Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ORIGINAL ARTICLE

Challenges in Assessing Privacy Impact: Tales from the Front Lines

Fenia Ferra[1] | Isabel Wagner[1] | Eerke Boiten[1] | Lee Hadlington[1,3] | Ismini Psychoula[1] | Richard Snape[2]

[1]Cyber Technology Institute, De Montfort University, UK
[2]Institute of Energy and Sustainable Development, De Montfort University, UK
[3]Division of Psychology, De Montfort University, UK

Correspondence
Isabel Wagner Email: isabel.wagner@dmu.ac.uk

Summary

Privacy impact assessments (PIAs) aim to identify, rank, and mitigate privacy risks. Even though PIAs are legally mandated in some cases and privacy professionals perform PIAs on a daily basis, it is an open problem how privacy risks can be measured systematically. Research on privacy risk measurement often does not take into account the practical needs and requirements for PIAs in real organizations. In this paper, we fill this gap by reporting on focus groups we held with a diverse group of privacy professionals. Through thematic analysis, we identify three themes that emerged from the focus groups: (1) how privacy in the contemporary society affects privacy risk assessment; (2) current practices and procedures in privacy risk assessment; and (3) common issues and challenges. Based on these themes, we identify future research directions for privacy risk measurement. Our paper can help to ground research on privacy risk measurement in practical challenges faced by privacy professionals.

KEYWORDS:
privacy impact, PIA, risk assessment, privacy metrics, privacy harm, focus groups, thematic analysis

1 | INTRODUCTION

For new computer based systems, particularly those which collect and process personal data, privacy impact assessment (PIA) is an essential process. Doing this early and continuously is a central idea within Privacy by Design (PbD)[1], and it is mandated in "risky" processing contexts in modern privacy legislation such as the GDPR (General Data Protection Regulation)[2].

Appropriately measuring such privacy risks is a prerequisite for managing them systematically[3]. Doing so allows prioritisation, aggregation, and triage of risks, and helps in justifying the selection of mitigating controls. Research into methods for privacy risk measurement needs to take into account the baseline of privacy professionals' activities in this sphere: data protection consultants and Data Protection Officers (e.g. in the GDPR sense) undertake privacy risk assessment in practice all the time. Professional practice can give grounding, inspiration and validation to the construction of systematic methods.

This paper reports on focus groups held by the authors with a self-selected group of privacy professionals, discussing privacy risk measurement and privacy risk assessment more broadly. The overall conclusion from the focus groups and subsequent plenary discussion was that the measurement of privacy impact is extremely challenging – but professionals also felt that it was very worthy of academics' continued attention.

The contribution of this paper is to present insights into why privacy professionals find privacy risk assessment difficult – both in terms of making the judgement and recording it in a meaningful way. This includes factors such as types of harm, number of

people affected, typical versus worst cases, organisational versus individual impacts, objectivity, and effects that are intrinsically hard to quantify such as violations of rights. The information collected feeds into research to improve privacy risk measurement and its surrounding processes, and identifies some clear research gaps.

2 | RELATED WORK

In this section, we briefly review academic work on the topic of privacy, privacy risk, and privacy impact assessments.

2.1 | Privacy

In the Universal Declaration of Human Rights, privacy was declared one of the fundamental human rights (Art. 12): "no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence"[4]. Many countries have ratified the universal declaration in their own laws, for example in the European Convention on Human Rights[5] and the OECD privacy framework[6].

Even though privacy is often said to be hard to define[7], it is recognized as a fundamental value especially in the digital world, where privacy is a precondition to enjoyment of other human rights including autonomy[8,9].

Many attempts have been made to unravel the concept of privacy. Nissenbaum's contextual integrity[10] ties privacy to *social norms*: any use or collection of data is considered a privacy violation if it is unexpected in terms of the social norms that apply to the current context. Solove's taxonomy of privacy[7] focuses on *how* privacy violations may occur: through information collection, information processing, information dissemination, and invasion. We have previously described *what* types of privacy can be violated[11]: location, state of body and mind, behavior and action, social life, and media. Each of these strands approaches privacy from a different angle. Taken together, they yield a more complete picture of the complex concept of privacy.

2.2 | Privacy Harms

Many privacy harms are well-known and well-documented. For example, IETF RFC 6973[12] describes privacy harms in the context of internet protocol engineering, including "harms to financial standing, reputation, solitude, autonomy, and safety". Regulators as well as researchers have published lists of known privacy harms[13,14,3], often grouped along Solove's taxonomy into harms caused by information collection, information processing, information dissemination, and invasion[7].

In addition to these well-documented harms, new types of privacy harms have emerged in recent years. *Predictive privacy harms* are caused by the collection and processing of fine-grained high-dimensional data as part of big data analysis. This data is not universally recognised as personal data or personally identifiable information, but can nevertheless be used to re-identify individuals[15] and in profiling to predict their behavior[16]. Therefore, predictive privacy harms are the "inappropriate inclusion and predictive analysis of an individual's personal data without their knowledge or express consent"[17]. For example, the retail chain Target used big data analysis on purchase patterns to predict whether customers were pregnant, and then disclosed that information to marketers[18].

Quantitative privacy harms are caused by excessive and ubiquitous data collection amounting to total surveillance. The key idea in quantitative privacy is that individuals have a reasonable expectation that large quantities of their data will remain private, even though they may not care about the disclosure of each individual data point[19,20]. For example, an individual might be happy to disclose one or a few of their spatio-temporal locations, but will still expect that the entirety of their daily geo-location trace remains private. Quantitative privacy harms through ubiquitous surveillance constitute threats to individual self-development and democratic culture[16].

Networked privacy harms (or interdependent privacy harms) occur in situations where one individual's privacy is affected by another's actions. As a result, even individual control over data is not sufficient to ensure an individual's privacy. Networked privacy issues are common in social networks[21], for example in photo tagging[22] or co-location sharing[23].

2.3 | Privacy Impact Assessments

Privacy impact assessments (PIAs) have been introduced as a means to anticipate these privacy harms and mitigate their negative impacts. Historically, privacy impact assessments originated from ideas and processes used for environmental and social impact

assessments as a means to identify "future consequences of a current or proposed action"[24]. PIAs should anticipate privacy impacts in a broad scope, and in particular use a broad scope when considering the types of privacy, the affected individuals and organizations, and their expectations. In addition to identifying privacy problems, PIAs should also propose solutions.

