Tool-Assisted Risk Analysis for Data Protection Impact Assessment

Salimeh Dashti, Silvio Ranise

Tool-Assisted Risk Analysis for Data Protection Impact Assessment

Salimeh Dashti[1,2] and Silvio Ranise[1]

[1] Security and Trust - Fondazione Bruno Kessler, Trento, Italy
[2] DIBRIS - University of Genoa, Genoa, Italy
{sdashti,ranise}@fbk.eu

Abstract. Unlike the classical risk analysis that protects the assets of the company in question, the GDPR protects data subject's rights and freedoms, that is, the right to data protection and the right to have full control and knowledge about data processing concerning them. The GDPR articulates Data Protection Impact Assessment (DPIA) in article 35. DPIA is a risk-based process to enhance and demonstrate compliance with these requirements. We propose a methodology to conduct the DPIA in three steps and provide a supporting tool. In this paper, we particularly elaborate on risk analysis as a step of this methodology. The provided tool assists controllers to facilitate data subject's rights and freedoms. The assistance that our tool provides differentiates our work from the existing ones.

Keywords: Data Processing Impact Assessment, Privacy Risk Analysis, Impact, Rights and freedoms

1 Introduction

Privacy Impact Assessment (PIA) has been gradually developed to assess the impacts on the privacy of a project which involves the processing of personal information [42]. It has been widely adopted and studied, e.g., [5, 6, 38, 9, 40, 28, 8, 35, 15, 24, 11]. The PIA is a tool to *help* [24] controllers who *wishes* [11] to demonstrate their compliance. It has not been an obligation until the General Data Protection Regulation (GDPR) came to enforcement.

The GDPR aims to protect data subjects concerning the processing of personal data. Recital 84 says "in order to enhance compliance with this Regulation where processing operations are likely to result in a high risk to the rights and freedoms of natural persons, the controller should be responsible for the carrying-out of a Data Protection Impact Assessment (DPIA) to evaluate, in particular, the origin, nature, particularity, and severity of that risk". Later in article 35, it provides a minimum guideline for carrying out the DPIA, and do not lay down any further set of requirements. The GDPR is unclear how to implement it [3, 13, 38, 18, 7, 39]. Different legal bodies and academics, e.g., [12, 25, 4, 7, 13, 1, 32, 2], started to introduce guidelines and tools to help controllers conduct the DPIA. These

works are short in either providing assistance (they work more like a checklist), including all steps of the DPIA (e.g., monitoring), or applying to all domains.

We propose an iterative tool-assisted methodology to assist controllers in different steps of DPIA. We organized our methodology in three steps: Processing Analysis, Risk Analysis, and Run-time Analysis. Each step generates a document. For the DPIA, the asset is the data subject's rights and freedoms. That is to respect the right to data protection, stated in recital 78, and the right of data subjects to fully control their personal data, stated in article 12 to 25.

The focus of this paper is the *Risk Analysis* step. We elaborate on how our methodology assists controllers to facilitate data subject's rights and freedoms. For which, we introduce the main features of the first step, *Processing Analysis*, and its output necessary for the second step. The reader is referred to [17] for further reading on the first step. We leave the description of the third step out, as it is an on-going work.

Plan of the paper. Section 2 recalls the tools and techniques we have adopted to our methodology. Section 3 discusses the first two steps of our methodology. Section 4 discusses the related work, and Section 5 concludes the paper.

2 Background

We briefly recall some notions and techniques that we have used in our work. The descriptions are partial and focus only on the relevant concepts.

The authors in [33] have employed a tool-supported framework for risk modeling and evaluation, called RiskML. The framework is supported by a modeling language and a quantitative reasoning algorithm to analyze models. It allows the user to observe how risks propagate and perform what-if analysis. The RiskML language relies on the following components: a *situation* illustrates the context in which a system is used; an *event* illustrates the situation that has a negative impact on goals; and a *goal* illustrates what the stakeholder aims to protect. These components connect to one another with relations. A situation can connect to an event with the following relations: *expose* when it opens the system to the event, *protect* when it decreases the likelihood of the event, and *increase* when it magnifies impact of the event. Events and goals connect by relation *impact*. All relations can be weighted to determine the significance of their effect. For example, the weight of an *expose* relation determines how likely is that the situation leads to the event.

