Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

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

*Georgios Georgiadis**, *Geert Poels*

Ghent University, Faculty of Economics and Business Administration, Tweekerkenstraat 2, B-9000 Gent, Belgium

ARTICLE INFO

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 enables today's businesses and organisations to process and utilise the raw data that is generated on a daily basis. While Big Data Analytics has improved efficiency and created many opportunities, it has also increased the risk of personal data being compromised or breached. The General Data Protection Regulation (GDPR) mandates Data Protection Impact Assessment (DPIA) as a means of identifying appropriate controls to mitigate risks associated with the protection of personal data. However, little is currently known about how to conduct such a DPIA in a Big Data Analytics context. To this end, we conducted a systematic literature review with the aim of identifying privacy and data protection risks specific to the Big Data Analytics context that could negatively impact individuals' rights and freedoms when they occur. Based on a sample of 159 articles, we applied a thematic analysis to all identified risks which resulted in the definition of nine Privacy Touch Points that summarise the identified risks. The coverage of these Privacy Touch Points was then analysed for ten Privacy Impact Assessment (PIA) methodologies. The insights gained from our analysis will inform the next phase of our research, in which we aim to develop a comprehensive DPIA methodology that will enable data processors and data controllers to identify, analyse and mitigate privacy and data protection risks when storing and processing data involving Big Data Analytics.

1. Introduction

Continuous and fast-growing advances in ubiquitous computing and social media dominate the current technology landscape. These advances have made it possible to collect vast amounts of structured data (e.g. geolocations, stock information), semi-structured data (e.g. emails, XML) and unstructured data (e.g. videos, clickstreams, text). This 'big data' is often too complex to handle with existing data processing frameworks, which has led to the emergence of Big Data Analytics, which combines big data with analytical techniques. Big Data Analytics techniques analyse seemingly unrelated massive data using computer algorithms capable of turning raw data into actionable information (i.e. knowledge) to promote better decision-making (J. Chen et al., 2013), along with a wide

* Corresponding author: Georgios Georgiadis Ghent University, Faculty of Economics and Business Administration, Tweekerkenstraat 2, B-9000 Gent, Belgium.
E-mail addresses: Georgios.Georgiadis@UGent.be (G. Georgiadis), Geert.Poels@UGent.be (G. Poels).

https://doi.org/10.1016/j.clsr.2021.105640
0267-3649/

array of other purposes. Common to all these purposes is the derivation of patterns that can be used to understand and predict behaviour.

While Big Data Analytics has improved efficiency and created significant opportunities, especially for providing new information-intensive services, it has unintentionally increased the risk of data compromise (Klievink et al., 2017) or breaches. In big data, virtually any data element is identifiable and can reveal sensitive information when used in conjunction with other data elements stored in the same or a different location (McMahon et al., 2020; Mourby et al., 2018). Apart from putting their reputations at stake through the incorrect or inappropriate use of data or failing to comply with the General Data Protection Regulation (GDPR) (EP, 2016), organisations incapable of preventing serious data breaches may suffer punitive financial measures as high as 20 million euros or 4% of their annual turnover due to substantial legal repercussions (Salleh and Janczewski, 2016).

The GDPR imposes a 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 as specified in Article 35 (EP, 2016). To this end, organisations must take concrete steps to comply with the legal obligations imposed through the notion of 'data privacy by design and by default' (PbD)[1] in the early stages of designing their products and services. In essence, organisations must embed data protection throughout the life cycle of their data processing whilst remaining aware from the outset of all possible negative impacts that personal data processing operations can have on 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 (Clarke, 2009). Although several studies have been conducted on the use of PIAs, little attention has been paid to understanding whether these assessments require a different approach in the context of Big Data Analytics. Moreover, to the best of our knowledge, the published research on Big Data Analytics has not been systematically reviewed from the point of view of data protection or privacy risk assessments, not even after the GDPR came into force. Other than a description of the privacy issues and the privacy-preserving techniques used in two case-studies of government-funded research projects that required analysing big data and sensitive personal information (Gruschka et al., 2019), the state of the art regarding personal data protection or conducting PIAs in a Big Data Analytics context is therefore unknown. A preceding literature mapping of addressing GDPR requirements in a Big Data Analytics context through Enterprise Architecture Management practices confirmed this lack of procedural knowledge regarding risk management and data governance Georgiadis & Poels, 2021).

Our study fills this gap by conducting a systematic literature review on privacy and personal data protection in environments with Big Data Analytics applications. With our study we make two contributions. First, we systematically review all privacy and data protection risks identified in the literature that are relevant to Big Data Analytics and summarise them into nine Privacy Touch Points. Second, we analyse the coverage of these Privacy Touch Points in ten PIA methodologies that received attention in our publications sample. The results of our study not only reveal strengths and weaknesses of existing PIA methodologies when used in Big Data Analytics environments, but also inform the development of an enhanced methodology for conducting a DPIA in contexts where vast amounts of personal and often sensitive data are processed.

The remainder of this paper is organised as follows: Section 2 provides background information on our study to enable a broader understanding of privacy issues related to Big Data Analytics, the concept of PIA methodologies and the DPIA imposed by GDPR. Section 3 formulates our research questions and describes our research method, which is based on the Systematic Literature Review methodology. Section 4 presents the results of the literature review and analysis to answer the research questions. Section 5 discusses our findings and suggests future research directions. Finally, section 6 concludes the paper.

2. Background

In this section, we first discuss privacy-related issues that have arisen with the use of big data. Next, we present the concept of PIA and briefly discuss and compare methodologies for conducting a PIA. Finally, we introduce the DPIA imposed by GDPR and reflect on how existing PIA methodologies could help in providing guidance for conducting a DPIA in a Big Data Analytics environment.

2.1. *Privacy in the era of big data*

Big Data Analytics requires a complex ecosystem of users, powerful infrastructure, services and applications that store, retrieve and process data from widely scattered sources (Shin and Choi, 2015). Several studies have shown or predicted its potential to create value for a wide range of actors in the global economy (Vesset et al., 2018; Drewer and Miladinova, 2017; Shirer, 2015; Krasnow Waterman and Bruening, 2014; McKinseyandCompany, 2011). Alongside these benefits, there are some growing concerns, the most pressing of which are related to the secure and lawful processing of personal data under the legal framework in force. The GDPR, in force since the beginning of May 2018, is the data protection regulatory framework across all European member states and

[1] The original intention of PbD was to ensure that privacy requirements are taken into account in very early stages of solution development. While PbD is a broad concept that encompasses several dimensions, the GDPR reduced its scope to matters of personal data protection (EDPS, 2018b).

[2] Note that in the literature, the acronyms 'PIA' and 'DPIA', and the terms they stand for are used interchangeably. In this paper, however, any reference to DPIA refers specifically to the GDPR, while we refer to methodologies for conducting privacy and data protection risk assessments as PIA methodologies. Here, we use the term 'methodology' to refer to all types and forms of guidance for conducting PIAs (e.g. processes, guidelines), as is common in the literature (e.g. (Friedewald et al., 2016, p. 22; ICO, 2013; Wright et al., 2014, p. 156; Wright & De Hert, 2012, p. 5)).

plays a fundamental role in shaping digital markets in the European Union (EU). Its general objective is to ensure the fair processing of personal data by setting an array of binding data protection principles, thereby enabling individuals – or data subjects – to exert better control over their data. In this sense, the GDPR aims to enable organisations to better understand and respond to data breaches that occur despite organisations' significant investments in acquiring new competences and changing business practices.

Complying to the GDPR is particularly challenging in environments with Big Data Analytics applications. For example, it has proven difficult to demonstrate the lawfulness of exploratory use of Big Data Analytics algorithms as 'necessary for the performance of a contract' under the GDPR (ICO, 2017). Data ownership and fidelity remain burning issues, with many unresolved cases ending up in court (Kaisler et al., 2013). Moreover, the impact of data breaches can be more severe as they may affect the data of many more people. The GDPR encourages organisations to implement protective measures to secure and control the processing of personal data. To this end, it imposes stringent requirements to the effect that processing operations are better avoided if they are likely to result in a high risk that could negatively impact the rights and freedoms of data subjects, while for activities with limited impact and lower likelihood of risk, organisations are expected to take measures proportionate to the level of risk (EDPS, 2018a).

Some large-scale breaches have received widespread press coverage. Equifax Inc., the oldest of the three largest American credit agencies, faced a serious cybersecurity incident in 2017 that affected the sensitive personal information of approximately 143 million consumers, including European residents (BBC, 2017; Bracy, 2017). The case of Cambridge Analytica involved a more recent large-scale data breach involving the illicit harvesting of 50 million users' personal data posted on Facebook. A recent data breach incident in Europe resulted from the online leak of nearly half a million French people's sensitive data (Libération, 2021) stolen from around 30 medical laboratories. Information published by the Privacy Rights Clearinghouse (PRC, 2017) shows that nearly 8000 data breaches have been made public since 2005 in the United States alone. For Facebook, 2018 was an unfortunate year, with one data breach following another (Frier, 2018; rtbf.be, 2019). The growth in the number and effects of data breach incidents across industry sectors is alarmingly high, as shown in a year-end review report of Theft Resource Centre (ITRC, 2020). Another study (Ponemon Institute, 2020) showed that the average data breach has a total cost of $3.86 million and takes 280 days to identify and resolve. In addition to drastically increasing the number of massive data initiatives benefitting from the power of Big Data Analytics (Sheridan, 2020), the COVID-19 pandemic has increased legal uncertainties regarding the potential penalties for non-compliance with the GDPR, given the large-scale digital surveillance and tracing of individuals' exposure through a wide assortment of hardware devices and app-based solutions (Bradford et al., 2020; Deloitte, 2020). Due to the pandemic, already affected sectors like healthcare and business may remain exposed as Big Data Analytics supported activities, such as customer interactions and clinical trials, have become accessible online due to mobility restrictions (Bages-Amat et al., 2020; Thorlund et al., 2020).

Despite the increasing occurrence of complex breaches, a growing number of companies have developed a keen interest in investing in new technologies whose associated business models build on the processing of large quantities of data (Guggenheim, 2016; IDC, 2019). In addition, many organisations that have outgrown in-house data centres, which suffer from scalability and flexibility limitations, have relocated their IT resources to the cloud (Alali and Yeh, 2012). Such decisions are inherently complex and risky, particularly when considering data flows across jurisdictions in transborder storage (D. Svantesson and Clarke, 2010); the possibility that the data controller will not exert control over data processing (D. J. B. Svantesson, 2012); the many vulnerabilities to which data are exposed (Theoharidou et al., 2013); the proliferation of multiple copies of data (Gloria González Fuster, 1989); and the direct or indirect loss or breach of personally identifiable data (Subashini and Kavitha, 2011), which the Cloud Security Alliance (CSA, 2017) has confirmed remains the top security concern for cloud customers.

State-of-the-art tools and measures are required to properly manage big data's so-called '5Vs' (volume, variety, velocity, value and veracity). This, along with the so-called 'paradoxical triad' (Richards and King, 2013) of Big Data Analytics (identity, transparency and power), implicates privacy concerns (Kaisler et al., 2013; Katal et al., 2013; Labrinidis and Jagadish, 2012) linked to the pervasiveness of data operations. Consequently, some authors have assumed that the very nature of big data contradicts protection goals for privacy engineering (Hansen et al., 2015) and that Big Data Analytics cannot adhere to all data protection principles under the GDPR. In other words, Munir et al. concur with the idea that '*privacy is dead because of [big data]*' (Munir et al., 2015, p. 358).

2.2. *Privacy impact assessments*

Broadly speaking, a PIA is a management tool that helps organisations identify, analyse and minimise privacy risks, whether organisational or technical. Privacy risks can arise from various sources within an organisation – for example, at an early stage in a project's design phase or from threats emanating from the external environment. Sometimes these risks simply arise from the continued use of existing systems, technologies, or processes. PIAs allow organisations to systematically assess privacy risks and their potential effects on an activity, proposal or project that involves the processing of personal data, thereby ensuring privacy by design, as organisations are aware of all possible consequences of their data processing operations from the outset (Cavoukian and others, 2009). By providing organisations with useful insights into their own data processing operations, a PIA can improve informed decision-making by revealing internal communication gaps and hidden assumptions. As a result, PIAs enable organisations to reduce costs in terms of management time, legal expenses, and potential negative media coverage.