The European data protection regulation GDPR[2] requires organizations to perform privacy impact assessments in cases that involve surveillance, data sharing, or new technologies. Several national regulators have published guidelines for privacy impact assessments, including the French CNIL[25] and the British ICO[26]. PIAs are seen as a good way to achieve another GDPR mandate, privacy by design and default[1].

In addition to these regulatory efforts, academic research has proposed improvements to PIA processes by making them more systematic and structured[27] and by proposing formal modeling techniques for privacy threats[14].

However, an analysis of the current practice of privacy impact assessments revealed several shortcomings[28]: First, organizations tended to focus on organizational risk instead of risk to the data subjects; Second, proposed countermeasures typically addressed the effect, instead of the cause, of privacy risk. Third, as a consequence, PIAs ensured that organizations were compliant with regulations, but did not ensure the creation of privacy-friendly products.

As a result, further research into more rigorous and transparent processes for privacy impact assessments is needed.

2.4 | Risk Assessment

After identifying privacy risks, PIAs typically assess each risk in terms of its impact and likelihood. This approach is very similar to traditional risk assessment processes, for example from computer security. For example, the NIST guidelines[29] use a five-point scale for the impact and likelihood of security risks, from "very high" to "very low", and provide abstract textual descriptions to indicate where individual risks might be located on the 5-point scale, for example indicating impact scores for "significant" versus "major" financial loss. The final risk level is determined from a table that lists risk levels for all combinations of impact and likelihood scores.

For privacy risk assessment, the only deviations from this generic process are the contents of the textual descriptions and the granularity of the scales. For example, the CNIL PIA guidelines[25] use a four-point scale and the OWASP top-10 privacy risks in web applications use a three-point scale[30].

The use of these coarse three-point or five-point scales means that the resulting assessment of privacy risk may be more like a rough guess than an accurate measurement. We have previously argued that these rough guesses for privacy risk are of limited value and not informative enough[3], in particular for the five purposes of (1) quantifying the effect of privacy controls, (2) comparing the effects of different controls, (3) analyzing trends in privacy risk over time, (4) computing a system's aggregate privacy risk from its components, and (5) ranking privacy risks.

2.5 | Measuring Privacy Impact

There is very little research on how these rough guesses could be turned into accurate measurements of privacy risk. Even though many privacy metrics have been proposed in the literature[31], they mainly measure properties of privacy-enhancing technologies instead of privacy impact.

We have previously proposed some initial steps towards privacy risk measurement[3], which clearly showed the need for more research in particular with regard to more fine-grained measurement, combination of metrics, and validation of metrics.

3 | METHODOLOGY

To find out how privacy risk assessments are carried out in practice and what challenges privacy professionals encounter "in the wild", we organized a one-day workshop with 17 external participants. The initial group of participants came from ongoing discussions on Twitter and existing research collaborations, and additional participants joined after open public invitations via Eventbrite and Twitter.

We started the workshop with four one-hour focus groups running in parallel (4-6 participants in each group). After the focus groups, most participants gave brief presentations in a plenary meeting in which they further clarified and expanded on thoughts from the focus groups. Prior to the workshop, we had obtained ethical approval for the focus groups and the research process

including the questions asked, see Appendix A. All participants received a detailed information sheet stating the main aims of the study, were informed of their right to withdraw at any time, and signed a consent form in advance.

Our participants included an information governance manager from the English National Health Service (NHS), committee members of the National Association of Data Protection and Freedom of Information Officers (NADPO) and the Information and Records Management Society (IRMS), a university information governance manager, academics and research students in privacy, a data protection officer for a large company with 80M customers, several independent IT security and data protection advisers, a former head of data protection of a major UK government department, a data protection expert from one of the big accountancy companies, a consultant in data protection in the charity sector, an information governance manager for a local authority, and a privacy campaigner. Some individual participants fit to several of these labels.

We recorded both focus groups and presentations and had them transcribed by a professional transcription service. Below, we identify quotes from the focus groups as *FG* and quotes from the plenary session presentations and discussions as *PS*. We analyzed the transcribed data using inductive thematic analysis. The analysis was based on methodology as described by Braun & Clarke[32] and included the following steps: 1. familiarization with the data, 2. generation of initial codes, 3. searching for themes, 4. reviewing themes, 5. defining and refining themes, 6. writing up the final analysis. Inductive thematic analysis is data-driven; hence theme development was not restricted by an existing coding framework or the researcher's interest in the area.

4 | RESULTS FROM THE THEMATIC ANALYSIS

Three themes have emerged from the thematic analysis of the data: (1) privacy in contemporary society and how this affects privacy risk assessment, (2) current practices and procedures in privacy risk assessments, and (3) issues and challenges. These three themes are explored in the following sections. The first of these throws up the fewest surprises for those familiar with general privacy issues – readers in that category may wish to skip to our findings on privacy risk assessment in practice (Section 4.2).

4.1 | Privacy in contemporary society and how this affects privacy risk assessment

4.1.1 | Defining privacy

One of the main issues raised during the study was definition and perceptions of privacy; how privacy is determined, defined and experienced by contemporary people and how this affects PIA practices. Defining privacy has proved to be a challenging task. Practitioners mentioned that there is a lack of a widely accepted definition of privacy, which has as a consequence the lack of a common language with regards to privacy related topics. Therefore, any discussion on privacy, or issues related to privacy protection and risk of privacy have been mainly based on people's perceptions of privacy, which however seem to be vastly different.

"People's view of privacy is very different." (R1, PS6)
"Some people's expectation of privacy is different to another's" (R3, FG3)

Participants mentioned that the recent unceasing development in technology and the increase in data collection and sharing has changed the dynamics of privacy, in a both individual and collective way. The electronic collection and sharing of data have made personal data more accessible than ever, and thus, privacy has been more threatened than ever.

"it's potential greater impact because obviously information, electronic information could just be spread worldwide, much more so than paper" (R2, FG2)
"by living your life in any digital way you're already at risk" (R1, PS3).