In [21, 29], authors propose a tool-based methodology and a technique to integrate legal compliance and security checks. Their tool checks the compliance of user-specified access control policy p concerning access control related articles in the European Data Protection Directive $p\prime$. Such that a policy p refines a policy $p\prime$ iff every authorization requests permitted or negated by p is also so by $p\prime$.

Figure 1. An overview of our methodology.

3 Our Methodology

A (D)PIA must do the following [41]: (1) identify the information flow; (2) identify privacy risks and implemented safeguards, and introduce new safeguards to address the identified risks; and (3) review and update throughout the life of a project. Accordingly, we propose an iterative methodology to conduct DPIA that comprises three steps: *Processing Analysis, Risk Analysis* and *Run-time Analysis*, shown in Figure 1. The controller needs to conduct the methodology for every data processing they operate in the organization.

Our tool generates a document at the end of each step, namely *Processing specification, Risk summary*, and *Asset and event mapping*. It is difficult to establish a consensus on what needs to be reported in a PIA report [42, 37]. In [37], authors have listed the elements that a PIA report needs to contain for different audiences. The list has four main categories: *general system information, assessment information, PIA quality signals*, and *accountability*. The documents that our tool generates cover the four mentioned categories as follows: the document *Processing specification* contains *general system information* and *accountability*; and the *Risk summary* contains *assessment information* and *PIA quality signals*. The document *Asset and event mapping* reports all changes and events that arise during the data processing, as requested in article 33.5 .

We do not provide any pre-assessment to check whether conducting the DPIA is necessary, because the minimum requirements that a DPIA shall contain, articulated in article 35.7, has already been stated in different articles of the GDPR. Article 12 requires a written description of the processing operation, as requested by article 35.7.a; article 5 requires purpose limitations and data minimization, as requested by article 35.7.b; article 32 requires to assess risks to rights and freedoms of the data subject, as requested by article 35.7.c,d. Yet, at the end of the first step, our tool informs the controller whether s/he needs to conduct a DPIA for the data processing in question, based on data type and data subject involved, the scale of processing, and usage of new technologies. We recommend the controller to continue, though.

Article 35.1 mandates controllers to conduct a DPIA. However, the GDPR obliges processors to comply with many of the requirements which apply to controllers, e.g., implement appropriate technical and organizational measures

to ensure the security of processing. For the sake of this paper, we only refer to controllers as the responsible role to follow the steps of our methodology.

Running Example. To illustrate the concepts of our methodology, we consider the following example, taken from [21].

An organization, called ITOrg, needs to compute the salary slips of its employees. Each Employee shall fill in a form, named profile, with some information such as name, surname, address, number of kids, type of contract, etc., and send it to ITOrg. Employees also need to give their consent to use their information to compute the salary slip. ITOrg delegates computing of the salary to its Fin(ancial) Dep(ar)t(ement). The Fin Dept receives selected parts of the profile in a document called fin_profile. In turn, Fin Dept sends it to a company called ACME to compute the salary. Once done, ACME sends the salary slip back to Fin Dept, in a document called salary, and Fin Dept forward it to Employees.

3.1 First Step: Processing Analysis

The first step of our methodology requires a written description of the processing operation. For that, the controller shall provide information, such as: who collects which data (data type), from who (data subject), for how long and why; how the data flow [41, 40], etc. Our tool asks the controller to provide this information, by filling a form and drawing a Message Sequence Chart (MSC).

We realized that controllers find it difficult to identify data type and the data subject. To assist them, we have associated the most likely data type and data subject to each economic sector. We have taken the sectors from the Standard Classification of Economic Activities in the European Community.[3] The controller need only to select the sector to which their organization belongs (e.g., education, Manufacturing, construction), and the tool provides a list of related data types and data subjects that the controller can select from. The controller can add others if they do not find what they need in the proposed list. But they need to justify the added items. For more information we refer the reader to [17].

Next, the controller shall draw the MSC of the data processing in question, to capture not only the flow of data but also to identify: (i) the granted permissions to each role who receives or collects the data; (ii) data subject's concerns in each stage of data processing; and (iii) supporting assets used to process data in each stage, for which they need to specify the name of vendor and product, and its version. The processing stages are: *collection, flow* and *processing*; and the permissions are: *r*ead, *w*rite, *u*pdate, and *t*ransfer.