Like other types of impact assessments, a PIA enables the identification and assessment of the potential consequences of an activity or proposal, but with a focus on the impact of privacy risks. Some definitions go further than presenting PIA as a risk impact assessment instrument. For example, the ISO/IEC 29,134:2017 guidelines for privacy impact assessment

consider PIA as an element of accountability for providing evidence that an organisation has instituted appropriate measures to avoid privacy risks and prevent breaches (ISO, 2017). We prefer to use in the context of our study the definition of Roger Clark, who defines PIA as "*a systematic process for evaluating the potential effects on privacy of a project, initiative or proposed system or scheme*" (Clarke, 2009, p. 1). Hence, we do not require that PIA is also an instrument for collecting evidence of legal compliance as a PIA does not necessarily refer to a particular legal framework like the DPIA imposed by the GDPR does.

As a concept, PIA emerged and fully developed between 1995 and 2005 (Clarke, 2009; Tancock et al., 2010b; Wright, 2011b). Currently, there are PIA methodologies published by governmental agencies in at least 10 jurisdictions (Clarke, 2011) and for different industries and sectors. Examples include the guidance documents developed by government agencies in English-speaking common law countries such as Australia, Canada, the United States (US), New Zealand (NZ) and the United Kingdom (UK) (OAIC, 2019b; OPC_NZ, 2007; OPC-CA, 2019; OPCL-US, 2012), the Privacy Impact Assessment Guideline for Radio-Frequency Identification applications (M. C. Oetzel et al., 2011) and the Guidance on Privacy Impact Assessment in health and social care (HIQA, 2017). This coexistence of multiple proposals – a phenomenon quite well known in other disciplines (e.g. frameworks in enterprise architecture (Schekkerman, 2004)) – was likely the motivation for the attempt by ISO to standardise the guidelines for conducting a PIA (ISO, 2017).

Some noticeable differences amongst PIA methodologies include the range of methods that can be considered, from a simple, small-scale privacy compliance audit to a full-scale, legally mandated assessment of privacy risks and liabilities with the key objective of earning public trust. Their common characteristics include their use at the outset of new programmes, services and technologies for keeping decision-makers and stakeholders aware of the disturbing implications of data mishandling, such as negative media attention and considerable financial losses (Tancock et al., 2010a). PIAs should not be considered one-off operations. Rather, they should be regularly reviewed and updated considering changes in technology or processing operations. As their underlying processes aim to identify risks and solutions adequate for protecting personal data, their scope is not strictly limited to producing a report that demonstrates compliance.

Opponents of PIAs consider them a costly bureaucratic hassle (Wright, 2011b) or forced exercise to appease the legal team (Easton, 2017). 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. However, a PIA may be revised multiple times during the life cycle of the project or data processing operation under consideration.

Although PIAs may be perceived as integral to project planning and risk management, they differ from other risk management and compliance techniques – such as privacy audits, privacy law compliance checks and privacy issue analysis – in their broader scope and aim as well as in that they are not applicable only to existing systems (Clarke, 2009; OPC_NZ, 2007; Wright, 2011c). Broadly speaking, the scale of a PIA largely depends on the amount and sensitivity of personal information being processed (which often reflects the number of people that may be impacted), the availability of the required resources, the potential for privacy invasion entailed by the involvement of new or additional technologies, and obligatory status (i.e. whether the PIA is mandated by law). For example, conducting separate PIAs for the SAVE and E-Verify programmes (DHS, 2007, 2011) was mandatory under US law.

In terms of actors' engagement in a PIA, a mixture of internal and external stakeholders with different areas of expertise and varying levels of participation is advised. It is often recommended to involve people in different parts of an organisation to ensure the strong endorsement of all interested or affected parties as well as a mix of types of expertise, such as legal, scientific and ethical (Scudder et al., 2018; Wright et al., 2011).

While there are numerous guides of varying origin and quality (Clarke, 2011), no PIA methodology explains in sufficient detail how to carry out a risk impact assessment for different types of privacy issues (De and Le Métayer, 2017; Meis and Heisel, 2016; M. C. Oetzel and Spiekermann, 2014; Wright et al., 2010). Along with the demonstrated reluctance to engage external stakeholders (Clarke, 2011; Puijenbroek and Hoepman, 2017), this has made conducting PIAs a challenging process (Bieker et al., 2018). Various researchers have noted issues with the PIA process, including (Puijenbroek and Hoepman, 2017; Wadhwa, 2012; Wadhwa and Rodrigues, 2013), who identified some common weaknesses while carrying out a field study on the use of PIAs in practical contexts.

2.3. *Data protection impact assessments under the GDPR*

In light of the discussion of PIAs in sub-section 2.2, we consider the 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. Much like a PIA, the DPIA aims to identify and address issues related to the handling of personal data early in a project's life cycle. In doing so, it raises awareness and builds trust by demonstrating an organisation's commitment to processing personal data by secure, transparent and lawful means (WP29, 2017). Accordingly, the reasons for conducting a DPIA and many of the underlying process steps are fundamentally similar to those of a PIA (Yordanov, 2017). It is therefore not uncommon to see both terms to be used indiscriminately in academic and professional articles. Nevertheless, DPIA and PIA are different concepts. Roger Clark (Clarke, 2017) has noted important differences in scope and basic value drivers. The broader scope of PIA goes beyond the protection of personal data to include four additional dimensions: bodily privacy, privacy of personal behaviour, privacy of personal communication, and privacy of personal experience (Bas Seyyar and Geradts, 2020; Clarke, 1999). In other words, a PIA is governed by societal values and has no exclusion or exemption in terms of privacy protection. Some PIAs also go beyond the scope of legal frameworks by addressing, for example, general ethical considerations about privacy and personal data handling (Binns, 2017).

According to Article 35(1) of the GDPR, a 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. Article

35(3) also sets outs three more types of processing operations requiring a DPIA:

1) systematic and extensive profiling with significant effects;
2) large scale use of special categories or 'sensitive' data; and
3) public systematic monitoring on a large scale.

A DPIA may also be required for cases whereby the risk of a processing operation previously labelled as low has changed to high as well as when it is uncertain where a certain processing operation belongs.

In the context of Big Data Analytics, the rights of the data subjects that must be protected and the types of high risks to avoid are not entirely clear, as they are usually dependant on how the law is interpreted (Yordanov, 2017). As a result, the scope of a DPIA has been somewhat uncertain (Gonçalves, 2017), while at the same time there are no concrete requirements for how an organisation should conduct a DPIA. For example, the reference to 'using new technologies' in Article 35(1) is very generic and does not necessarily entail risks arising from the processing of big data. The participation of various stakeholders is stressed in Article 35(9) but without an explanation of how such participation should occur.

In an attempt to improve clarity regarding the processing operations that require a DPIA, supervisory authorities such the ICO and European Data Protection Board (EDPB) have published guidelines (EDPS, 2019; ICO, 2019, pp. 199–200) that set forth a non-exhaustive list of criteria to be used in determining the processing operations of high-risk endeavours. These criteria were further refined into what was already presented in (WP29, 2017) and adopted in the 4 October 2017 revision. Given that certain emerging technologies and practices, such as Big Data Analytics, have significant influence on the rights and freedoms of individuals, there is a continual need to include further instances of data processing (Bieker et al., 2016). A good reason is that the EDPB[3] continuously reviews the lists proposed by the national Data Protection Authorities in accordance with Article 35(4).

Although Roger Clarke's taxonomy of privacy dimensions (Clarke, 2017), (Finn et al., 2013) has been expanded with two additional privacy types to capture the impact of emerging technologies (Finn et al., 2013), his point of view on the dissimilarity between privacy and data protection is consistent with the findings of other scholars (Kokott and Sobotta, 2013) and international Steering Committees papers (Council of Europe, 2014), which studied means by which individuals exercise their right to freedom of expression and information particularly in the context of the EU law. While the process for undertaking the PIA and DPIA types of impact assessment may be similar, their scope and contexts are different. Considering the DPIA's self-imposed limitation to personal data protection and that it is not a risk management approach sensu stricto – the GDPR fosters zero tolerance towards residual risks and does not consider justifiable risks in processing data subjects' data and entering in their personal space – we support (De Hert et al., 2012)'s remark that the terms DPIA and PIA must be used consistently and not interchangeably.

With this background on the specific challenges of safeguarding privacy and protecting personal data in environments with Big Data Analytics applications, PIA (methodologies), and the GDPR's DPIA, the research goal of our study can be more precisely formulated as investigating, based on literature review and analysis, what can be learned from PIA methodologies to perform a DPIA in a Big Data Analytics environment under the GDPR. For this purpose, the aim of our study is to analyse the extent to which existing PIA methodologies cover data protection and privacy risks specific to Big Data Analytics.

3. Research method

3.1. *Research approach*

This paper surveyed recent and publicly accessible papers from prominent journals, conferences and Data Protection Authorities using a Systematic Literature Review approach. The original aim was to study current and still-in-use PIA methodologies and thereby search for any privacy or data protection risk assessment approaches that have been developed for the Big Data Analytics context. As it soon turned out that this approach would not deliver much concrete results, we changed our strategy. Our new aims were to identify, based on the reviewed literature, the specific privacy and data protection risks incurred in a Big Data Analytics context and to determine the degree to which PIA methodologies covered these risks. The objective of our study was to obtain information that would enable us to develop a methodology for conducting a DPIA that addresses Big Data Analytics specific privacy and personal data protection risks while taking into consideration the underlying GDPR legal requirements. With this paper, we hope to advance knowledge relevant to real-life organisational problems of data governance in environments employing Big Data Analytics (Floridi, 2006).

3.2. *Research questions*

With the above aims and research objective in mind, our literature review and analysis addressed the following research questions:

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

Answering RQ1 requires the investigation of two sub-questions, RQ1.1 and RQ1.2, related to the identification of privacy risks and data protection risks, respectively. RQ1 will be answered based on what we find in the reviewed literature.

(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?*

The answer to RQ2 will require an analysis of our own as we found that this research question has not been addressed in the literature. However, for the selection of PIA methodologies, we used the reviewed literature.

[3] The EDPB is an independent European body composed of representatives of the national DPAs and the European Data Protection Supervisor (EDPS). Its purpose is to promote cooperation and ensure the consistent application of data protection rules across the EU.

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.

Table 1 presents a more elaborate motivation for each research (sub-)question.

3.3. *Systemic literature review process*

Our literature review was conducted systematically following the guidelines originally advocated by (Kitchenham, 2007; Tranfield et al., 2003) and subsequently refined by (Turner et al., 2010). The Systematic Literature Review methodology goes beyond an overview and annotated bibliography by suggesting a scientifically rigorous synthesis of primary studies. It delves deeply into the literature and is comprehensive in scope, including all relevant material. As such, it enables researchers to identify gaps in the current literature, offers explicit and full explanations of how the research was carried out and follows a systematic approach. In other words, it is repeatable and consistent. In addition to its narrower scope, which requires a more thorough search process and quality evaluation (Petersen et al., 2015), a distinguishing feature is its evidence-driven nature.

Our review involved several activities, shown in Fig. 1, and consisted of three phases: planning, conducting, and reporting. Schematically, Fig. 1 shows the associated paper's (sub-)section for each activity. The planning stage was partly reported in sub-sections 3.1 and 3.2, where we defined our research objective, aims and research questions. The review protocol comprises the list of data sources, search strategy (i.e. string(s) of search terms), the period coverage of the primary studies and the inclusion and exclusion criteria for paper screening. Next, in the conducting stage, the selected search strings were used to search the online databases based on the review protocol. The search results we obtained were analysed to identify pertinent studies based on the inclusion and exclusion criteria defined in the planning phase. Finally, in the reporting stage, we discussed the results of the review in sufficient detail by combining the facts gathered from the studies and reported the findings while answering the research questions.