In contrast to the increase of data collection and sharing, people's knowledge on the topics has remained comparatively stable. The level of people's awareness and understanding of data collection and sharing, as well as privacy, has been perceived to be low. It has been noted that there is an inadequate digital literacy in people, while there is limited awareness on privacy and data sharing.

"So and I think people's understanding of risk generally is just so low, that they're not really, it's just that fear, they don't understand what the risks are" (R2, FG2).
"how much people are aware of how much their lack of agency about their own data is actually costing them now?" (R4, FG1)
"I'm saintly on that score, but I have colleagues who have keys and passwords stuck to their screen, you know, so it's useless, isn't it?" (R8, PS9)

Participants felt that educational provision on privacy and privacy risks is limited. As a result people are unprepared and struggle to take appropriate measures to protect their privacy.

"so it's digital literacy, so I don't really want everyone in the world to be able to go and log in and look at, you know, stuff and know that you're away and have this conversation in public because we probably wouldn't shout it in a bar. We wouldn't probably shout in a bar to this, we'd probably have a conversation about it" (R4, FG1).

Moreover, the public has been perceived to be unable to understand the mechanisms behind data sharing, and in most cases, they are unable to understand or acknowledge the dimensions of data-sharing. This is due to the complexity of data-sharing that takes place. This lack of knowledge, both on an individual and organizational level, together with the over-collection of data noted, increases privacy risks; and this was one of the areas that privacy experts felt that they need to contribute to *"arming data subjects and organizations with a better understanding of data sharing implications."* (R1, PS1).

4.1.2 | The digital era

This inadequate digital literacy and the lack of knowledge towards privacy, data sharing and protection has cultivated a culture of ignorance towards privacy impact. People fail to understand the privacy impact and thus, are less cautious towards their data and privacy and end up ''*gifting their data away"* (R1, FG1) with no hesitation and without realizing the consequences of their actions.

"and frankly the general citizen couldn't care less" (R1, PS3)

Moreover, the increased use of technology in everyday life also has an effect on people's awareness and cautiousness towards privacy and privacy impact. It has been mentioned that the use of technology in everyday life (*"surprised just how much data a car takes..."* (R1, FG1)), as well as the increased surveillance, together with the normalization of it and the increased use of social media and smart devices has made people even less cautious of privacy protection and risks. People are being conditioned in ceding their data and see this as being the norm.

"And that is the biggest part I think that has yet to be transparent to this generation, is how they are used and exploited as a digital persona. And that part of sharing data is totally hidden." (R1, PS6)

Another point that was made by numerous participants is the 'hidden part' of how data are being used or exploited by the collecting and sharing companies. As mentioned above, there are misconceptions and lack of adequate knowledge and awareness when it comes to the use and flow of data, which is an alerting issue, when it comes to privacy and privacy impact.

"Because the evidence behind it, even when you talk about children using things like social media, Facebook, etc., the children that understand how to use privacy settings use them. [...] The EU Kids Online project has clear evidence that children believe privacy settings work. Which means that they think their personal data they put on are seen by the people that they understand that they are friends with. The hidden part of how that is used is how the system uses them." (R1, PS6)
"but they're now going to enable that distribution to commercial companies, journalists, charities, think tanks, etc. [...] Somewhere in the chain something is failing in terms of understanding either privacy impact or data protection requirements or that distribution isn't possible" (R1, PS6)

Another issue discussed was that people are commonly unable to acknowledge the actual power data might have. In several instances, they have been using them as a currency for gaining access to services, without however being able to estimate their

value. Most people are commonly willing to provide personal data for getting access to a huge range of services, from ordering food online to public services and social media. This is again something that is being related to both the normalization of those actions, providing data, as well as to the low public awareness.

R3: *"Yeah, I'd say sort of like whether I want to read the privacy policy of Domino's pizza or–"* R4: *"No, you just want a pizza, just give me the pizza."* R3: *"I just want a pizza, just give me the pizza. Stop trying to sell my data, just give me pizza. (Laughs)"* (FG3)

Age has been described as a detrimental factor to this ignorance culture and any related behaviour towards privacy and privacy protection and risk. Younger people have been brought up in the digital era, where the use of the online world is inevitable. The increased use of social media for various purposes, from socialization to entertainment and career purposes, leads to the normalization of data sharing and thus commonly to an increased willingness to share data. Younger people's perceptions are being shaped by this culture and actions, and younger individuals end up developing an increased carelessness to privacy impact and protection.

"Because generation X is brought up with the fact that Facebook is the norm" (R3, FG3)

4.1.3 | Ethics, human rights and the law

However, it is not always up to people's willingness to share private data. People's control over data, as previously mentioned, is sometimes limited, experts have noted, due to the increased data sharing between different organizations and corporations. The volume of information an individual is required to process, in order to figure out the journey of data collection and sharing is beyond human cognitive abilities; and thus it is questionable whether data sharing could be an informed choice for an individual.

"trying to get people to make informed decisions is really difficult because of the sheer insane fire-hose volume of information they'd have to pass in order to make – and the cognitive load of making decisions all the time, every day, about things which are trivial until they're not and you don't know what, in advance, what is going to be trivial and what's going to be significant is just – we've basically created a world way outside our biological capability to handle" (R3, FG1).

In addition, it has also been noted that data sharing is indeed not always an informed choice, for various other reasons, such as lack of information and consent regarding the actual collection and/or use of data. For instance, one of the participants has mentioned that there is a lot of data collection and sharing within the educational sector, from children who commonly remain uninformed.

"especially in the education sector it's based on public interest and they don't have a choice about it. A lot of it also is signed by teachers, by those in the education sector or given by parents. So again they have no ability to give autonomy or consent, or perhaps even told how the data is being used. So how do you assess that privacy impact for children who perhaps don't even know their data is being collected?" (R1, PS6).

Taking a step backward, privacy is a concept very much depending on the way that data is perceived as a concept. According to participants, people tend to falsely define data as just numbers and statistics, while they fail to understand that most of the times data consist of bits of private information. They perceive data as information that are commonly *"depersonalized or deidentified or anonymized or pseudonymized"* (R1, PS6) and thus, they feel that they cannot be linked to actual individuals and traced back to them; something however that is not true, as data sharing between different sources has become so wide that the identification of individuals through connecting and analyzing different databases is more often possible than not.

