest improvements that would enable organisations and data controllers alike to identify weaknesses and possible legal infringements in their processing operations involving Big Data Analytics.

THANK YOU!



Q & A

Since we saw from the paper that the PIA bodies are mostly concentrated in the North America and Europe. Let’s dig a bit deeper into our surrounding.

Are you aware of any Data Privacy Regulatory Bodies in Asia ?

ANSWER

Many countries in Asia have data protection laws or regulations that require organizations to assess the privacy implications of their data processing activities. They might not be as prevalent as the western countries but its growing.

- THAILAND:
  - Name of law: Personal Data Protection Act 2019 (PDPA).
  - Supervisory authority: Personal Data Protection Committee (PDPC).

- SINGAPORE
  - Name of law: Personal Data Protection Act 2012 (2020 Revised Edition) (PDPA).
  - Supervisory authority: Personal Data Protection Commission (PDPC), under the Infocomm Media Development Authority.

- VIETNAM:
  - Name of law: Decree on Personal Data Protection (Decree).
  - Supervisory authority: Ministry of Public Security (MPS).

- JAPAN
  - Name of law: The Act on the Protection of Personal Information (2003) and amendments (last implementation April 2023). Revised guidelines released September 2022.
  - Supervisory authority: Personal Information Protection Committee.

REFERENCES

https://www.dataguidance.com/
\
https://www.privacyworld.blog/privacy-asia-pacific/

https://www.mofo.com/resources/insights/230130-new-wave-of-privacy-laws-in-the-apac-region

https://www.mineos.ai/articles/the-state-of-data-protection-in-asia

https://www2.deloitte.com/content/dam/Deloitte/sg/Documents/risk/sea-risk-unity-diversity-privacy-guide.pdf