3.4. *Defining the review protocol and conducting the search*

Due to the multidisciplinary nature of our research topic, we considered the databases with the most returned records in our preceding literature mapping of addressing GDPR requirements in a Big Data Analytics context (Georgiadis & Poels, 2021). We therefore decided to include the following digital databases: ACM,[4] EBSCOhost,[5] ScienceDirect,[6] HeinOnline,[7] IEEE Xplore,[8] International Data Privacy Law,[9] Web of Science,[10] ProQuest,[11] Scopus[12] and Taylor & Francis Online.[13]

Our research questions were used to determine the search keywords. To develop the search strings, we combined the terms 'privacy', 'assessment', 'risk', 'evaluation', 'impact' and 'big data' or their abbreviations with Boolean (AND, OR) operators, yielding the following two search strings:

1) *'privacy impact assessment' OR pia OR 'privacy impact statement' OR 'data protection impact assessment' OR dpia OR 'impact assessment' OR 'privacy risk assessment' OR 'privacy risk' OR 'privacy evaluation' OR 'data protection risk assessment' OR 'data protection risk' OR 'data protection evaluation'*
2) *('privacy impact assessment' OR pia OR 'privacy impact statement' OR 'data protection impact assessment' OR dpia OR 'impact assessment' OR 'privacy risk assessment' OR 'privacy risk' OR 'privacy evaluation' OR 'data protection risk assessment' OR 'data protection risk' OR 'data protection evaluation') AND 'big data'*

[4] ACM (dl.acm.org)
[5] EBSCOhost (www.ebsco.com)
[6] ScienceDirect (www.sciencedirect.com)
[7] HeinOnline (heinonline.org)
[8] IEEE Xplore (ieeexplore.ieee.org)
[9] International Data Privacy Law (academic.oup.com)
[10] Web of Science (apps.webofknowledge.com)
[11] ProQuest (search.proquest.com)
[12] Scopus (www.scopus.com)
[13] Taylor & Francis Online (www.tandfonline.com)

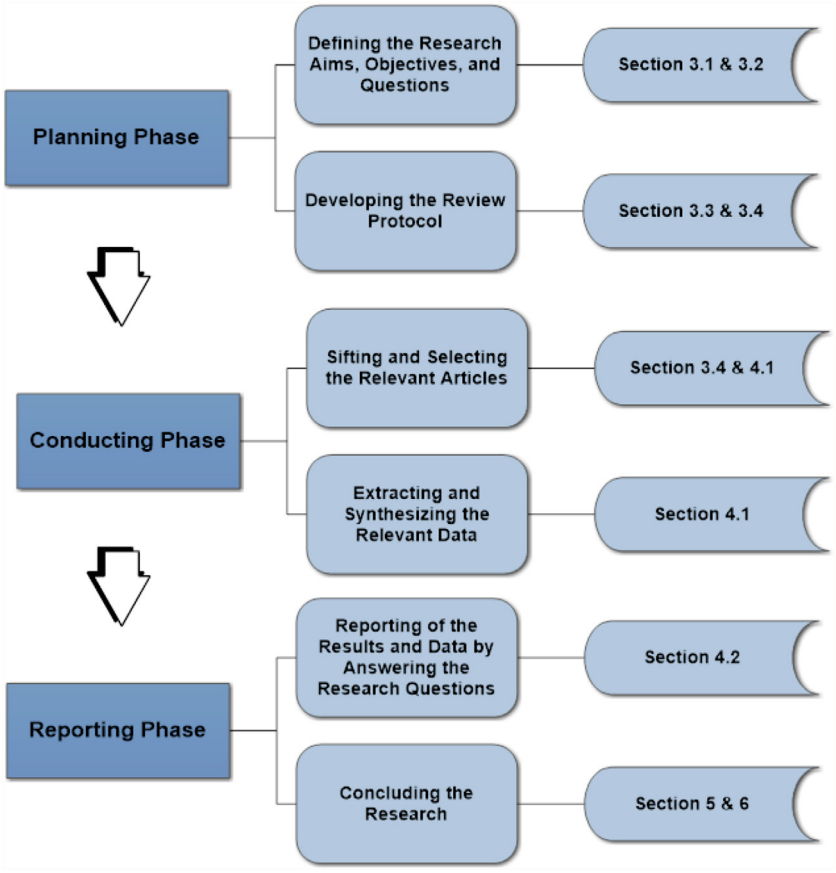

Fig. 1 – Systematic literature review process phases and activities (based on (Kitchenham, 2007; Tranfield et al., 2003), (Turner et al., 2010)).

Table 2 – Inclusion and exclusion criteria.

Type	Description	Reason
Inclusion Criteria	Peer-reviewed publications or documents available on the Data Protection Authorities websites	We mostly targeted journal papers and conference proceedings. Because most PIA methodologies are promoted and typically maintained by Data Protection Authorities, we also added pertinent article references found on their respective websites.
	Contributions published after January 1995	This point in time coincides with when the PIA gained momentum as a tool. The DPIA appeared much later, in 2013, given the history of the GDPR.
	Contributions with a focus on the GDPR or similar legal frameworks, privacy or data protection risks and big data	To study only articles relevant to our domains of interest.
Exclusion Criteria	Paper not written in English	Our language of understanding.
	Focus is mainly on technologies/ tools (and hence not related to the research questions)	We are interested in studying approaches or methodologies and therefore excluded technical solutions that did not provide guidance on how to conduct a PIA.
	Duplicated contributions	To identify and eliminate all papers with the same title or content.
	Papers with no full-text version	A Systematic Literature Review requires the complete content of a publication to be reviewed.

The purpose of the first string was to obtain an overview of the research area and publication space, whereas the second looked specifically for risk assessments of any type involving big data.

Where a database's functionality allowed, we selected peer-reviewed articles from the disciplines of computer science, the social sciences, engineering, business, finance, management, information science and law. We also carried out a parallel manual search using the names of authors and the

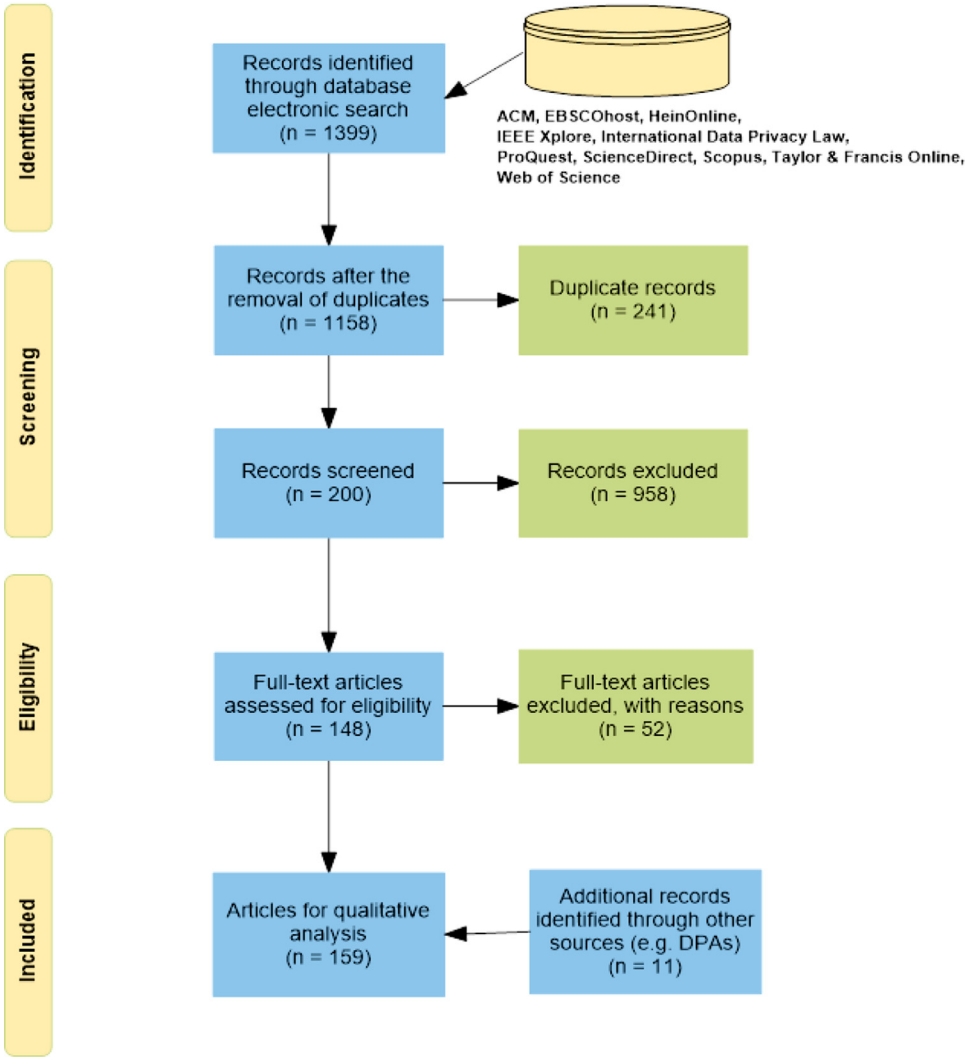

Fig. 2 – PRISMA flowchart.

works cited in the reference section of the retrieved papers and applied the backward and forward snowballing technique described by (Jalali and Wohlin, 2012). Due to the historic nature of the PIA methodologies, we also looked for articles and other pertinent online material, such as guides or white papers, on the websites of their origin Data Protection Authorities.

Table 2 depicts our inclusion and exclusion criteria. Papers that were irrelevant to the research questions were excluded, as were those in a language other than English. We considered publications from January 1995 onwards, as this point coincides with when the PIA gained momentum as a privacy tool (Clarke, 2009).

4. Review and analysis results

4.1. *Sifting and selecting relevant articles*

The paper screening followed the steps outlined in the PRISMA statement (Liberati et al., 2009), illustrated in Fig. 2. PRISMA is an evidence-based methodology that is often used for reporting systematic literature reviews. Following PRISMA in the reporting improves transparency and completeness. Each step aims to ensure that the systematic literature review can be replicated or reiterated in the future.

Our search returned 1399 papers from the 10 databases listed in Table 3, which presents the papers found in each database along with the number retained after each phase. The PRISMA process began by storing the articles yielded by the electronic database searches in separate groups in the EndNote 20 reference management application. As we used a mix of publisher-dependant databases, such as Taylor & Francis Online, and independent databases, such as Scopus, we encountered various duplicate papers. EndNote allowed us to easily detect and eliminate duplicates within each group (i.e. internal duplicates) and across groups (i.e., external duplicates). After eliminating duplicates, 1158 documents remained.

During screening, we examined article titles and abstracts against the inclusion and exclusion criteria listed in Table 2 to ensure that they matched our topic of interest. In total, 200

Table 3 – List of online databases and number of papers retrieved from each database.

Database	# Initially Found (including duplicates)	# After Screening based on title and abstract	# After Eligibility Check based on full paper
ACM	38	6	1
EBSCOhost	4	2	0
ScienceDirect	233	47	28
HeinOnline	8	8	7
IEEE Xplore	176	25	17
International Data Privacy Law	13	11	9
Web of Science	199	15	4
ProQuest	174	29	25
Scopus	168	48	53
Taylor & Francis Online	386	9	4
	1399	**200**	**148**

unique articles remained after screening. Next, in the eligibility check step, we re-examined the inclusion criterion 'Contributions with a focus on the GDPR or similar legal frameworks, privacy or data protection risks and big data' and the exclusion criterion 'Focus is mainly on technologies/tools (and hence not related to the research questions)' based on the full-text versions of the papers. After this eligibility check, 148 papers remained in our publications sample. An additional 11 articles were then added to this sample, resulting in a final sample of 159 articles. These 11 articles were discovered on the Data Protection Authorities websites and after applying the backward and forward snowballing technique described in (Jalali and Wohlin, 2012).

Since analysing the full texts of these 159 articles required a lot of coding work, we used MAXQDA 2020 software to organise the PDF documents of the articles, which had previously been stored in EndNote along with their metadata, into (sub)groups based on their source of origin and areas of interest, database name, privacy and data protection issues, Big Data concerns, regulation & guidelines, impact assessment methodologies, and other concerns. We then used the features of MAXQDA to extract relevant text passages from each document and categorise them under thematic groups (codes) relevant to our research questions. These codes were identified using the five generic value chain phases for Big Data Analytics defined and mapped in (ENISA, 2015). This made it easier for us to identify the specific privacy (RQ1.1) and data protection risks (RQ1.2) for environments where Big Data Analytics applications are deployed, group the papers that justify our definition of each of these risks, and select the PIA methodologies that we analyse in terms of the extent to which these risks are covered (RQ2).