"But then when people start linking your datasets together for research or not, they're beginning to build up a bigger picture about you purely through circumstance." (R3, FG3)
"That jigsaw matching" (R2, FG2)

Practitioners have also been really skeptical towards privacy and human rights. All of them have made a clear distinction between individuals' and organizations' rights on privacy, and they were more concerned about the individual rather than organizational side. They noted that individuals' rights should always be kept in mind and safeguarded, which however might not always be the case.

"So I think there is kind of an ethical duty to consider the impact of privacy theoretically on people even if they're not all that bothered themselves because they don't know what they're talking about most of the time" (R3, FG1).

Questions were raised in terms of privacy and the potential ways of privacy impact having an effect on individuals' rights. One of the participants gave the following example:

"I don't know if you've seen that story this morning about how people who have any record of having visited or having sought out mental health care are being denied life insurance because they have decided that there is a correlation between risk of, you know, having to pay out essentially and accessing mental health services. So, yeah, you know, see what I mean about people in groups being evil?" (R1, PS10)

The lack of a comprehensive legislation was another issue discussed. Similarly to ethics, concerns were raised by participants in relation to legislation and available policies. They noted that policies and legislation around privacy, together with the multiple jurisdictions involved in privacy talks – as the collection and sharing of information has no geographical boundaries – are not practical, effective and considerate of its global nature.

R4: *"Because the law is so badly written that PSD2*[1] *says they cannot say to people you cannot do that any more, and it's just – it's the most insane..."* R3: *"I know, crazy."* R4: *"..two bits of legislation working together."* (FG1)
"These fines are less than what you're talking about in terms of GDPR. So it's cheaper to kill somebody than it is to have a data breach. Is that realistically equitable?" (R1, PS11)
"It's the mushroom approach, keep them in the dark and feed them bullshit and they'll be fine, which is why most privacy policies are utterly unfit for purpose because, as you say, if you were really honest with people about what you were doing, there would be an absolute outcry" (R3, FG1)
"Chain of contract is required now under GDPR. We laugh at that because a lot of our suppliers are in America. They may or may not be within Privacy Shield and we're slowly getting to terms with that" (R1, PS2)

4.2 | Current practices and procedures of PIAs

4.2.1 | Quantifying privacy impact

"And it's really hard to then put all that together and actually come up with something vaguely in the right area." (R1, PS3)

It has been widely agreed that assessing privacy impact is a challenging task, which is still in its infancy and needs a lot of input and guidance, by both privacy experts and practitioners.

"what is it, where is it, who are you doing it, how long have you had it, how old is it, where is it going, what system is it going on, where is that stored? All those things, those questions that we're all asking and putting forward, well, it depends, it depends. Any true data protection person is going to say, well it depends, on – so what is privacy? Well, it depend (laughs)" (R3, FG3).

Assessing privacy impact has been described as a highly complex procedure, taking into account several factors, which might change, based on the purpose and perspectives of the assessment (e.g. organizational vs individual). It requires the selection of the factors that should be measured, and the data required for that, as well as the application of those factors into several different scenarios.

Similar to privacy, there has been a lot of discussion on the vagueness of the definition of PIA. Some participants mentioned that it is impossible to work on PIAs effectively, if they cannot even define them. Moreover, except the lack of well-established

[1]The EU Payment Services Directive, Directive (EU) 2015/2366

definitions for both PIAs and privacy, there is also a lack of a common language in relation to those practices, as well as a lack of shared patterns and templates.

"If you can't define, how do you measure it?" (R2, FG3)

The lack of 'basic principles and templates' was a core issue raised during the study. According to practitioners, having a common language could result in experts establishing better communication with each other and being able to share issues of concern and perspectives around PIAs, both within and outside the organizations/institutions they are in.

In addition, there is also limited training provided in the field, and there is a wide range of different current practices found within different sectors. Practitioners have been trying to create various metrics for assessing privacy impact and each one is viewed as an isolated attempt to approach the issue. As previously mentioned, one thing that has been agreed by all participants is the difference between impact on the organization and the individual.

"Privacy risk assessment, whose privacy and whose harm?" (R6, FG3)
"Am I going on sort of impact as in how I would deal with it in an organizational sense? Or on a personal sense?" (R1, PS7)

Most experts mentioned that PIAs are usually focused on the organizational impact instead of the individual, which according to them should be of greater concern; however the focus direction depends mainly on the sector.

"Predominantly risk assessments are all about business impact, business impact assessments, what's going to happen to the business [...] when it comes to that privacy impact, it's only ever going to be theoretical" (R1, FG1)

As noted before, numerous factors have been identified as contributing to privacy impact. Three factors have been mentioned by all participants as being taken into account when conducting PIAs, the type of impact, the severity of it, and the likelihood of this impact actually occurring. Based on those, together with others, experts have been trying to quantify impact in various ways:

"I try and pull out two impact scores, one of a typical and one of what's the worst case" (R4, FG1)
"we've got an overall risk framework in the NHS, which has a very sort of five-five-five matrix, from one to five, impact if something bad happened, times probability of that happening" (R2, FG2)
"I put together a five-point scale on about six or seven fundamental rights, so loss of privacy agency, loss of control of data, physical integrity, mental integrity, employment, finance, and then I've tried to assign some values to it." (R4, FG1)
"I built a scale of — like a one-to-five scale based on – I mean statements of risk to data subjects on a number of categories, you know, so discrimination at work, lots of discrimination at work, losing job, those type of things." (R4, FG1)
"I've got a scale that says, my low, medium, high and critical and minor is in terms of privacy and financial, physical integrity, measurable integrity and employment." (R1, PS5)
"I designed my own data protection impact assessment that I can send out to the people who are wanting to input into the system, rather than me having to do the impact assessment, I send it to them and they fill it out" (R1, PS7)

Some have also been relying on standardised PIA products:

"We're currently using OneTrust to do our DPIAs on. It's a huge upgrade from spreadsheets, email and my memory." (R1, PS2)

Some practitioners have mentioned that while quantified measures are being designed and used in practice, privacy impact is something that possibly cannot be effectively quantified.