Data subjects have different concerns about their data in different stages of processing. For example, during data collection, they may be concerned about giving too much data; once collected, they worry about the ways their data can be misused. Recital 75 of the GDPR lists some damages caused by data processing: discrimination, identity theft, and damage to the reputation. Other works [12, 27, 19, 18, 36] have introduced some further concerns, such as: induced disclosure,

[3] `https://ec.europa.eu/competition/mergers/cases/index/nace_all.html`

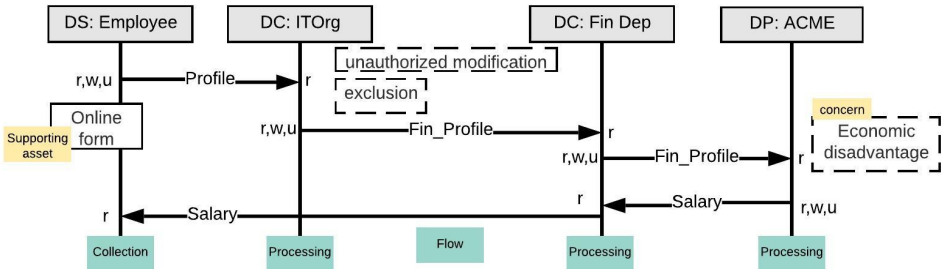

Figure 2. Annotated Message Sequence Chart of the running example.

secondary purpose and identifiability. In this paper, we use the term *concern*, when referring to both damages and concerns. We have used all the concerns mentioned in the literature and the recital, and mapped them to the relevant stage of processing. Table 1 lists an excerpt of the mapping. We have embedded the mapping into the tool. While the controller is drawing the MSC, the tool provides a list of concerns related to that stage.the controller needs to specify the most relevant concerns they believe data subjects might have, according to the type of data and data subject involved, the purpose of data processing (e.g., using new technology), how data are processed, and who else gets to see the data. They need also to specify the supporting assets used in each stage of the processing. That is to be used later in the second step of our methodology.

Once the controller completes the MSC, our tool outputs the user-specified access control, supporting assets, and list of concerns in JSON format. It also generates the *Processing specification* document from information provided by the controller.

Figure 2 illustrates the MSC of our running example, where data subject Employee can read, write and update the document Profile, while controller ITOrg can only read it. Data subject Employee gives profile data using the supporting asset online form (captured by solid rectangles). Data subjects concern (captured by dashed-border rectangles) Unauthorized modification, Exclusion and Economic disadvantage in *processing* stage.

Table 1. An Excerpt of the Mapping

Privacy Concern	Problematic Data Action	Privacy Goal	Stage
Identity theft	Combination Appropriation	Unlinkability	Processing
Induced Disclosure	Interrogation	Integrity	Processing
Unauthorized modification	Distortion	Integrity	Collection Processing
Economic disadvantage	Distortion poor judgment	Transparency	Collection
Exclusion	Unauthorized purpose	Transparency	Processing

3.2 Second Step: Risk Analysis

Before describing the second step, we introduce the protection goals. We have employed six protection goals to incorporate privacy and data protection [23], that are: confidentiality, integrity, availability (the CIA), unlinkability, transparency, and intervenability. Recent research efforts have come up with these goals [31, 22, 16, 30], to provide an interdisciplinary standard model to asses and judge the consequences of utilizing complex IT systems concerning privacy and data protection [23]. As the CIA are well-known, we discuss only the last three. *Unlinkability* relates to the requirements of necessity and data minimization as well as purpose determination, purpose separation, and purpose binding [23], as requested by article 6.4.e and 32.1.a. *Transparency* enables direct controls by entitled entities, such as the data-processing organization itself, a supervisory authority, or the affected human individual whose personal data is processed [23], as requested by article 5.1.a and 12.2. *Intervenability* refers to the requirements that data subjects are effectively granted their rights to notification, information, rectification, blocking and erasure at any time, and that the controller is obliged to implement the appropriate measures [14], as requested by article 12.2.

Classical risk analysis protects assets that organizations value, while for a DPIA assets are data subject's rights and freedoms. Controllers shall respect the data subject's rights and freedoms with regards to (1) data protection, and (2) having full control and knowledge about data processing concerning them. To respect (1), controllers are required to take appropriate technical and organizational measures (recital 78), including the ability to ensure the ongoing confidentiality, integrity, availability, and resilience of processing systems and services (article 32.1.a). To respect (2), controllers shall facilitate the exercise of data subject rights under articles from 15 to 22. Accordingly, we divide the second step of our methodology into two sub-steps: (1) supporting data subject's right to data protection, (2) supporting data subject's right to have full control and knowledge about data processing concerning them.