4.2. *Answers to research questions*

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

In our literature review, we identified nine key risks or harms associated with the storage of Big Data and the application of Big Data Analytics. Four of these risks relate more to privacy and five to data protection. The identification of these risks was done using an iterative research technique called Template Analysis,[14] a generic form of thematic analysis that offers a structured approach to hierarchical data coding. Initially, we started with an a priori code structure consisting of five codes corresponding to the value chain phases of Big Data Analytics defined in (ENISA, 2015). The five value chain phases for Big Data Analytics are: data acquisition and collection, data analysis, data curation, data storage, and data usage. These were mapped 'many-to-many' to the following Privacy by Design strategies: minimize, aggregate, hide, inform, control, separate, enforce, demonstrate (e.g. strategies relevant to data storage are hide and separate). We then assigned codes to segment the text of the examined articles. In this process, we identified additional codes and subcodes corresponding to recurring themes. Lastly, we organised themes into clusters to produce the most general higher-order themes, which provided the classification dimensions of our schema. To avoid the abundant use of the term 'risk' in the paper, we named these themes 'privacy touch points' (PTPs). So, PTPs are what we defined as the four privacy risks (PTPs (4), (6), (8) and (9)) and five data protection risks (PTPs (1), (2), (3), (5) and (7)) specific to Big Data Analytics, based on our analysis of the reviewed literature.

Although privacy and data protection risks are not entirely equivalent, there is no definitive agreement on the definition of the risks that fall under these two terms. For the purposes of this paper, we have relied on (Clarke, 2017), (Finn et al., 2013) taxonomy of privacy dimensions and the definition of Recital 75 of the GDRP. For example, we classified PTP (1) as a data protection risk because data controller is a core role in the GDPR while one could argue that PTP (3) is both. The reason why we classified it as (more) data protection related is purely subjective and related to the Recital 75, which states that '…here the processing may give rise to discrimination, identity theft or fraud, financial loss, damage of reputation, …'.

PTP(1) Unclear data controllership. The 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. Existing mechanisms that rely on privacy by pol-

[14] https://research.hud.ac.uk/research-subjects/human-health/template-analysis/what-is-template-analysis/

icy – such as (informed and explicit) consent and privacy notices or 'notice-and-consent' models (Mantelero, 2016) – fail to provide transparency and control sufficient for individuals to understand how their data have been collected and used (Al-Fedaghi, 2012; Gruschka et al., 2019; Mantelero, 2014; Zarsky', 2017). In this regard, Barocas and Nissenbaum (Barocas and Nissenbaum, 2014b) argued that notices should cover information that Big Data Analytics might yield.

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[15] (Anindya et al., 2017; Bertino, 2015; Crawford and Schultz, 2014; Kumar, 2012; Mantelero, 2016; Sampson, 2014). Such features have made re-identification easier, arguably to the point of undermining the value of anonymisation (Altman et al., 2018; Barocas and Nissenbaum, 2014a; Gruschka et al., 2019; Ohm, 2010). (Narayanan and Shmatikov, 2008; Sweeney, 2002), for instance, demonstrated how the linking of data sets can render the benefits of anonymisation unviable. This has had a direct impact on medical research, in which huge quantities of data are aggregated from different sources, as it is important for practitioners to have a full and accurate picture of several people's medical histories (Birnhack, 2019).

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 (Barocas and Selbst, 2016; Favaretto et al., 2019; MacCarthy, 2018; White House, 2016) or 'data determinism' (Ramirez, 2013, p. 7), a by-product of Big Data Analytics, may cause non-material damage to data subjects in accordance with Recital 75 of the GDPR. To illustrate this phenomenon, IDC used the concept of the 'digital shadow' (Gantz and Reinsel, 2011) to describe the amount of data about a particular individual that is collected, organised and analysed. Similar to 'smart' surveillance (Wright et al., 2010) and data-sharing suggestions to avoid the so-called 'digital dark age' (Jeffrey, 2012), the problem with digital shadowing lies in the inability of privacy laws to control how information may be used by enforcement agencies, companies, groups or individuals themselves in different geographical locations.

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 (Gloria González Fuster, 1989; Rubinstein, 2012). 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 (German Federal and State Data Protection Commissioners, 2016). The loss of data subjects' trust in the processing of their data could result in diminished data quality (ENISA, 2015), a critical factor in producing reliable analytics (Jugulum, 2016).

PTP(5) Increased scope leading to further processing incompatible with the initial purpose. New types of privacy threats could result from the ability of big data to recontextualise data (Birnhack, 2019), combined with the anticipated 11.9% growth rate of further investments in Big Data Analytics solutions over the 2017–2022 period (Vesset et al., 2018) and the increasing exposure of personal data in data-fertile fields such as healthcare, manufacturing, banking and the public sector (Kumar, 2012; Quinn and Quinn, 2018). 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.[16] Unlike other types of personal data, an individual's genetic or human genomic sequence data are immutable (Gostin, 1995). Even with today's computing power, the processing of these data may reveal abundant sensitive[17] information about an individual that bears no relation to what was originally intended. Worse still, attempts to protect these data using techniques like anonymisation are not invincible to re-identification attacks (El Emam et al., 2011).

PTP(6) Improper treatment of different types of privacy 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 (Bertot and Choi, 2013; Buhr and Kleiner, 2012), big data–sharing activities are becoming commonplace. The same applies for the new Big Data Analytics approaches that have gradually appeared in the market following the as-a-Service model (Ardagna et al., 2016). These could lead to new types of data breaches and, in turn, inadequate handing of personal information (Alshehri and Drew, 2010; Otjacques et al., 2007). In addition, as field researchers and practitioners have remarked, existing security schemes and technologies strive to satisfy pertinent requirements and expectations (H. Chen and Yan, 2016; Nelson and Olovsson, 2016; Sagiroglu and Sinanc, 2013). However, current research confirms that the security and privacy of big data have received relatively less focus despite being strategically vital for organisations (Akoka et al., 2017; Liu et al., 2017). As a rich organisational asset with distinct features, big data 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 (Mai, 2016). This poses a new type of privacy and security threat in the context of the GDPR (Akma et al., 2018; Nadimpalli and Kumari, 2012). 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) of the GDPR considers optional the participation of external (to the organisation) stakeholders whose data are predominantly processed in a Big Data Analytics context. However, to ensure fairness and transparency, the perspectives and concerns of a broad range of stakeholders are vital in risk assessment

[15] Sometimes referred to as auxiliary information (Ohm, 2010).

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

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

(Wright, 2013), as they allow for the building of less privacy-invasive systems and processing operations (Wright, 2011b).

PTP(8) Practical issues due to procedural vagueness. Existing 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 (Ahmadian et al., 2018; M. Oetzel and Spiekermann, 2017), including an objective way to assess and present privacy impacts and monitor the efficiency of countermeasures. To properly assess privacy 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 (Ferra et al., 2020). 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 of the GDPR.

PTP(9) Treatment of indirect privacy harms.[18] (Crockett et al., 2018; C. D. Raab, 2020; Wright and De Hert, 2016) 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 (EESC, 2016; Ienca and Vayena, 2020; Zwitter and Gstrein, 2020). 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 (Mantelero and Vaciago, 2013). 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 (Barocas and Nissenbaum, 2014b). In this context, the European Data Protection Supervisor and EDPB (EDPS, 2015) have strongly endorsed ethics in the processing of big data to emphasise (as have many scholars) that the real problem arises in use (Krasnow Waterman and Bruening, 2014). The rapid emergence of Big Data Analytics has made it practically impossible to explain all the possible uses of data at the time they are initially gathered.

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?

We first discuss our selection of PIA methodologies and next investigate the extent to which they cover the nine PTPs that answer RQ1.

4.2.1. *Selection of PIA methodologies*

We identified 13 established PIA methodologies in our publication sample (Table 4). Twenty articles referred to the DPIA imposed by the GDPR (i.e. (Bu-Pasha, 2020; Bisztray and Gruschka, 2019; Coles et al., 2018; Crockett et al., 2018; Custers et al., 2018; Raphaël Gellert, 2018; Drewer and Miladinova, 2017; Easton, 2017; Raphael Gellert, 2017; Gonçalves, 2017; Edwards et al., 2016; Mantelero, 2014; Notario et al., 2015; Puijenbroek and Hoepman, 2017; Quinn and Quinn, 2018; Todde et al., 2020; van Dijk et al., 2016; Wei et al., 2020; Wright and Raab, 2014; Yordanov, 2017)). The EU DPIA has likely received interest with the introduction of the GDPR as the new data protection regulation in Europe and because it mandates impact assessments for privacy-vulnerable data processing operations. More than half of the 159 articles in our publication sample were published after 2015, which corresponds to the part of the timeline[19] where the GDPR's final text was more well known. As the EU DPIA is merely a general guideline for conducting a data protection impact assessment, and not a methodology, it is not included in Table 4.

The most referenced PIA methodology in Table 4 is the ICO PIA methodology from the UK government. The UK's ICO has been a longstanding advocate and pioneer in Europe for making the first guidance documents for conducting a PIA publicly available (Clarke, 2009). Nowadays, the ICO PIA methodology – as well as many other European PIA methodologies, excluding those that are sector specific – is practically similar to the EU DPIA in terms of content, however, its level of methodological guidance is different.

Apart from the ICO PIA methodology, the other twelve established PIA methodologies in Table 4 were mentioned in one to nine papers. Some papers (i.e. (Ahmadian et al., 2018; Di Iorio et al., 2020; Edwards et al., 2016; Himmel et al., 2015; Joyee De and Le Métayer, 2016; Mantelero, 2018; Meis and Heisel, 2016; Notario et al., 2015; Sion et al., 2019; Sun and Lee, 2013; Wei et al., 2020; Wright and Raab, 2012)) discussed custom-built methodologies or processes for improving existing PIA frameworks. These papers clearly show that researchers have already identified the need to complement or improve certain parts of the impact assessment methodologies, which, according to them, were not sufficiently developed to cover the needs of emerging technologies (Scudder et al., 2018) or were somewhat obscured and not made to address complex cases (Meis and Heisel, 2015). For example, the developers of PRIAM[20] (Joyee De and Le Métayer, 2016) argued that privacy risk analysis (which they described as the 'technical part' of a PIA methodology) is not well specified or developed in any PIA methodology. To fill this gap, PRIAM put forth a detailed model comprising seven components: system, stakeholders, data, risk sources, privacy weaknesses, feared events and harms. The model

[18] 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 norms (Joyee De and Le Métayer, 2016). Privacy harms are often used to assess privacy risks based on risk sources and weaknesses (Sion et al., 2019).

[19] https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en

[20] Privacy Risk Analysis Methodology

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.

is supported by a two-step process: information gathering and risk assessment. Along the same lines were proposed PRIPARE[21] (Notario et al., 2015; Garcia et al., 2015), pISRA[22] (Wei et al., 2020), PIAMS[23] (Sun and Lee, 2013), Privacy Points (Himmel et al., 2015) and the model-based privacy analysis approach discussed in (Ahmadian et al., 2018). These research proposals address risk analysis through software engineering techniques and quantitative methods. From their authors' point of view, PIAs are primarily intended to cover legal aspects, whereas compliance with PbD principles requires an engineering approach based on systems design. We believe this is not entirely correct. Recently published approaches for conducting DPIAs, such as CNIL's February 2018 edition (see Table 4), have been accompanied by supporting materials that cover technical matters, although these are predominantly served in the form of compliance checklists.

Other authors, such as (Barocas and Nissenbaum, 2014b; Edwards et al., 2016; Mantelero, 2018; Tene and Polonetsky, 2013; Wright and Friedewald, 2013; Wright and Raab, 2012), argue that, parallel to technological assessment, it is imperative to consider rights and values by assessing the impact of other factors, such as ethical and societal values. Due to the particularities of big data, it is important to ensure transparency in the design of algorithms, sustainability through sound resource usage, lawful data minimisation and sharing in the future reuse or generation of (new) data and ultimately the impact on groups and society at large (C. Raab and Wright, 2012; Rhoen and Feng, 2018). Given the lack of adequate coverage of a wide range of possible issues and concerns, including those discussed above, some authors have argued that existing PIA methodologies must be broadened in scope to embrace a more comprehensive standpoint (Agarwal, 2016; Kloza et al., 2018; Mantelero, 2018; Tancock et al., 2013; Wright, 2011a). For instance, (Di Iorio et al., 2020) and (Alnemr et al., 2015) extended the scope of the DPIA to address privacy risks inherent to data governance, ethics and cloud services. In the same vein, (Wright and Friedewald, 2013) advocated the integration of the DPIA with Ethical Impact Assessments (EIAs), as both follow similar processes and legal issues are often intertwined with ethical matters in the development of new technologies.