"But algorithms for me just don't seem to work." (R1, PS8)

Even though all participants have been designing PIAs and trying to come up with ideas to measure privacy impact in various ways, they have shown a low confidence in their work. During the study a lot of hesitation has been expressed by the greatest majority of participants, who were in many instances questioning the credibility and effectiveness of their work.

R3: *"to sort of be able to measure yourself against other people as well is that you can say, 'okay, I'm really shit at this.' "*
R1: *"But so is everybody else"* (PS9)

In addition, it has been mentioned that sharing PIA practices could increase practitioners' current knowledge, as well as their confidence, while it could also increase PIAs' actual effectiveness.

"And I think we've beaten ourselves up as an industry, we often beat ourselves up an awful lot about not having the answers. And I love coming to days like this because it's like, hey, everyone's as shit as I am." (R1, PS4)

4.2.2 | Content and context

Due to the complexity of assessing impact, participants have frequently noted that PIAs are 'very specific to a particular scenario'. As a result, PIAs may have little that is transferable to different situations and contexts. This is something, however, that comes in addition to the attempts to quantify privacy impact and, it is the main reason why practitioners are expressing concerns about the effectiveness of any metrics in real practice.

"so we need to not totally get married to the idea that the metrics rule." (R3, FG1)

More precisely, participants have noted that every PIA is conducted for a specific scenario, and is based on a specific content, in terms of data and context, while it is usually set into a specific time-frame. Different risk factors arise in different situations, contexts and time-frames and there is a great difference between different sets of data, and what they contain, in terms of privacy.

"You need to identify the inherent risk factors, the hot spots, in your risk universe of different datasets under different conditions being processed for different purposes." (R1, PS1)

A couple of participants mentioned that even the data labelled as 'sensitive' may vary, in terms of level of sensitivity and impact. In different contexts and time-frames the impact of this information changes in severity and thus, the impact might either increase or decrease likewise. According to participants there are circumstances where the impact of sensitive information might be quite low. One participant mentioned that even though religion is labelled as sensitive data (GDPR "special category"), it might not be probably for the Pope.

"one person's sensitive data is another person's, you know, so-what." (R5, FG2)

Quantifying the context has been perceived as a highly challenging task. Participants have noted that quantifying the impact for specific contexts might be a manageable task, but using the same means for assessing the impact in different contexts is commonly ineffective. Participants have also seemed to be puzzled about how context could be quantified.

"So this data in one context might be perfectly innocuous, but in another context it could be really significant, but quantifying context is for greater minds than mine." (R3, FG1)

Moreover, the instability of PIAs over time was another issue raised during the study. Most participants were really concerned of this over time instability of privacy impact, and the potential effects that this might have on PIAs. The interesting thing is that there is no time frame or limit that could be set, as the circumstances might change.

"Something might change in the day after you've done the risk assessment which puts the data at higher risk." (R2, FG1)
"Risk profiles can change, threat profiles can change." (R1, PS1)

As a result, PIAs' accountability is highly short lived, which is something that needs to be also taken under consideration.

4.2.3 | Process rather than template

"PIA is a living document" (R4, FG3) and it works more effectively *"as a process and not a template"* (R5, PS9). Practitioners have acknowledged the content, context and time-frame specificity of PIAs and they described PIAs as a continuous process,

which should be updated and informed of potential changes in all these parameters. This is the only way, proposed by practitioners, to avoid PIAs losing their credibility and effectiveness in capturing the actual impact. However, in actual practice PIAs tend to be treated as just a standard procedure and as metrics and templates; something that causes the so-called 'black-hole effect' of PIAs. Practitioners noted that PIAs are commonly completed as a standard procedure, then being put on a shelf or in a drawer and just being referenced as a line in metrics.

"And all of the rich data is either entirely free-form, non-standardized and subjective or does have potential to get analyzed for severity and trend, but is never put together and that's all lost." (R1, PS1)

According to practitioners, in order to produce meaningful and credible PIAs, all these factors need to be included in metrics and quantifying measures used, while a greater focus need to be given to the constant need for updating any change. One of the participants has mentioned that the complexity of PIAs might be captured by AI, and AI could provide an insight into taking account different conditions; while another described how the procedure could be more individually driven.

"So creating maybe a virtual person and seeding data and tracking and modelling how it's used and, you know, estimating what the effect on an actual human being might be, I mean that's a long-term thing, but maybe that's an approach to studying it." (R1, PS10)
"So we would like increasingly to consider in privacy impact assessment, can you give some sort of usage measurement to the individual at the point of collection? And can you enable them throughout the lifetime of the data, the life cycle of the data, be able to see and assess continually [...] But that we have some sort of autonomy returned to the individual about can I assess myself that I'm still comfortable with the impact on my privacy at a later point in time, other than the point of collection?" (R1, PS6)

4.2.4 | Types of impact

"So impact is a spectrum, ranging from annoyance to being tortured to death essentially." (R1, PS10)

There has been a lot of discussion around impact and how this is defined. All participants distinguished between different types of impact and severity, while it has been agreed that impact is highly 'person-dependent' .

Physical, mental and economic impact were the main categories suggested by most experts. However, numerous other subcategories have been mentioned during the study. Interestingly, practitioners mentioned that there are so many differences between these types of impact, making it really challenging to have a formula that fits them all. Some of the different types of impact mentioned by practitioners were:

"loss of agency, emotional distress, loss of certain freedoms: speech, movement, association." (R3, FG1)
"discrimination in the workplace due to data leakage [...] mental integrity [...] Stress is quite common." (R1, PS5)
"risk to the individual's sense of identity [...] from knowing that there has been that breach" (R3, FG2)
"breach of your rights in technicality but no actual harm." (R3, FG1)
"But because it was the BNP[2]*, you know, people were discriminated against and lost their jobs and whatever. And whatever you think of that [...] morally and, you know, ethically, if you stand back and look at it just on the facts, it's discriminating against people for their political opinions, which we're not supposed to do, even if their political opinions are repugnant."* (R3, FG1)

In lots of cases impact is not single sided, and different types of impact might be included in a single case. Linking different types of impact together makes PIAs even more challenging and demanding. In real life practice it is usually a blend of different types of impact that practitioners are dealing with and have to measure and calculate, however all under the same metrics.