1. Supporting data subject's right to data protection. Controllers are required to take appropriate technical and organizational measures to ensure confidentiality, integrity, availability, and resilience of processing systems and services. Thus, we need to assess whether the implemented security control in the organization in question is appropriate. We need also to identify the vulnerabilities of supporting assets [12, 11, 10] and security controls, as they are the source of attacks; that would impact not only the controller but also the involved data subjects. For example, fax-ready HP OfficeJet inkjet printers come out of the box with static buffer overflow vulnerability, which allows remote code execution (CVE-2018-5925). Another example is Microsoft ADFS 4.0 Windows Server 2016 and previous versions (Active Directory Federation Services) which suffer from SSRF vulnerability (CVE-2018-16794). Below, we discuss our approach to (i) assess the appropriateness of security controls and (ii) identify the vulnerabilities of supporting assets and security controls.

(i) Assess the appropriateness of security controls. Knowing what is an appropriate security level is not trivial. But, we need to ensure that there is

a minimum level of security according to the type of data resident on the organization under consideration, as stated in FIPS publication 200 [20]. In this publication, they introduce three sets of security control baselines, named: low-impact, moderate-impact, and high-impact.

Our tool suggests the controller the baseline according to the most sensitive data type that the organization handles. According to the identified baseline, the organization needs to implement one of the appropriately tailored security control baselines from NIST Special Publication 800-53, forth revision [26]. As such, we assign the baseline to data type, as follows: low impact to *personal data*, moderate-impact to *personal data with high-risk*, and the high-impact to a *special category of personal data*. While the first and last data types are taken from the GDPR, we have introduced *personal data with high risk* [17], to make the classification more fine-grained. The data assigned to this category are more sensitive than personal data but less that special category of personal data. For example, geographical location data and financial data are categorized as data with high risk.

To assess whether the baseline controls are implemented, we use a questionnaire, called *Control Questionnaire*, based on the Publication 800-53r4. The *Control Questionnaire* has three sections for the CIA. For example, it asks 'Do you have dual authorization?' (confidentiality), 'Do you review or restrict inputs to trusted sources?' (integrity), 'Do you have a daily backup?' (availability). The controller's response could be: "Yes", "Partially implemented", "No", and "Not applicable". Whenever the answer is one of the first two, the controller shall specify the security mechanism of the control. According to the number of security controls implemented for each goal, we evaluate the likelihood of them to be compromised. The *Control Questionnaire* reports the implemented security controls, what remains to be implemented, and the likelihood of each goal to be compromised. The likelihood—and impact— evaluation is on a scale from 1 (the lowest) to 5 (the highest). It also produces a JSON file, to inputs to RiskML tool (introduced in Section 2), that we explain later in this section.

(ii) Identify the vulnerabilities of supporting assets and security controls. We use the National Vulnerability Database (NVD).[4] The NVD is the U.S. government repository of standards-based vulnerability management, which includes databases of security checklist references, security related software flaws, misconfigurations, and impact metrics. It specifies the vulnerable object (either an application, an operating system, or a hardware), by name of its vendor and product, and its version. We have made a list from these objects, to show to the controller when s/he needs to specify the supporting asset (see Section 3.1) and the mechanism while answering the *Control Questionnaire*.

Our tool has already generated two JSON files which list the supporting asset, from the MSC chart, drawn in step 1 (see section 3.1), and the security mechanism, from the *Control Questionnaire*. We query the NVD with these two files. If it finds any matches, it retrieves from the database vulnerability type,

[4] https://nvd.nist.gov/general

and the exploitability score.[5] Its output is a JSON file, that is also an input for the RiskML tool. Note that, the *Control Questionnaire* and NVD assist risk analysis and controllers—they do not offer a comprehensive risk analysis.

Controllers may define and enforce access control policies in their organizations, which is enough from a classical risk analysis perspective, but it is not enough to ensure compliance with the GDPR. For example, the classical risk analysis dose not see any risk if a controller uses data subject's personal data for any purpose than the primary purpose for which data are collected, or if they do not give access to data subject to delete or transfer data. However, under the GDPR they put data subject privacy at risks. To avoid such risks, we have used the tool proposed by [21, 29], introduced in Section 2. We have adapted their tool to the GDPR and extended it to generate its input which are user-specified access control policies, automatically from the MSC (Section 3.1). Giving the input, the tool checks it against the access control requirements of the GDPR and reports any non-compliance.