As a final note, we wish to observe that an ethical approach to personal data processing goes beyond legal compliance. Enterprises and organisations need to ensure that data is processed in a way that is not ethically questionable and does not result in unfair outcomes for individuals or adverse effects on society, thereby infringing on fairness principles. Following the set of guidelines issued by the European Commission (EC, 2019), and due to the multifaceted nature of ethics and its relevance to big data's underlying technologies, we noticed in our review that some Data Protection Authorities have started to raise awareness of this matter (PCPD, 2019).

4.2.2. *Coverage of privacy and data protection risks specific to big data analytics*

Regarding the assessment of the extent of coverage by the PIA methodologies in Table 4 of the nine PTPs that answer RQ1, it is important to note that the findings are based on our own analysis and are thus not directly found in the reviewed literature. Table 5 shows our results in terms of the following three classes:

1) PTP addressed in core process or supporting material (+)
2) PTP referred to but not sufficiently addressed (*)
3) PTP neither documented nor implied (-)

To assess the coverage of the PTPs for each PIA framework, we analysed their texts or other relevant materials posted on the associated websites. The PIAF methodology was not included in our assessment, as its original purpose was to identify elements from existing PIA methodologies to develop an EU model framework (De Hert et al., 2012). We also excluded the two sector specific PIAs methodologies for smart grids and RFID. The analysis of the remaining PIA methodologies is explained next.

4.2.3. *CNIL*

CNIL's PIA methodology offers a good coverage of the more privacy risks related PTPs. Because of its well-rounded set of supporting materials[24] in the form of templates, checklists and a software tool, it adequately addresses the processing operations related to PTPs (4), (6) and particularly (8), despite the issue of integrating some of the materials with its guiding process (Bisztray and Gruschka, 2019). However, given that it insufficiently addresses PTPs (1)-(3), (5) and (8), it must undergo further development for addressing data protection risks in the Big Data Analytics context. Finally, PTP (9) is not covered at all although ethical and societal concerns for the use of algorithms were presented in one of its reports back in 2017 (CNIL, 2017).

4.2.4. *UK ICO*

The UK ICO PIA methodology can be regarded as an elaborated version of the DPIA. It addresses many of the identified PTPs in one of its complementary guides (Butterworth, 2018; ICO, 2017), where it provides general instructions on conducting a PIA for big data. In the same guide, it offers strategies to mitigate certain big data risks in each DPIA process step. However, we believe that this document does not cover all PTPs in sufficient depth, in particular PTPs (5) to (9), and falls short of providing a set of tools (e.g. templates) or a detailed method for carrying out a privacy risk assessment in a Big Data Analytics context. Concerning the PTP (9), we noted that ICO had launched an online consultation about views on data handling ethics across all sectors and organisational sizes with a closing date of 8 January 2021. The description published on ICO's

[21] PReparing Industry to Privacy by design by supporting its Application in Research

[22] privacy considered Information Security Risk Assessment Model

[23] Privacy Impact Assessment Management System

[24] This comprises four documents: methodology, knowledge bases, application to IoT devices and templates. The methodology is also accompanied with a open source application for carrying out the DPIA electronically.

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.

website[25] specifically mentioned the importance of considering ethics for 'particularly complex situations' when undertaking a DPIA. We may therefore expect a better coverage of this PTP in future releases of this PIA methodology.

4.2.4.1. *ISO/IEC 29,134:2017* ISO/IEC 29,134:2017 focuses heavily on defining processes closely resembling those found in project management methodologies. This likely explains the meticulous attention paid to consulting stakeholders to determine the scope of the PIA (i.e. PTP (7)). ISO/IEC 29,134:2017 offers practical guidance by specifying the objective, input, expected output and actions for each step, including a well-structured report. Hence, it received a good evaluation for PTP (8). It also includes example scales and criteria for estimating the level of impact and likelihood of generic privacy risks but does not address risks specific to big data. Hence, PTPs (3) and (9) are neither implied nor documented.

4.2.4.2. *OPC* OPC's guide to PIAs focuses solely on federal public sector institutions. The guide refers to all privacy matters and includes an extra phase for risk analysis. Although it addresses matters related to data collection, use, disclosure, retention, quality (through accuracy) and transparency (through openness), it is highly generic, as none of its core topics refers to the handling of big data. Nonetheless, articles posted on the OPC website discuss some broad implications of big data, including those linked to societal and ethical matters (Bennett and Bayley, 2015; Kosseim, 2016).

4.2.4.3. *OAIC* OAIC's guide is geared towards the private sector and government agencies. Its methodology is quite broad and covers several data privacy concerns, including data/information quality, retention, security and risk management (i.e. PTP (6)). However, it does not incorporate processes related to big data or Big Data Analytics, although pertinent information is found in separate guides available on its website (OAIC, 2018, 2019a) .

[25] https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-consultation-on-the-role-of-data-ethics-in-complying-with-the-gdpr/

4.2.4.4. *US homeland security* The PIA methodology from the US Department of Homeland Security is primarily geared towards government departmental programmes and systems. As such, it focuses on information stored in government systems and new or substantially changed technologies used to process personally identifiable information. It introduces a number of roles parallel to those in the DPIA whilst placing greater emphasis on the use of the term 'privacy' as opposed to 'data protection'. For example, the tasks and responsibilities of the Data Protection Officer in the DPIA are carried out by the so-called Chief Privacy Officer in the Department of Homeland Security. In terms of content, the methodology is fairly high-level and thus does not provide in-depth guidance for assessing the impact of Big Data Analytics projects. PTPs (2), (5), (7), and (8) are briefly discussed in some of its PIA example documents, such as (DHS, 2014) and (DHS, 2020), available on the US Department of Homeland Security website.

4.2.4.5. *NZ privacy commissioner* The PIA toolkit of the NZ Privacy Commissioner targets both businesses and government organisations and is more detailed than the US Department of Homeland Security document in terms of content, making it relatively easy to apply in a practical context. On the other hand, apart from short online articles and references to other websites, we could not find any substantial document explaining how to conduct a PIA in the Big Data Analytics context. Similar risks to those found in Big Data Analytics environments were discussed in a paper on the Commissioner's position on biometric data (NZPC, 2021) published late in 2021, but again, the role of PIAs and the necessity of considering privacy concerns were simply noted.

4.2.4.6. *LINDDUN* LINDDUN was designed as the privacy-related equivalent of the STRIDE method (Johnstone, 2010). It is a privacy threat analysis methodology used to identify privacy-related vulnerabilities in the early stages of a product development process (Al-Momani et al., 2019; Deng et al., 2011). As such, its primary focus is on the technical and engineering process aspects of privacy, paying lesser attention to legal compliance perspective (Bisztray and Gruschka, 2019; Wuyts and Joosen, 2015). This shortcoming has been noted by the DistriNet research group,

which maintains the methodology and a proposal for how to rectify the issue is outlined in one of the group's research papers (Sion, Dewitte, et al., 2019). We believe that this kind of improvement requires more development work including finding its way into an updated version of the official tutorial, which at the time of this writing dates back to 2015. Nonetheless, we expect that it will help to better explore the symbiotic relationship between technical and legal privacy risks for the DPIAs. In terms of strengths, LINDDUN enables the identification of privacy risks related to de-identification, inference and disclosure, thanks to its extensive elicitation of data flow diagrams and threat modelling, thereby directly addressing PTPs (2), (3) and (6). Similarly, PTPs (4) and (5) are also addressed by examining data detectability and non-compliance. Something that LINDDUN has done quite well – and which is not found in other PIAs is the steering of threat mitigation towards solutions based on privacy-enhancing technologies. However, we cannot say the same for PTPs (7) and (8). Although LINDDUN uses visual representations of information flows, it is less intuitive and perhaps cumbersome for non-privacy stakeholders. Moreover, the steps related to documentation are not well developed (Wuyts et al., 2020). Finally, we could not find anything directly relevant to addressing societal and ethical concerns; hence, we consider PTP (9) to be not covered.

4.2.4.7. IRL data protection commission The information about the PIA methodology of the IRL Data Protection Commission is almost identical to that on the EU DPIA. It provides high-level coverage of PTPs (1) - (5), and (7) but does not particularly address the treatment of processing operations in the Big Data Analytics context. Hence, we classified its coverage of these PTPs as not sufficient. We found very little information about the remaining PTPs - (6), (8) and (9) – despite that they are largely privacy-related.

4.2.4.8. HK PCPD The PIA methodology of the HK PCPD is quite abstract, with the entire process contained in a document of just four pages. The impact of big data including ethical matters are discussed in articles scattered on the organisation's website, but the concerns or findings expressed are not reflected in the PIA methodology.

5. Discussion and future work

The primary purpose of this literature review was to provide 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. While imposing a DPIA, the GDPR does not reference a concrete data protection risk assessment method beyond describing the minimal content of a DPIA in Article 35(7). Notably, this could cause many problems in the Big Data Analytics context, where the volume and type of processing operations are inherently larger in scale and more complex. We were therefore interested in finding out whether we could learn from PIA methodologies how to conduct a DPIA in an environment with Big Data Analytics applications.

Given the multidisciplinary nature of our research topic, we searched ten online databases, including general academic ones (e.g. Web of Science, Scopus), and some more engineering-orientated (e.g. IEEE, Xplore, ACM) or legal science focused (e.g. International Data Privacy Law), to retrieve relevant publications. We discovered papers discussing established PIA methodologies aiming to enable organisations to conduct generic privacy or data protection risk assessments as well as privacy threat analysis frameworks maintained by research institutes, with a strong focus on the analysis and mitigation of data risks based on privacy engineering techniques and principles. These frameworks were often proposed as extensions of existing PIA methodologies.

Our research findings show that no single PIA methodology appears capable of addressing all Big Data Analytics-specific privacy and data protection risks. To demonstrate this, we identified nine Privacy Touch Points (PTPs), as categories of Big Data Analytics-specific privacy and data protection risks. Next, we analysed to what extent these PTPs are covered by the PIA methodologies discussed in our publication sample. Here we notice 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, including the ISO/IEC 29,134:2017 standard, barely cover any specific privacy or data protection risks that are specific to the Big Data Analytics context.

Our study not only allowed to reveal existing PIA methodologies' weaknesses and strengths with respect to privacy and data protection impact assessment in Big Data Analytics environments, but also to obtain a broad understanding of how to better address Big Data Analytics-specific privacy and data protection risks in DPIAs imposed by the GDPR. To go forward, one option would be to devise a best-of-breed methodology for conducting a DPIA for big data by combining the best-evaluated parts of each assessed methodology. However, we believe this approach is overly simplistic, given that PIAs are viewed by many authors as more of a compliance check than a process. Moreover, combining parts of different PIA methodologies does still not guarantee effective coverage of all risks. Our analysis of the nine PTPs revealed weaknesses in virtually all existing methodologies (e.g. PTP (9) is not sufficiently covered by any of the reviewed methodologies).

An alternative and preferred path would be 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 we defined in this paper as Privacy Touch Points and how these risks were addressed in PIA methodologies that cover them well. Further, to confirm our findings based on the reviewed literature and to fill in missing gaps, we are planning a study that uses the Delphi technique (Cantrill et al., 1996) by designing questionnaires in a similar manner to (Alnemr et al., 2015) and sending them to a panel composed of both privacy and big data experts. By consulting and confronting the opinions of both types of experts, we intend to confirm, and if needed modify, our PTPs and suggest specific guidance to make the GDPR's DPIA more fit for the Big Data Analytics context.

6. CONCLUSION

Big data, and specifically Big Data Analytics, have improved efficiency and created significant opportunities for today's information-intensive services. However, their benefits are restricted by the unintentionally heightened risk they carry of breaking the law, as they are relatively easy to compromise due to the vast amounts of personal and sometimes sensitive data that they process.

Based on our analysis of 159 systematically reviewed articles, we conclude that, 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. As our research interest is the GDPR, we are naturally keen to further develop methodological guidance for conducting DPIAs as required by the GDPR and make this assessment methodology useful for organisations and users with limited knowledge of privacy and data protection. As this is not something to be accomplished hastily by simply compiling the best-evaluated parts of each reviewed PIA methodology, we will employ the Delphi technique as a decision-making and consensus-seeking technique to make an informed choice about the most relevant risk assessment criteria in the PIA methodology sample. These criteria will be selected by analysing the responses collected from scholars and experts in the fields of privacy and big data. Our goal for the next phase of our research is to develop a more comprehensive DPIA methodology wherein we 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. The current systematic literature review is a crucial step in building the knowledge base required for achieving our research goal.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

REFERENCES

Agarwal S. Developing a structured metric to measure privacy risk in privacy impact assessments. IFIP Advances in Information and Communication Technology; 2016 Vol. 476.