"how I relate financial loss to privacy loss to physical and mental loss effects, on the same one to five scale." (R1, PS5)

[2]The British National Party (BNP) is a far-right political party in the UK. Their membership data was published on the internet in 2008 and found to contain police staff and people in other public roles.

Moreover, practitioners made many references to societal impact and how this is linked to individual impact. Societal impact sometimes needs to be considered, alongside individual. Societal impact, while consisting of individual privacy impact, might be fundamentally different from it, as it focuses on the greater picture instead of individuality.

"Well, what would be the impact to the individual? Well, if it's 80 million individuals, what I'd call a community impact is actually pretty high." (R4, FG1)
"10 medical records versus 100 medical records. The impact to the individual, obviously contextualized to each individual is different, but the general impact is the same to the individual; it's the breadth of individuals you could impact upon is one scale." (R1, FG1)

4.2.5 | Impact vs probability

"It's how repetitive your processes are. So it's probably how often you do an activity and how often that is likely to go wrong. If you wanted to actually calculate it out, that's the way I would do it." (R4, FG2)

However, PIAs do not rely just on impact measurement and predictions but also on likelihood. Most participants mentioned likelihood as one of their main factors in PIAs equations and metrics. According to them, the probability of an estimated impact to occur is a factor that determines PIAs as much as impact does. This is something that adds to the complexity of PIAs and makes them even more person, context and time-frame dependent.

"Privacy I think is a bit weird, because you might have a really, really low probability of something bad happening, but the impact if it did, would be really, really high" (R2, FG2)

Some practitioners proposed that probability is even more significant than impact itself, and that this is the main factor that determines PIAs. Moreover, probability has been distinguished into two different categories, the likelihood of data being exposed, and the likelihood of the data being exploited. Likelihood is more likely to be impacted by controls than impact,

"when we're talking about data privacy risk, the kinds of things that we can impact are mainly about probability. It's very rarely about impact." (R1, PS1)

Some experts noted that probability is a factor that is more easily defined, but also, quantified and effectively used in metrics. In addition to impact, they described probability as being less person-dependent and complex, as probability has not so many different types as impact.

"I think human likelihood you can bring faux methodology into it, but with impact it has to be, what would it feel like if I had no shoes, or..." (R1, FG2)

4.3 | Issues and challenges

4.3.1 | Over-complicated and time-consuming

Most of the discussion on issues and challenges of PIA practices focused on the effectiveness of current practices and techniques. More precisely, many practitioners expressed concerns about the effectiveness of both techniques and approaches followed by experts. Except for coming up with more effective metrics and measuring techniques, privacy experts thought that the approach towards PIAs needs to change and needs to be wider, more inclusive and tied to real practice.

"...the inability to quantify anything in a useful, consistent way. And the inability to aggregate, trend analyze, scale, the outputs of that, rather than just having bunches of pieces of subjective paper in a drawer that nobody puts together later." (R1, PS1)

Otherwise, PIAs will become more of a bureaucratic procedure, with no practical application and need.

One of the main challenge of PIAs was that it is perceived to be over-complicated and too time-consuming. Some practitioners mentioned that PIAs require a great deal of time to process, while the outcome is commonly questionable in terms of credibility, due to its complexity.

4.3.2 | It is all about guts

Another issue discussed by practitioners was that due to its complex nature and lack of effective methodology, PIA remains commonly *"guesses and averages"* (R3, FG1) blended with guts, personal biases and prejudices.

Practitioners have shown an increased concern in the use of personal intuition in a job that they felt should be more based on consistent metrics and methodology.

"So it's a lot of blunt gut instinct" (R3, FG2)
"a lot of gut feeling and very little room for nuance" (R3, FG2)
"I don't think I'm going it a good way, I'm confessing. I use gut." (R1, FG2)

4.3.3 | Moving fast

Moreover, privacy, following technology, is moving fast and concerns have been raised on whether current PIA practices manage to keep up with the technological changes and advancements.

"And people are — it's almost like technology and privacy is just going — they're not working at the same speed as each other." (R4, FG3)
"Because you don't know (laughs). Sometimes you don't know until it happens. You can do the best data protection impact assessment in the world and you think you've covered everything. It's a working document, and within two weeks of doing one, something will happen that you never contemplated at all." (R4, FG3)

Privacy experts believed that there is a gap between practices and technology, especially in terms of changes and improvements, which should be reduced in order to be able to produce effective and credible PIAs. For instance, a few participants mentioned smart devices and how these are linked to privacy of both the owners as well as their social connections.

"But I have friends and family who choose to have these [smart speakers]. So when I go over to theirs for a glass of wine, do I expect them to turn it off? Or do I expect them to tell me there is one in the room? Those kind of things." (R3, FG3)

5 | DISCUSSION

Based on the thematic analysis, and in addition to the issues and challenges described above, we identified four main areas where our expert participants felt that more work is necessary: a common language, metrics for privacy and privacy risk, pragmatic support for getting the job done, and regulation vs. choice.

5.1 | Common Language

As already described in Section 4.2, the participants identified a need for a shared language.

"Identifying a common language set that we can all use to describe the same things that we're talking about." (R3, FG2)

The data protection discourse in the EU is grounded in legal terminology that is relatively stable. However, the language in e.g. the GDPR text does not extend into the realm of privacy impacts and the characterisations of the various associated harms. Such extensions may be useful, and could possibly be derived on the basis of linguistic and thematic analysis of data protection case law and regulators' notices.

"in the actual ICO ... notes, in the monetary penalties, you notice certain particular words, so having read a few of them in the last few weeks you get a trend, so I think there is a book, a language book that they use for certain things." (R4, FG2)

However, communication of and around PIAs should take place in a broader context than just a legal one. Stakeholders should be involved in consultation on PIAs, and business decisions potentially based on the outcomes of PIAs; framing PIAs in legalistic language might have the counterproductive effect of dressing them as compliance exercises. Ultimately there should be a

practical and jargon-free language to discuss privacy impacts with all stakeholders. This might even be supported by pictorial representations, such as privacy icons[33]. P3P for web privacy policies[34] may be taken as a scope-limited precursor of such a language, defining a common language for the legal components of a privacy policy, with an attempt to represent them in a "human-readable" format.