In our methodology, we focus on the processing that involves an individual, not the ones that aggregates data of several individuals by using big data or machine learning techniques. Indeed, there is a lot of data processing which uses neither of them and yet privacy critical. Such as public administrations' process to deliver certificates, e-government processes, or financial transactions that are inherently centered around the personal data of a single individual. For such processing, we consider the necessary access control permissions.

2. Supporting data subject's rights and freedoms. Article 5.1.a requires controllers to inform the data subject about data operation concerning them and their rights (transparency); articles 15 to 22 require controllers to facilitate the exercise of their rights (intervenability). Controllers need to ensure that their actions do not infringe data subject's rights and freedoms, e.g, to ensure they do not probe for more information [34] or use data for an unauthorized purpose (unlinkability). Some of the concerns could be addressed by security controls, such as insecurity [34], identifiability [19], detectability [19]; but not all. In [14, 23], authors have listed some controls to address these protection goals, which are more like actions to take than security mechanisms. Thus, we cannot address them in the *Control Questionnaire*. Sometimes, it is even hard to understand all possible concerns. For example, an organization uses an automated decision-making process to decide on employees' performance. The set of data that has trained the process happened to be biased towards some nationalities. The chance of capturing such bias is really low. Thus, they keep having *poor judgment* towards some employees, which leads to *discrimination* (recital 75). While drawing the MSC (see section 3.1), our tool provides a list of concerns and asks the controller to specify the most likely ones that data subjects involved in data processing in question have. The list could help controllers to rethink what data subject's concern can be.

[5] Please refer to `https://www.first.org/cvss/specification-document` for more information on how the scores are evaluated. Note that we use the CVSS v3.0.

In our methodology, we map the data subject's concerns to possible actions that cause them, called *problematic data actions* [27]. We extend the mapping to privacy goal to introduce proper control for counteracting the problematic data action. Table 1 shows the mapping between them. Using a questionnaire, called *Right Questionnaire*, we assess whether the controller takes any of the problematic data actions, and how they facilitate the exercise of data subject's right. For example, it asks 'can data subjects withdraw their consent?', 'May you use the personal data for a purpose other than what you have specified in the document *Processing specification*?', 'do data subjects know how their data are being processed?'. The protection goals mapped to the questions above are interveinability, unlinkability, and transparency, respectively. The questionnaire report which data subject's concerns (from the MSC) and rights are supported, what has remained to act upon, and how likely it is for them to be compromised. It also produces a JSON file from the report to input into the RiskML tool. The *Right Questionnaire* can be seen as a guideline for controllers to ensure they do not miss anything.

To evaluate the impact of data processing on data subjects, we consider the following indicators: (1) type of data; as the impact increases if sensitive data are involved, such as health data, financial data and political opinion; (2) data subject category; as the impact increases, if vulnerable data subjects are involved, such as minors, asylum seekers and employees [4]; (3) the scale; as the impact increases if a large amount of personal data or a large group of data subjects are involved (article 35.3.b, c); (4) usage of new technology (article 35.3.a); as it increases the scale of the impact. Even if only one of these indicators hold, the impact is high on the data subject. In the first step we asked the status of these indicators. Our tool generates a JSON file to report the existing indicators for the data processing in question and the evaluated impact.

Note that, the tool meant to assist the controller and risk analyst, not to be replaced. For example, the control questionnaire is to assess the implemented security mechanism. The controller and risk analyst are responsible to check they are implemented correctly.

Modeling the risk using RiskML Tool. In summary, our tool has generated the following JSON files, for the data processing in question: 1) from the MSC, user-specified access control policy, list of concerns and supporting assets; 2) from the *Control Questionnaire*, list of implemented and remained security mechanism and the likelihood of the CIA to be compromised; 3) from the *Right Questionnaire*, list of taken actions to address the concerns/rights, remaining ones, and the likelihood of three protection goals (transparency, intervinability, and unlinkability) to be compromised; 4) from NVD result, list of the vulnerable supporting assets and security mechanisms, and their exploitability score and vulnerability type; 5) list of existing indicators and the impact. The first JSON file is used to check compliance of access control policy. The last four are inputs into the RiskML to generate the risk model.