Ahmadian AS, Strüber D, Riediger V, Jürjens J. Supporting privacy impact assessment by model-based privacy analysis. Proceedings of the 33rd Annual ACM Symposium on Applied Computing - SAC '18; 2018. p. 1467–74.

Akma N, Rafhi M, Rahman SA, Rafhi NAM, Rahman SA. Factors of big data analytics in enabling the knowledge management practice. Int J Acad Res Bus Soc Sci 2018;7(11):917–27. doi:10.6007/ijarbss/v7-i11/3526.

Akoka J, Comyn-Wattiau I, Laoufi N. Research on Big Data – A systematic mapping study. Comput Standards Interfaces 2017;54:105–15. doi:10.1016/j.csi.2017.01.004.

Al-Fedaghi S. Engineering privacy revisited. J Comput Sci 2012;8(1):107–20. doi:10.3844/jcssp.2012.107.120.

Al-Momani A, Kargl F, Schmidt R, Kung A, Bosch C. A privacy-aware V-model for software development. 2019 IEEE Secur Privacy Workshops (SPW) 2019:100–4. doi:10.1109/SPW.2019.00028.

Alali FA, Yeh C-L. Cloud computing: overview and risk analysis. J Informat Syst 2012;26(2):13–33 internal-pdf://185.191.121.39/Alali-2012-Cloud Computing_ Overview and Risk.pdf.

Alnemr R, Cayirci E, Corte LD, Garaga A, Leenes R, Mhungu R, et al. A data protection impact assessment methodology for cloud. Annual Privacy Forum. Springer; 2015. p. 60–92.

Alshehri M, Drew S. E-government fundamentals. *Proceedings of the IADIS International Conference ICT, Society and Human Beings 2010, Part of the IADIS Multi Conference on Computer Science and Information Systems 2010, MCCSIS 2010, 2001; 2010.* p. 35–42.

Altman M, Wood A, O'Brien DR, Gasser U. Practical approaches to big data privacy over time. Int Data Privacy Law 2018;8(1):29–51. doi:10.1093/idpl/ipx027.

Anindya IC, Roy H, Kantarcioglu M, Malin B. Building a dossier on the cheap: integrating distributed personal data resources under cost constraints. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Part F1318*; 2017. p. 1549–58.

Ardagna CA, Ceravolo P, Damiani E. Big data analytics as-a-service: issues and challenges. 2016 IEEE International Conference on Big Data (Big Data); 2016. p. 3638–44.

Bages-Amat A, Spillecke D, Stanley J. These eight charts show how COVID-19 has changed B2B sales forever. McKinsey & Company 2020 https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/these-eight-charts-show-how-covid-19-has-changed-b2b-sales-forever.

Barocas S, Nissenbaum H. Big data's end run around procedural privacy protections. Commun ACM 2014a;57(11):31–3. doi:10.1145/2668897.

Barocas S, Nissenbaum H. Big data's end run around anonymity and consent. Privacy, Big Data, and the Public Good: Frameworks for Engagement. Cambridge University Press; 2014b. p. 44–75.

Barocas S, Selbst A. Big data's disparate impact. Calif Law Rev 2016;104(3):671. doi:10.15779/Z38BG31.

Bas Seyyar M, Geradts ZJMH. Privacy impact assessment in large-scale digital forensic investigations. Forensic Sci Int 2020;33. doi:10.1016/j.fsidi.2020.200906.

BBC. Equifax Says Almost 400,000 Britons hit in Data breach, 15 September 2017. News: Tech; 2017 https://www.bbc.com/news/technology-41286638.

Bennett, C.J., & Bayley, R.M. (2015). Privacy protection in the era of " big data": response to office of privacy commissioner's discussion paper on " consent and privacy." In *Exploring the Boundaries of Big Data*.

Bertino E. Big data-security and privacy. Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015; 2015. p. 757–61.

Bertot JC, Choi H. Big data and e-government. Proceedings of the 14th Annual International Conference on Digital Government Research; 2013. p. 1–10.

Bieker F, Friedewald M, Hansen M, Obersteller H, Rost M. A process for data protection impact assessment under the European general data protection regulation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); 2016. p. 21–37.

Bieker F, Martin N, Friedewald M, Hansen M. Data protection impact assessment: a hands-on tour of the GDPR's most practical tool. IFIP Adv Inf Commun Technol 2018;526:207–20. doi:10.1007/978-3-319-92925-5_13.
Binns R. Data protection impact assessments: a meta-regulatory approach. Int Data Privacy Law 2017;7(1):22–35. doi:10.1093/idpl/ipw027.
Birnhack M. A process-based approach to informational privacy and the case of big medical data. Theoretical Inquiries Law 2019;20(1):257–90. doi:10.1515/til-2019-0009.
Bisztray T, Gruschka N. Privacy Impact Assessment: comparing Methodologies with a Focus on Practicality. 2019 Nordic Conference on Secure IT Systems. Springer International Publishing; 2019. p. 3–19.
Bracy J. The Equifax breach, response, and fallout. Iapp.Com 2017 https://iapp.org/news/a/equifax-data-breach-affects-143-million-consumers/.
Bradford L, Aboy M, Liddell K. COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes. J Law Biosci 2020;7(1). doi:10.1093/jlb/lsaa034.
Bu-Pasha S. The controller's role in determining 'high risk' and data protection impact assessment (DPIA) in developing digital smart city. Inf Commun Technol Law 2020;29(3):391–402. doi:10.1080/13600834.2020.1790092.
Buhr C-C, Kleiner T. European open data policy: challenges and opportunities. Zeitschrift Für Politikberatung 2012;5(3):141–6. doi:10.5771/1865-4789-2012-3-141.
Butterworth M. The ICO and artificial intelligence: the role of fairness in the GDPR framework. Comput Law Secur Rev 2018;34(2):257–68. doi:10.1016/j.clsr.2018.01.004.
Cantrill JA, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. Int J Pharmacy Practice 1996;4(2):67–74. doi:10.1111/j.2042-7174.1996.tb00844.x.
Cavoukian A, others. Privacy by design: the 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada; 2009 https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf.
Chen H, Yan Z. Security and privacy in big data lifetime: a review. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10067 LNCS*; 2016. p. 3–15.
Chen J, Chen Y, Du X, Li C, Lu J, Zhao S, et al. Big data challenge: a data management perspective. Front Comput Sci 2013;7(2):157–64. doi:10.1007/s11704-013-3903-7.
Clarke, R. (1999). *Introduction to dataveillance and information privacy, and definitions of terms*. Roger Clarke's Dataveillance and Information Privacy …. http://www.cse.unsw.edu.au/~cs4920/resources/Roger-Clarke-Intro.pdf
Clarke R. Privacy impact assessment : its origins and development. Comput Law Secur Rev 2009;25(2):123–35. doi:10.1016/j.clsr.2009.02.002.
Clarke R. An evaluation of privacy impact assessment guidance documents. Int Data Privacy Law 2011;1(2):111–20. doi:10.1093/idpl/ipr002.
Clarke R. Privacy and Free Speech; 2014 http://www.rogerclarke.com/DV/PFS-1408.html.
Clarke R. Privacy impact assessments as a control mechanism for Australian counter-terrorism initiatives. Comput Law Secur Rev 2016;32(3):403–18. doi:10.1016/j.clsr.2016.01.009.
Clarke R. The Distinction between a PIA and a Data Protection Impact Assessment (DPIA) under the EU GDPR. Roger Clarke's Web-Site; 2017 http://www.rogerclarke.com/DV/PIAvsDPIA.html.
CNIL. (2017). *How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence*. https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf
Coles J, Faily S, Ki-Aries D. Tool-supporting data protection impact assessments with CAIRIS. Proceedings - 2018 5th International Workshop on Evolving Security and Privacy Requirements Engineering, ESPRE 2018; 2018. p. 21–7.
Council of Europe. Guide to *Human Rights* for Internet *Users*; 2014 https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016804d5b31.
Crawford K, Schultz J. Big data and due process: toward a framework to redress predictive privacy harms. Boston College Law Rev 2014;55(1):93. doi:10.1525/sp.2007.54.1.23.
Crockett K, Goltz S, Garratt M. GDPR impact on computational intelligence research. 2018 International Joint Conference on Neural Networks (IJCNN); 2018. p. 1–7.
CSA. The Treacherous 12: Top threats to Cloud Computing. Top Threats Working Group; 2017 https://cloudsecurityalliance.org/working-groups/top-threats.
Custers B, Dechesne F, Sears AM, Tani T, van der Hof S. A comparison of data protection legislation and policies across the EU. Comput Law Secur Rev 2018;34(2):234–43. doi:10.1016/j.clsr.2017.09.001.
De Hert, P., Kloza, D., & Wright, D. (2012). *PIAF Project Deliverable 3: recommendations for a privacy impact assessment framework for the European Union*. http://piafproject.eu/ref/PIAF_D3_final.pdf
De SJ, Le Métayer D. A refinement approach for the reuse of privacy risk analysis results. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10518 LNCS; 2017. p. 52–83.
Deloitte. Privacy and Data Protection in the age of COVID-19. Deloitte; 2020 https://www2.deloitte.com/be/en/pages/risk/articles/privacy-and-data-protection-in-the-age-of-covid-19.html.
Deng M, Wuyts K, Scandariato R, Preneel B, Joosen W. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requirements Eng 2011;16(1):3–32. doi:10.1007/s00766-010-0115-7.
DHS. (2007). *Verification Information System Supporting Verification Programs* (Issue 571).
DHS. (2011). *Systematic Alien Verification for Entitlements (SAVE) Program* (Issue 202).
DHS. (2014). *Privacy Impact Assessment for the DHS Data Framework*.
DHS. (2020). Privacy impact assessment for the Data Analytics Technology Center. In *DHS*.
Di Iorio CT, Carinci F, Oderkirk J, Smith D, Siano M, De Marco DA, et al. Assessing data protection and governance in health information systems: a novel methodology of Privacy and Ethics Impact and Performance Assessment (PEIPA). J Med Ethics 2020:1–8. doi:10.1136/medethics-2019-105948.
Drewer D, Miladinova V. The BIG DATA Challenge: impact and opportunity of large quantities of information under the Europol Regulation. Comput Law Secur Rev 2017;33(3):298–308. doi:10.1016/j.clsr.2017.03.006.
Easton C. Analysing the role of privacy impact assessments in technological development for crisis management. J Contingen Crisis Manag 2017;25(1):7–14. doi:10.1111/1468-5973.12140.
EC. (2019). *High-level expert group on artificial intelligence: ethics guidelines for trustworthly AI*. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
EDPS. (2015). *Opinion 4 /2015: towards a new digital ethics - Data, dignity and technology*. https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_en.pdf
EDPS. Guidelines on personal data breach notification For the European Union Institutions and Bodies. EDPS; 2018a https://edps.europa.eu/data-protection/our-work/publications/guidelines/guidelines-personal-data-breach-notification_en.
EDPS. Opinion 5/2018: preliminary Opinion on privacy by design. EDPS; 2018b