The different types of language around PIA – legal, regulatory, business, consultative, practical – are interconnected, and this might be formalised by defining ontologies and their relationships through some kind of refinement hierarchy.

5.2 | Privacy and Privacy Risk Metrics

There is an inherent conflict between the numeric measurement of privacy risk and the rich, nuanced situations that real people in the real world find themselves in because numeric measurements are necessarily abstractions from the complexity of real-word scenarios. As such, we cannot expect that numeric privacy risk metrics accurately reflect every aspect of privacy risk. For example, realistic privacy risks that are hard to quantify include cases when "*by chance you get a celebrity in your hospital and all of a sudden people want to know and they go and look*" (R2, FG2), or cases when third parties can re-identify individuals because they happen to have specific additional information.

In addition, even if a numeric metric was found that included all possible cases weighted by their probability, this metric would tend towards the average and would thus hide low-probability worst-case risks. Organizations that rely on these numeric average-case metrics may then make ill-informed decisions about how to handle privacy risks and as a result may leave important risks un-mitigated. It is important to keep in mind that worst-case privacy impact can destroy lives, and numeric measurement that conceals worst-case risks is therefore unlikely to be acceptable.

More work is needed to find effective ways to "assign numbers" to privacy risk while preserving the nuance and detail of real-world worst-case risks. One way forward could be to combine average-case metrics with textual descriptions of worst-case risks. These descriptions could be structured similarly to misuse cases in privacy engineering[14]. An effective combination of metrics and misuse cases would allow to adequately consider non-quantifiable harms and would eliminate the need to find (possibly bad) proxy metrics for every aspect of privacy risk.

5.3 | Pragmatism

The naive approach to risk management suggests comprehensive analysis and treatment of all risks. In practical situations, however, there is resource limitation across the board. There certainly will not be finance to pay for controls to mitigate every risk; there may not even be time to explore every risk; company boards may only show an interest in risks above certain financial or reputational limits, and may not have time to sign off on any residual risks in the risk register.

All this asks for a pragmatic approach, where privacy risk management focuses quickly on the most serious risks, which implies the need for *triage*[35].

> *"That's why triage is absolutely vital as [...] was talking about. You need to identify the inherent risk factors, the hot spots, in your risk universe of different data sets under different conditions being processed for different purposes. Same for your suppliers. And flag the things that invite the most impact from the most frequent incidents."* (R1, PS7)

One approach to triage is to do a rough assessment of risky *areas* rather than individual risks first, for example by looking at individual subsystems.

> *"what I want to do is find out – and also be able to say to my management, these are my riskier systems, this one is more risky than that one, those ones aren't very risky."* (R1, PS12)

This still implies a need for *ranking* of risk measures, and possibly also for aggregate estimate of risk.

> *"at the end of the day, I'm actually trying to just prioritise. I'm not trying to get an absolute figure of risk; I'm trying to just work out which my riskiest stuff is and which my least risky stuff is so I can put my limited resources to where the most risky stuff is."* (R4, FG1)

Triage may also be on the basis of criteria or measured values, rather than on relative ranking only. In this case, *threshold* values in some risk measurement framework need to be established, which provide some objective level of privacy risk above which risks definitely need to be mitigated.

All these pragmatic approaches stand a chance of reducing the requirements on privacy impact metrics, but still present a need for at least an ordinal scale for privacy risk.

5.4 | Regulation vs. Choice

A frequently made argument is that individuals should make their own privacy choices instead of paternalistically being "provided" with privacy. As a result, it is the individuals' own responsibility when they make the choice to give up their privacy. However, this argument fails to acknowledge that the playing field has been rigged against individuals by large corporations – this is the surveillance capitalism described by Zuboff[16].

To illustrate why the argument for choice and against regulation is faulty, we draw an analogy to public safety. Let's assume a busy road that people frequently need to cross, and let's assume that they can choose one of two road sections to do so. Accident statistics show that one of the crossings results in 100 accidents with pedestrian injuries per year, and the other in only two. However, the individual persons crossing the road do not necessarily know about these statistics, so for them both sections are equal and they might choose one or the other randomly or based on convenience. In contrast, governing bodies and regulators are aware of the accident statistics and can thus decide to implement safety measures such as pedestrian lights or zebra crossings.

The situation is similar in privacy: individuals are often not aware of global data flows and what the consequences of data sharing might be. In addition, even if individuals are aware, their means of controlling global data flows are limited. This lack of awareness and lack of control limit the individuals' ability to make fully informed choices.

To address the lack of awareness, more work is needed in digital literacy. For example, work on *explainable privacy* might develop ways to convey privacy concepts and risks to users and thus enable individuals to make informed privacy decisions. In addition, the lack of control points to a need for regulatory action with the aim of reducing privacy risks for everybody.

6 | CONCLUSION

The academics involved went into these discussions with privacy professionals with a slight hope that practical privacy impact metrics could be extracted from or inspired by current professionals' activity. It is clear that the state of practice is not near this point yet. This is partially because privacy risk itself is heterogeneous, badly understood, and certainly far removed from successful objective quantification. Nevertheless the discussions brought a few areas to the fore where further research could contribute to making privacy impact assessment more systematic and transferable, by uncovering in which areas the professionals are currently aware of what they are unable to know and do.

Even though the discussions made it clear that privacy impact is hard to pin down, they also emphasized the importance of the assessment process in a society where privacy impacts are both ubiquitous and near-invisible. Considering the experiences of how PIAs are being dealt with in practice by organisations, we are of the view that the process privacy impact assessment itself could be a lot more transparent and accountable. PIAs should not end up in a drawer. The GDPR and regulators do not currently demand or even strongly advise the publication of PIAs, and quality of PIAs is assured only implicitly and indirectly via the quality and independence of the GDPR role of data protection officer. For those processing activities where the GDPR and associated national legislation mandate privacy impact assessment, regulation could be strengthened to demand independent assessment, (redacted) publication, and external audit.