As mentioned in Section 2, RiskML model comprises of situations, events and goals. Situations and events are connected by the *expose* relation, *increase* and

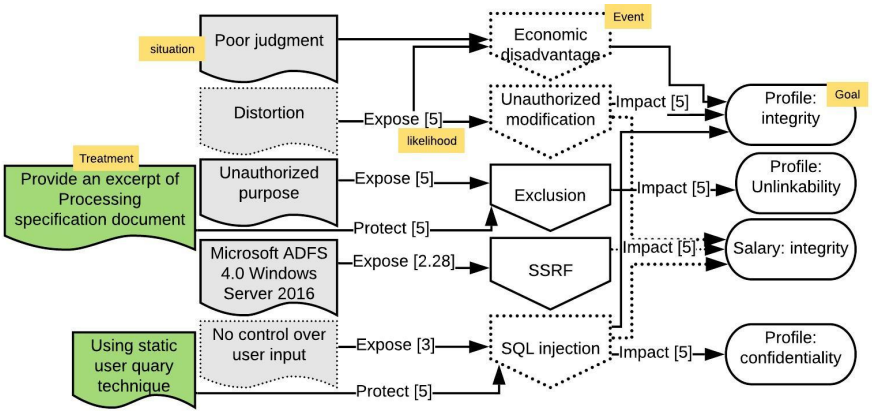

Figure 3. Annotated Risk model of the Running example.

protect. Events and goals are connected by relation *impact*. We have extended the RiskML tool to automatically generates the risk model from the four JSON files mentioned above. *Situations* will be created from the problematic data actions, vulnerable supporting assets and security mechanism. *Events* will be created from the concerns mapped to the problematic data actions, vulnerability types, or threats raise from missing security mechanism. The weight on *expose* relation, that is the likelihood, comes from the Control Questionnaire, Right Questionnaire and the NVD. The weight on *impact* comes from the last aforementioned JSON file. The ontroller can modify the model, add treatments, and conduct what-if analysis.

Our running example deals with financial data, that belongs to *Personal data with high risk*. Thus, controller ITOrg needs to implement security controls for the *moderate-impact baselines*. The impact of such processing is evaluated 5 because the data type involved is financial data (that is personal data with high risk) and the data subjects involved are among vulnerables [4].

The MSC of this example shown in Figure 2 depicts that data subjects use an online form. Answering the *Control Questionnaire* we know that there is no control over user input. This may lead to the *event* SQL injection. They also specify in the questionnaire that they use Microsoft ADFS 4.0 Windows Server 2016 to control accesses in the system, depicted as *situation*. Querying the NVD, the tool has retrieved the SSRF vulnerability, depicted as *event*. The vulnerability has an exploitability score of 2.28.[6] Furthermore, the MSC also shows that the data subject Employee concerns that controller ITOrg modifies their data, or excludes them from knowing other purposes for which they may use their data. Employee also concerns that the processor ACME makes them experience economic disadvantage. From the *Right Questionnaire*, we know that there is no action taken to address such concerns. The mapped problematic actions to

[6] We have normalized the score to 1 to 5.

these concerns are derived from the embedded list into the tool. The controller addresses the event Exclusion by providing a document, namely providing an excerpt of *Processing Specification document* (the output of the first step, see Figure 1). The document informs the data subject about all the possible purposes, their data may be used for. The controller also addresses the event SQL injection, by using static user query technique. The RiskML tool captures how the risk propagates. Figure 3 shows the propagation by dashed lines. In our running example, the situation Distortion which impacts integrity of profile will also impact the integrity of salary. This propagation is captured through the structure of the documents. The processor ACME computes the salaries based on data he receives by document Fin_profile; that is extracted from profile. Thus, any distortion in the profile will affect the Fin_profile, and consequently affect the salary.

We input the user-defined access policy to the tool introduced in Section 2 to report any non-compliance in access control policy. Our scenario is not compliance, as the controller does not grant Transfer access to data subject Employee. From the security point of view, this does not make any problem. However, the GDPR grant right to data portability to data subjects (article 20).