https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf.
EDPS. (2019). *Decision of the European Data Protection Supervisor of 16 July 2019 on DPIA Lists Issued Under Articles 39(4) and (5) of Regulation (Eu) 2018/1725*. https://edps.europa.eu/data-protection/our-work/publications/guidelines/data-protection-impact-assessment-list_en
Edwards L, McAuley D, Diver L. From privacy impact assessment to social impact assessment. 2016 IEEE Symposium on Security and Privacy Workshops, SPW 2016; 2016. p. 53–7.
EESC. The ethics of big data: balancing economic benefits and ethical questions of big data in the eu policy context. EESC; 2016 https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data.
El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of *Re*-identification attacks on health data. PLoS ONE 2011(12):6. doi:10.1371/journal.pone.0028071.
ENISA. Privacy and Data Protection by Design - from policy to engineering. ENISA; 2014 https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design.
ENISA. Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. Enisa; 2015.
EP. Regulation (EU) 2016/679 of the European Parliament and of the coucil of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC. Official Journal of the European Union; 2016 https://doi.org/http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_285/l_28520031101en00330037.pdf.
Favaretto M, De Clercq E, Elger BS. Big Data and discrimination: perils, promises and solutions. A systematic review. J Big Data 2019;6(1):12. doi:10.1186/s40537-019-0177-4.
Ferra F, Wagner I, Boiten E, Hadlington L, Psychoula I, Snape R. Challenges in assessing privacy impact: tales from the front lines. Secur Privacy 2020;3(2):1–19. doi:10.1002/spy2.101.
Finn RL, Wright D, Friedewald M. Seven types of privacy. European Data Protection: Coming of Age. Springer Netherlands; 2013. p. 3–32.
Floridi L. The ontological interpretation of informational privacy. Ethics Inf Technol 2006;7(2005):185–200. doi:10.1007/s10676-006-0001-7.
Friedewald M, Schiffner S, Serna J, Ikonomou D, Rannenberg K, Bieker F, et al. Springer; 2016. p. 21–37. http://link.springer.com/10.1007/978-3-319-44760-5.
Frier, S. (2018). *Facebook CEO Zuckerberg Says Problems Will Take Years to Fix*. Bloomberg.Com. https://www.bloomberg.com/news/articles/2018-12-28/facebook-ceo-zuckerberg-says-problems-will-take-years-to-fix
Gantz, J., & Reinsel, D. (2011). *EMC corporation: extracting value from Chaos*.
Garcia AC, McDonnell NN, Troncoso C, Le Metayer D, Kroener I, Wright D, et al. PRIPARE Privacy-and Security-by-Design Methodology Handbook; 2015.
Gbadeyan A, Butakov S, Aghili S. IT governance and risk mitigation approach for private cloud adoption: case study of provincial healthcare provider. Ann Telecommun 2017;72(5–6):347–57. doi:10.1007/s12243-017-0568-5.
Gellert Raphael. European Union • the article 29 working party's provisional guidelines on data protection impact assessment. Eur Data Protect Law Rev 2017;3(2):212–17. doi:10.21552/edpl/2017/2/11.
Gellert Raphaël. Understanding the notion of risk in the general data protection regulation. Comput Law Secur Rev 2018;34(2):279–88. doi:10.1016/j.clsr.2017.12.003.
German Federal and State Data Protection Commissioners. (2016). *The Standard Data Protection Model: a concept for inspection and consultation on the basis of unified protection goals*.
Gloria González Fuster AS. Big data and smart devices and their impact on privacy. J Chem Inf Model 1989;53 https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU%282015%29536455.
Gonçalves ME. The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward. Inf Commun Technol Law 2017;26(2):90–115. doi:10.1080/13600834.2017.1295838.
Gostin LO. Genetic privacy. J Law, *Med Ethics* 1995;23(4):320–30. doi:10.1111/j.1748-720X.1995.tb01374.x.
Gruschka N, Mavroeidis V, Vishi K, Jensen M. Privacy issues and data protection in big data: a case study analysis under GDPR. IEEE International Conference on Big Data; 2019. p. 5027–33.
Guggenheim. (2016). *Technological Innovation Portfolio, Series 11*. ps://www.guggenheiminvestments.com/uit/trust/atec011
Hansen M, Jensen M, Rost M. Protection goals for privacy engineering. Proceedings - 2015 IEEE Security and Privacy Workshops, SPW 2015; 2015. p. 159–66.
Himmel J, Siebler N, Laegeler F, Grupe M, Langweg H. Privacy points as a method to support privacy impact assessments. 2015 IEEE/ACM 1st International Workshop on TEchnical and LEgal Aspects of Data PRivacy and SEcurity; 2015. p. 50–3.
HIQA. Guidance on Privacy Impact Assessment in Health and Social Care. HIQA; 2017 (Issue 2.0) http://www.hiqa.ie/ .
ICO. (2013). *Privacy Impact Assessment executive summary*. https://ico.org.uk/media/1042837/trilateral-report-executive-summary.pdf
ICO. Big data, artificial intelligence, machine learning and data protection. ICO; 2017 https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.
ICO. (2019). *Guide to the General Data Protection Regulation (GDPR)*. https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/
IDC. (2019). *Worldwide Public Cloud Services Spending Forecast to Reach $160 Billion This Year, According to IDC*. International Data Corporation. https://www.businesswire.com/news/home/20190228005137/en/Worldwide-Public-Cloud-Services-Spending-Forecast-to-Reach-160-Billion-This-Year-According-to-IDC
Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020;26(4):463–4. doi:10.1038/s41591-020-0832-5.
ISO. ISO/IEC 29134:2017 Guidelines for privacy impact assessment. ISO; 2017 https://www.iso.org/standard/62289.html.
ITRC. (2020). *Data Breach Report 2020*. https://notified.idtheftcenter.org/s/
Jalali S, Wohlin C. Systematic literature studies : database searches vs. backward snowballing. ESEM'12: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement; 2012. p. 29–38.
Jeffrey S. A new digital dark age? Collaborative web tools, social media and long-term preservation. World Archaeol 2012;44(4):553–70. doi:10.1080/00438243.2012.737579.
Johnstone MN. Threat modelling with STRIDE and UML. *Proceedings of the 8th Australian Information Security Management Conference, November*; 2010. p. 18–27.
Joyee De S, Le Métayer D. PRIAM: a privacy risk analysis methodology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9963 LNCS*; 2016. p. 221–9.
Jugulum R. Importance of data quality for analytics. In: Sampaio P, Saraiva P, editors. Quality in the 21st Century. Springer International Publishing; 2016. p. 23–31.
Kaisler S, Armour F, Espinosa JA, Money W. Big data: issues and challenges moving forward. 2013 46th Hawaii International Conference on System Sciences; 2013. p. 995–1004.

Katal A, Wazid M, Goudar RH. Big data : issues, challenges, tools and good practices. International Conference on System Sciences (HICSS); 2013. p. 1–27.
Kitchenham B. Guidelines for performing systematic literature reviews in software engineering. Software Eng Group School Comput Sci Math 2007;65. doi:10.1145/1134285.1134500.
Klievink B, Cunningham S, Romijn BJ, Cunningham S, de Bruijn H. Big data in the public sector : uncertainties and readiness. Inf Syst Front 2017;19(2):267–83. doi:10.1007/s10796-016-9686-2.
Kloza D, van Dijk N, Casiraghi S, Vazquez Maymir S, Roda S, Tanas A, et al. Data protection impact assessments in the European Union: designing an appraisal method towards a more robust protection of individuals. D.Pia.Lab Policy Brief, VUB 2018;2:4.
Kokott J, Sobotta C. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. Int Data Privacy Law 2013;3(4). doi:10.1093/idpl/ipt017.
Kosseim, M. (2016). *Speech: my Data Made Me Do It: ethical Considerations of Big Data*. Office of the Privacy Commissioner of Canada. https://www.priv.gc.ca/en/opc-news/speeches/2016/sp-d_20160930_pk/
Krasnow Waterman K, Bruening PJ. Big Data analytics: risks and responsibilities. Int Data Privacy Law 2014;4(2):89–95. doi:10.1093/idpl/ipu002.
Kumar A. Distributed and big data storage management in grid computing. Int J Grid Comput Appl 2012;3(2):19–28. doi:10.5121/ijgca.2012.3203.
Labrinidis A, Jagadish HV. Challenges and opportunities with big data. Proceedings of the VLDB Endowment; 2012. p. 2032–3.
Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009;62(10):e1–e34. doi:10.1016/j.jclinepi.2009.06.006.
Libération. (2021). *Les informations confidentielles de 500 000 patients français dérobées à des laboratoires et diffusées en ligne*. Libération. https://www.liberation.fr/checknews/les-informations-confidentielles-de-500-000-patients-francais-derobees-a-des-laboratoires-medicaux-et-diffusees-en-ligne-20210223_VO6W6J6IUVATZD4VOVNDLTDZBU/
Liu Q, Srinivasan A, Hu J, Wang G. Preface: security and privacy in big data clouds. Future Generat Comput Syst 2017;72:206–7. doi:10.1016/j.future.2017.03.033.
MacCarthy M. Standards of fairness for disparate impact assessment of big data algorithms. SSRN Electron J 2018;48:67–148. doi:10.2139/ssrn.3154788.
Mai J-E. Big data privacy: the datafication of personal information. Inf Soc 2016;32(3):192–9. doi:10.1080/01972243.2016.1153010.
Mantelero A. The future of consumer data protection in the E.U. *Re*-thinking the "notice and consent" paradigm in the new era of predictive analytics. Comput Law Secur Rev 2014;30(6):643–60. doi:10.1016/j.clsr.2014.09.004.
Mantelero A. Personal data for decisional purposes in the age of analytics: from an individual to a collective dimension of data protection. Comput Law Secur Rev 2016;32(2):238–55. doi:10.1016/j.clsr.2016.01.014.
Mantelero A. AI and Big Data: a blueprint for a human rights, social and ethical impact assessment. Comput Law Secur Rev 2018;34(4):754–72. doi:10.1016/j.clsr.2018.05.017.
Mantelero A, Vaciago G. The "dark side" of big data: private and public interaction in social surveillance. Comput Law Rev Int 2013;14(6):161–9. doi:10.9785/ovs-cri-2013-161.
McKinsey&Company. Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute; 2011 https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation.
McMahon A, Buyx A, Prainsack B. Big data governance needs more collective responsibility: the role of harm mitigation in the governance of data use in medicine and beyond. Med Law Rev 2020;28(1):155–82. doi:10.1093/medlaw/fwz016.
Meis R, Heisel M. Systematic identification of information flows from requirements to support privacy impact assessments. 10th International Conference on Software Paradigm Trends, Proceedings; Part of 10th International Joint Conference on Software Technologies, ICSOFT 2015; 2015. p. 43–52.
Meis R, Heisel M. Supporting privacy impact assessments using problem-based privacy analysis. International Conference on Software Technologies, ICSOFT 2015; 2016. p. 79–98.
Mourby M, Mackey E, Elliot M, Gowans H, Wallace SE, Bell J, et al. Are pseudonymised data always personal data? Implications of the GDPR for administrative data research in the UK. Comput Law Secur Rev 2018;34(2):222–33. doi:10.1016/j.clsr.2018.01.002.
Munir AB, Mohd Yasin SH, Muhammad-Sukki F. Big data : big challenges to privacy and data protection. WASET Int J Soc, Educ, Econ Manag Eng 2015;9(1):355–63 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609229.
Nadimpalli SV, Kumari VV. Detecting dependencies in an anonymized dataset. Proceedings of the International Conference on Advances in Computing, Communications and Informatics - ICACCI '12, 2012.
Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. Proceedings - IEEE Symposium on Security and Privacy; 2008. p. 111–25.
Nelson B, Olovsson T. Security and privacy for big data: a systematic literature review. Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016; 2016. p. 3693–702.
Notario N, Crespo A, Martin YS, Del Alamo JM, Metayer DLe, Antignac T, et al. PRIPARE: integrating privacy best practices into a privacy engineering methodology. Proceedings - 2015 IEEE Security and Privacy Workshops, SPW 2015; 2015. p. 151–8.
NZPC. (2021). *Office of the Privacy Commissioner position on the regulation of biometrics*. https://www.privacy.org.nz/publications/guidance-resources/biometrics-and-privacy/
OAIC. (2018). Guide to Data Analytics and the Australian Privacy Principles. In *Office of the Australian Information Commissioner*. https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-data-analytics-and-the-australian-privacy-principles/
OAIC. (2019a). *Data breach preparation and response: a guide to managing data breaches in accordance with the Privacy Act 1988 (CTH)*. https://www.oaic.gov.au/resources/agencies-and-organisations/guides/data-breach-preparation-and-response.pdf
OAIC. (2019b). *Privacy Impact Assessment: consumer Data Right*. https://www.oaic.gov.au
Oetzel MC, Spiekermann S. A systematic methodology for privacy impact assessments: a design science approach. Eur J Inf Sys 2014;23(2):126–50. doi:10.1057/ejis.2013.18.
Oetzel, M.C., Spiekermann, S., Grüning, I., Kelter, H., & Mull, S. (2011). *Privacy Impact Assessment Guideline for RFID Applications*. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/ElekAusweise/PIA/Privacy_Impact_Assessment_Guideline_Langfassung.pdf?__blob=publicationFile
Oetzel M, Spiekermann S. Privacy-by-design through systematic privacy impact assessment: a design science approach. Eur J Inf Syst 2017;2:128–50 http://epub.wu.ac.at/5495/1/EJIS_PIA_vs9_final_circ.pdf.
Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev 2010;57(6):1701–77.
OPC_NZ. (2007). *Privacy Impact Assessment Handbook*. https://www.privacy.org.nz/