ACKNOWLEDGMENTS

This work was supported in part by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/P006752/1. We thank the expert participants of our workshop on "Privacy risk: harm, impact, assessment, metrics" in January 2018.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

References

1. Cavoukian A. Privacy by Design: The 7 Foundational Principles. https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf; 2011.
2. European Parliament and Council of the European Union . Regulation (EU) 2016/679 of the European Parliament and of the Council 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 (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC; 2016.
3. Wagner I, Boiten E. Privacy Risk Assessment: From Art to Science, by Metrics. In: 13th International DPM Workshop on Data Privacy Management. Barcelona, Spain: Springer. 2018 (pp. 225-241).
4. United Nations . The Universal Declaration of Human Rights. Resolution 217 A, United Nations; 1948.
5. Council of Europe . *European Convention on Human Rights*. Strasbourg: Council of Europe . 2010.
6. OECD . Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data. Article C(2013)79, Organisation for Economic Co-operation and Development; 2013.
7. Solove DJ. A Taxonomy of Privacy. *University of Pennsylvania Law Review* 2006; 154(3): 477-564.
8. Bernal P. *Internet Privacy Rights: Rights to Protect Autonomy*. Cambridge Intellectual Property and Information LawCambridge University Press . 2014
9. Wachter S. Privacy: Primus Inter Pares — Privacy as a Precondition for Self-Development, Personal Fulfilment and the Free Enjoyment of Fundamental Human Rights. SSRN Scholarly Paper ID 2903514, Social Science Research Network; Rochester, NY: 2017.
10. Nissenbaum H. Privacy as Contextual Integrity. *Wash. L. Rev.* 2004; 79: 119.
11. Eckhoff D, Wagner I. Privacy in the Smart City – Applications, Technologies, Challenges and Solutions. *IEEE Communications Surveys & Tutorials* 2018; 20(1): 489-516.
12. Cooper A, Tschofenig H, Aboba B, et al. Privacy Considerations for Internet Protocols. Tech. Rep. RFC 6973, Internet Architecture Board (IAB); 2013.
13. Commission Nationale de l'Informatique et des Libertés . Privacy Impact assessment (PIA) 3: Knowledge Bases. https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-3-en-knowledgebases.pdf; 2018.
14. Deng M, Wuyts K, Scandariato R, Preneel B, Joosen W. A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements. *Requirements Engineering* 2011; 16(1): 3–32.
15. Narayanan A, Shmatikov V. Robust De-Anonymization of Large Sparse Datasets. In: IEEE Symposium on Security and Privacy. IEEE. 2008 (pp. 111–125)
16. Zuboff S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs. 1st ed. 2019.
17. Crawford K, Schultz J. Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review* 2014; 55: 93.
18. Duhigg C. How Companies Learn Your Secrets. *The New York Times* 2012.

19. Gray D, Citron D. The Right to Quantitative Privacy. *Minnesota Law Review* 2013; 98: 62.

20. Citron DK, Gray D. Addressing the Harm of Total Surveillance: A Reply to Professor Neil Richards. *Harvard Law Review Forum* 2012; 126: 262.

21. Marwick AE, boyd d. Networked Privacy: How Teenagers Negotiate Context in Social Media. *New Media & Society* 2014; 16(7): 1051-1067.

22. Biczók G, Chia PH. Interdependent Privacy: Let Me Share Your Data. In: Financial Cryptography and Data Security. Springer. 2013 (pp. 338–353).

23. Olteanu AM, Huguenin K, Shokri R, Hubaux JP. Quantifying the Effect of Co-Location Information on Location Privacy. In: Cristofaro ED, Murdoch SJ., eds. *Privacy Enhancing Technologies*No. 8555 in Lecture Notes in Computer Science. Springer International Publishing. 2014 (pp. 184-203).

24. Clarke R. Privacy Impact Assessment: Its Origins and Development. *Computer Law & Security Review* 2009; 25(2): 123-135.

25. Commission Nationale de l'Informatique et des Libertés . Privacy Impact assessment (PIA) 1: Methodology. https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf; 2018.

26. Information Commissioner's Office . Data Protection Impact Assessments (DPIAs). https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/; 2018.

27. Oetzel MC, Spiekermann S. A Systematic Methodology for Privacy Impact Assessments: A Design Science Approach. *European Journal of Information Systems* 2014; 23(2): 126-150.

28. Puijenbroek vJPM, Hoepman JH. Privacy Impact Assessments in Practice: Outcome of a Descriptive Field Research in the Netherlands. In: 3rd International Workshop on Privacy Engineering (IWPE). San Jose, CA, USA: IEEE. 2017.

29. National Institute of Standards and Technology (NIST) . Guide for Conducting Risk Assessments. *NIST special publication* 2012; 800-30 r1.

30. Stahl F, Burgmair S. OWASP Top 10 Privacy Risks Project. https://www.owasp.org/index.php/OWASP_Top_10_Privacy_Risks_Project; 2017.

31. Wagner I, Eckhoff D. Technical Privacy Metrics: A Systematic Survey. *ACM Computing Surveys (CSUR)* 2018; 51(3).

32. Braun V, Clarke V. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 2006; 3(2): 77-101.

33. Mehldau M. Iconset for Data-Privacy Declarations v0.1. https://netzpolitik.org/wp-upload/data-privacy-icons-v01.pdf; 2007.

34. W3C . The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. 2006.

35. Devey C. *A Triage Playbook: Privacy Harm and Data Incident Response in the UK*. PhD thesis. City University of London, 2019. Forthcoming.

APPENDIX

A FOCUS GROUP QUESTIONS

- How do you assess privacy risk?

- How do you assess the likelihood of privacy risks?
- How do you assess the impact of privacy risks?
- How does the scale of privacy risks influence the impact?
- How do you assess the scale of privacy risks?
- How does the sensitivity of data influence the impact?
- How do you assess the sensitivity of data?
- How do user expectations influence the impact?
- How do you assess user expectations?
- How does the harm caused influence the impact?
- How do you assess the harm caused by privacy violations?
- What other factors influence the impact of privacy risks?
- What gaps do you see in privacy risk assessment?