4 Related Work

Privacy Impact Assessment (PIA) has gradually developed from the 1990s onward [38] and adopted in Australia, Canada, New Zealand, and the United States. The UK introduced the first PIA methodology in Europe in 2007. The European Data Protection Board (EDPB) set general requirements for PIAs for the RFID [6] in 2009, and smart metering [5] in 2012. Academic works have largely studied it,CNIL e.g., [9, 40, 28, 8, 35, 15]. D.Wright defines PIA as "a methodology for assessing the impacts on privacy of a project, policy, program, service, product or other initiative which involves the processing of personal information and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimize negative impacts" [42]. In his book, he has proposed a sixteen-step optimized PIA process based on a review of various existing PIA methodologies. When these works were developed, conducting a PIA was not a legal obligation. The UK Information Commissioner's Office (ICO) introduced it as "a tool which can help organizations identify the most effective way to comply with their data protection obligations and meet individuals' expectations of privacy" [24]; and the French Commission Nationale de l'Informatique et des Libertés (CNIL) refer to PIA as the process that "[..] controllers who wish to demonstrate their compliance approach and the controls they have selected, as well as for product providers wishing to show that their solutions do not breach privacy" [11].

The General Data Protection Regulation (GDPR) mandates controllers to conduct a Data Protection Impact Assessment (DPIA), stipulated by article 35, when data processing is likely to result in a high risk to the rights and freedoms of natural persons. Article 35(7) provides a minimum standard for carrying out the DPIA and do not lay down any further requirements. It is unclear how to

implement it [3, 13, 38, 18, 7, 39]. National privacy authorities—such as CNIL [12] and ICO [25]—have proposed guidelines to conduct a DPIA. ICO provides a self-assessment toolkit[7], which offers checklists for different roles, namely controller and processor. The toolkit helps to lead a discussion to find out what needs to be done to make to keep people's personal data secure. On the other hand, the CNIL introduces a tool to help controllers build and demonstrate compliance. It requires controllers to respond to some questions in three categories: 1) to describe the context of processing, e.g., applicable standards, data types involved, supporting assets; 2) fundamental principles, e.g., the proportionality of purpose to collected data, accuracy of data, storage duration; 3) risks, e.g., existing controls, possible risks, and the likelihood and impact. Both of these works resemble a check-list that do not provide assistance—which is not an assessment[8]—while our tool assists controllers in different step of the assessment. For instance, our tool assists the user in evaluating risk in two phases by assessing the implemented security countermeasures and the vulnerabilities that may exist in the supporting assets and also implemented security countermeasures. In contrast, the CNIL tool asks the user how they estimate the risk level with no further guidance.

EDPB proposes an iterative process for carrying out a DPIA [4] consisting of the following seven steps organized in a cycle: (1) description of the envisaged processing, (2) assessment of the necessity and proportionality, (3) data protection measures already envisaged, (4) assessment of the risks to the right and freedoms of data subjects, (5) data protection measures envisaged to address the risks, (6) documentation, (7) monitoring and reviewing. We organized our methodology in three steps, namely: Processing Analysis, Risk Analysis, and Run-time Analysis. Each step generates a document. These steps cover those by the EDPB, as follows: Processing Analysis corresponds to steps (1) and (2); Risk Analysis to steps (3), (4), and (5); Run-time Analysis to step (7); and the three documents generated by our tool cover step (6). The authors of [7] use a three-stage process to conduct the DPIA, namely: preparation, evaluation, and report and safeguard. They use the Standard Data Protection Model (SDM) [14] to demonstrate compliance with the requirements of data protection and identify appropriate safeguards. SDM systematizes technical and organizational measures to protect the rights of the data subjects based on protection goals. The authors use a catalog of data protection measures developed by the technical working group of the conference of German data protection authorities (AK Technik). These works show the requirements and steps to take to comply, while our work assists the user to meet the DPIA requirements and be compliant.

In the academic world, the authors in [13] use UML class diagrams to specify crucial requirements underlying various aspects of a DPIA, such as consent and necessity. Their focus is to integrate security and privacy requirements engineering processes into a DPIA and understand how a previously developed tool for risk

[7] https://ico.org.uk/for-organisations/data-protection-self-assessment/

[8] http://www.isaca.org/Knowledge-Center/Blog/Lists/Posts/Post.aspx?ID=864\&utm_referrer=direct\%2Fnot\%20provided\&utm_referrer=direct\%2Fnot\%20provided\&utm_referrer=