OPC-CA. Privacy Impact Assessments: Frequently asked Questions. Office of the Privacy Commissioner of Canada; 2019 https://priv.gc.ca/en/privacy-topics/privacy-impact-assessments/.
OPCL-US. (2012). Privacy Impact Assessments: official Guidance. In *Office of Privacy and Civil Liberties - United States Department of Justice (Revised*. https://www.state.gov/
Otjacques B, Hitzelberger P, Feltz F. Interoperability of E-government information systems: issues of identification and data sharing. J Manag Inf Syst 2007;23(4):29–51. doi:10.2753/MIS0742-1222230403.
PCPD. (2019). *Data Ethics for Small and Medium Enteprises*. https://www.pcpd.org.hk/english/resources_centre/publications/information_leaflet/information_leaflet.html
Petersen K, Vakkalanka S, Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: an update. Inf Softw Technol 2015;64:1–18. doi:10.1016/j.infsof.2015.03.007.
Ponemon Institute. (2020). *Cost of a Data Breach Report 2020*. https://www.ibm.com/security/digital-assets/cost-data-breach-report
PRC. (2017). *Privacy Rights Clearinghouse: data Breaches*. Privacy Rights Clearinghouse. https://www.privacyrights.org/data-breaches
Puijenbroek JVan, Hoepman J-H. Privacy impact assessment in practice - the results of a descriptive field study in the Netherlands. IWPE 2017: International Workshop on Privacy Engineering: Proceedings of the 3rd International Workshop on Privacy Engineering; 2017. p. 1–8.
Quinn P, Quinn L. Big genetic data and its big data protection challenges. Comput Law Secur Rev 2018;34(5):1000–18. doi:10.1016/j.clsr.2018.05.028.
Raab CD. Information privacy, impact assessment, and the place of ethics. Comput Law Secur Rev 2020;37. doi:10.1016/j.clsr.2020.105404.
Raab C, Wright D. Surveillance: extending the limits of privacy impact assessment. Privacy Impact Assessment. Springer Netherlands; 2012. p. 363–83.
Ramirez, E. (2013). *The privacy challenges of Big Data: a view from the lifeguard's chair*. http://www.ftc.gov/os/caselist/1023136/111024/googlebuzzcmpt.pdf;
Rhoen M, Feng QY. Why the "computer says no": illustrating big data's discrimination risk through complex systems science. Int Data Privacy Law 2018;8(2):140–59. doi:10.1093/idpl/ipy005.
Richards NM, King JH. Three paradoxes of big data. Stanford Law Rev Online 2013 https://heinonline.org/HOL/P?h=hein.journals/slro66&i=41.
rtbf.be. (2019). *Après une annus horribilis, Facebook va devoir rassurer pour l'avenir*. https://www.rtbf.be/info/economie/detail_apres-une-annus-horribilis-facebook-va-devoir-rassurer-pour-l-avenir?id=10132038
Rubinstein I. Big data: the end of privacy or a new beginning? Int Data Privacy Law 2012:12–56. doi:10.2139/ssrn.2157659.
Sagiroglu S, Sinanc D. Big data: a review. 2013 International Conference on Collaboration Technologies and Systems (CTS); 2013. p. 42–7.
Salleh KA, Janczewski L. Technological, organizational and environmental security and privacy issues of big data: a literature review. Procedia Comput Sci 2016;100:19–28. doi:10.1016/j.procs.2016.09.119.
Sampson F. 7 rights of individuation: the need for greater protection of individual rights in big data. Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing; 2014. p. 677–80.
Schekkerman J. How to survive in the jungle of enterprise architecture frameworks: creating or choosing an enterprise architecture framework. Framework. Trafford Publishing; 2004 http://www.amazon.com/Survive-Jungle-Enterprise-Architecture-Frameworks/dp/141201607X.
Scudder N, McNevin D, Kelty SF, Walsh SJ, Robertson J. Forensic DNA phenotyping: developing a model privacy impact assessment. Forensic Sci Int Genet 2018;34:222–30. doi:10.1016/j.fsigen.2018.03.005.
Sheridan C. Massive data initiatives and AI provide testbed for pandemic forecasting. Nat Biotechnol 2020;38(9):1010–13. doi:10.1038/s41587-020-0671-4.
Shin DH, Choi MJ. Ecological views of big data: perspectives and issues. Telemat Inf 2015;32(2):311–20. doi:10.1016/j.tele.2014.09.006.
Shirer, M. (2015). *Double-digit growth forecast for the worldwide big data and business analytics market through 2020 Led by Banking and Manufacturing Investments, According to IDC*. Idc.Com. https://doi.org/10.1207/S15327051HCI16234
Sion L, Dewitte P, Van Landuyt D, Wuyts K, Emanuilov I, Valcke P, et al. An architectural view for data protection by design. *2019 IEEE International Conference on Software Architecture, ICSA 2019*, i; 2019. p. 11–20.
Sion L, Van Landuyt D, Wuyts K, Joosen W. Privacy risk assessment for data subject-aware threat modeling. Proceedings - 2019 IEEE Symposium on Security and Privacy Workshops, SPW 2019; 2019. p. 64–71.
Subashini S, Kavitha V. A survey on security issues in service delivery models of cloud computing. J Netw Comput Appl 2011;34(1):1–11. doi:10.1016/j.jnca.2010.07.006.
Sun J, Lee S. A study on the implementation of the effective privacy impact assessment management system. International Conference on Information Science and Applications, ICISA 2013; 2013. p. 1–4.
Svantesson D, Clarke R. Privacy and consumer risks in cloud computing. Comput Law Security Rev 2010;26(4):391–7. doi:10.1016/j.clsr.2010.05.005.
Svantesson DJB. Data protection in cloud computing – the Swedish perspective. Computer Law & Security Review 2012;28(4):476–80. doi:10.1016/j.clsr.2012.05.005.
Sweeney L. k-anonymity: a model for protecting privacy. Int J Uncertain, Fuzz Knowledge-Based Syst 2002;10(05):557–70.
Tancock D, Pearson S, Charlesworth A. Analysis of privacy impact assessments within major jurisdictions. PST 2010: 2010 8th International Conference on Privacy, Security and Trust; 2010a. p. 118–25.
Tancock, D., Pearson, S., & Charlesworth, A. (2010b). *The emergence of privacy impact assessments*.
Tancock D, Pearson S, Charlesworth A. A privacy impact assessment tool for cloud computing. Privacy and Security For Cloud computing. Springer; 2013. p. 73–123.
Tene O, Polonetsky J. Judged by the Tin Man: individual Rights in the Age of Big Data. Journal on Telecommunications & High Technology Law 2013;2:351–68 https://heinonline.org/HOL/P?h=hein.journals/jtelhtel11&i=379.
Theoharidou M, Papanikolaou N, Pearson S, Gritzalis D. Privacy risk, security, accountability in the cloud. Proceedings of the International Conference on Cloud Computing Technology and Science, CloudCom; 2013. p. 177–84.
Thorlund K, Dron L, Park J, Hsu G, Forrest JI, Mills EJ. A real-time dashboard of clinical trials for COVID-19. The Lancet Digital Health 2020;2(6). doi:10.1016/S2589-7500(20)30086-8.
Todde M, Beltrame M, Marceglia S, Spagno C. Methodology and workflow to perform the data protection impact assessment in healthcare information systems. Inf Medicine Unlocked 2020;19. doi:10.1016/j.imu.2020.100361.
Tranfield D, Denyer D, Smart P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manag 2003;14(3):207–22. doi:10.1111/1467-8551.00375.
Turner M, Linkman S, Pretorius R, Budgen D, Niazi M, Pearl

Brereton O, et al. Systematic literature reviews in software engineering – a tertiary study. Inf Softw Technol 2010;52(8):792–805. doi:10.1016/j.infsof.2010.03.006.
van Dijk N, Gellert R, Rommetveit K. A risk to a right? Beyond data protection risk assessments. Comput Law Secur Rev 2016;32(2):286–306. doi:10.1016/j.clsr.2015.12.017.
Vesset, D., Olofson, C.W., & Fleming, M. (2018). *Worldwide big data and analytics software forecast, 2018 –2022* (Issue September).
Wadhwa K. Privacy impact assessment reports: a report card. Info 2012;14(3):35–47. doi:10.1108/14636691211223210.
Wadhwa K, Rodrigues R. Evaluating privacy impact assessments. Innovation 2013;26(1–2):161–80. doi:10.1080/13511610.2013.761748.
Warren A, Bayley R, Bennett C, Charlesworth A, Clarke R, Oppenheim C. Privacy impact assessments: international experience as a basis for UK Guidance. Comput Law Secur Report 2008;24(3):233–42. doi:10.1016/j.clsr.2008.03.003.
Wei Y-C, Wu W-C, Lai G-H, Chu Y-C. pISRA: privacy considered information security risk assessment model. J Supercomput 2020;76(3):1468–81. doi:10.1007/s11227-018-2371-0.
White House. (2016). *Big Data: a Report on Algorithmic Systems, Opportunity, and Civil Rights*.
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
Wright D. A framework for the ethical impact assessment of information technology. Ethics Inf Technol 2011a;13(3):199–226. doi:10.1007/s10676-010-9242-6.
Wright D. Should privacy impact assessments be mandatory? Commun ACM 2011b;54(8):121. doi:10.1145/1978542.1978568.
Wright D. The state of the art in privacy impact assessment. Comput Law Secur Rev 2011c;28(1):54–61. doi:10.1016/j.clsr.2011.11.007.
Wright D. Making privacy impact assessment more effective. Inf Soc 2013;29(5):307–15. doi:10.1080/01972243.2013.825687.
Wright D. How Good are PIA Reports – and where are they? Eur Bus Law Rev 2014;25(3):407–26.
Wright D, De Hert P. Privacy impact assessment. Privacy Impact Assessment In D. Wright & P. De Hert (Eds.). Springer Netherlands; 2012.
Wright D, De Hert P. Enforcing Privacy: Regulatory, Legal and Technological Approaches, 25. Springer International Publishing; 2016 D. Wright & P. De Hert (eds.).
Wright D, Finn R, Rodrigues R. A comparative analysis of privacy impact assessment in six countries. J Contemporary Eur Res 2013;9(1):160–80.
Wright D, Friedewald M. Integrating privacy and ethical impact assessments. Sci Public Policy 2013;40(6):755–66. doi:10.1093/scipol/sct083.
Wright D, Friedewald M, Gutwirth S, Langheinrich M, Mordini E, Bellanova R, et al. Sorting out smart surveillance. Comput Law Security Rev 2010;26(4):343–54. doi:10.1016/J.CLSR.2010.05.007.
Wright D, Gellert R, Gutwirth S, Friedewald M. Minimizing technology risks with PIAs, precaution, and participation. IEEE Technol Soc Mag 2011;30(4):47–54. doi:10.1109/MTS.2011.943460.
Wright D, Raab C. Privacy principles, risks and harms. Int Rev Law 2014;28(3):277–98 *Computers & Technology*. doi:10.1080/13600869.2014.913874.
Wright D, Raab CD. Constructing a surveillance impact assessment. Comput Law Secur Rev 2012;28(6):613–26. doi:10.1016/j.clsr.2012.09.003.
Wright D, Wadhwa K, Lagazio M, Raab C, Charikane E. Integrating privacy impact assessment in risk management. Int Data Privacy Law 2014;4(2):155–70. doi:10.1093/idpl/ipu001.
Wuyts K, Joosen W. LINDDUN Privacy Threat Modeling: a Tutorial. Report CW 685; 2015 https://lirias.kuleuven.be/.
Wuyts K, Sion L, Joosen W. LINDDUN GO: a lightweight approach to privacy threat modeling. Proceedings - 5th IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2020; 2020. p. 302–9.
Yordanov A. Nature and ideal steps of the data protection impact assessment under the general data protection regulation. Eur Data Prot Law Rev 2017;3(4):486–95. doi:10.21552/edpl/2017/4/10.
Zarsky' TZ, 995; 2017 https://heinonline.org/HOL/License.
Zwitter A, Gstrein OJ. Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. J Int Humanitarian Action 2020;5(1):4. doi:10.1186/s41018-020-00072-6.