analysis of UML diagrams can be effectively used in this context. The authors of [1] propose UML-based security and privacy analysis. They annotate the UML with the security requirements and check if they are satisfied using UMLsec. To specify the level of privacy, they have extended the Privacy Level Agreement (PLA) to include the following four privacy preferences: purpose, granularity, visibility, and retention. A PLA is an appendix to a service level agreement and provides a structured means to specify privacy preferences and data security requirements. The authors map harmful activities (introduced in [34]), to the privacy threats and appropriate checks. Thus, by identifying the harmful activities in the diagram, they react accordingly. To evaluate impact, they extend the mapping to the privacy target, to identify what has to be protected, and its impact on data subjects and data processor. The authors do not consider to specify likelihood, as they believe that if a privacy threat exists, we control it. They also associate security controls to the mapping. The security controls are based on the NIST 800-53r4, ISO 27001, and German IT baseline. Although we share a similar approach, we acknowledge the fact that privacy and security are two sides of the same coin; thus, we check whether the security controls or supporting assets are vulnerable. Moreover, we check compliance of access control policy, as having them implemented cannot guarantee compliance. Lastly, DPIA is not a one-time activity; instead, it has to continue as long as the life-cycle of the data processing, which the authors of [1] have not considered. In [32], authors have introduced a tool supported by a formal method to verify whether privacy properties are met or not. The authors believe that using such a tool makes the validity of the DPIA more understandable and manageable; also, it verifies whether the DPIA performed by the controller has captured the risks raised from excessive data collection and unauthorized purpose. While using such a tool is useful, they still cannot capture all the DPIA requirements, and instead, focus on purpose and consent. The authors in [2] offer a tool-supported DPIA for Cloud Service Provider (CSP). They have two questionnaires: one to assess whether the DPIA is necessary, and one to establish the effect of interactions among data subjects and CSPs on their rights to data protection. The questions have some pre-defined answers which are associated with privacy indicators and weighted according to the impact they have on them. A global privacy indicator is then calculated based on the indicators. While we share some similarities in the risk analysis phase, our tool is agnostic concerning the technology used to implement the data processing activities. The Risk Analysis step of our methodology is parametric concerning the particular technique used for risk evaluation

5 Conclusion and Future work

The DPIA shall assess the risk to data subject's rights and freedoms (article 35.7.c). The GDPR grants to the data subjects the right to data protection and the right to be fully in control of their data and be fully aware of the processing operation concerning them. In this paper, we discussed our three-step DPIA methodology and its supporting tool. In particular, we detailed the second step dedicated to

risk analysis. This step assists the controller to exercise data subject's rights and freedoms, in two sub-steps: (1) exercising data subject's right to data protection, and (2) exercising data subject's rights and freedoms. For the former, we ensure that the controller has implemented an appropriate level of security controls and that the supporting assets and the security mechanism are not vulnerable. While for the latter, we ensure that the controller has facilitated the exercise of data subject's rights and has informed them about the data processing that concerns them. To conduct effective DPIA, we need to consider legal, organizational, social, and technical aspects. Applying our methodology on the running example showed that our tool helps to govern the process and tame the complexity.

Working with the public administration of the province of Trento, Italy, we realize that different people are involved in risk analysis in particular in evaluating the impact of data processing. This has led to different impact evaluations for similar processing. To address this issue, we plan to propose a methodology to suggest the most related impact to the data processing according to data type, data subject, and type of processing. For the last step of the methodology, titled *Run-time Analysis* (see Figure 1), we plan to integrate Inventory Management and Security Information and Event Management to carry out monitoring of compliance as stated in article 40.4.

References

1. Ahmadian, A., Strüber, D., Riediger, V., Jürjens, J.: Supporting privacy impact assessment by model-based privacy analysis. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. pp. 1467–1474 (2018)
2. Alnemr, R., Cayirci, E., Dalla Corte, L., Garaga, A., Leenes, R., Mhungu, R., Pearson, S., Reed, C., de Oliveira, A.S., Stefanatou, D.: A data protection impact assessment methodology for cloud. In: Annual Privacy Forum. Springer (2015)
3. Alshammari, M., Simpson, A.: Towards an effective PIA-based risk analysis: an approach for analysing potential privacy risks. Tech. Rep. CS-RR-18-01, Department of Computer Science, University of Oxford (2017)
4. Article 29 Working Party: Guidelines on data protection impact assessment and determining whether processing is "likely to result in a high risk" for the purposes of regulation 2016/679. `https://ec.europa.eu/newsroom/document.cfm?doc_id=47711`, accessed: 2019-19-06
5. Article 29 Working Party: Opinion 07/2013 on the data protection impact assessment template for smart grid and smart metering systems ('DPIA template') prepared by expert group 2 of the commission's smart grid task force. `https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp209_en.pdf`, accessed: 2019-19-06
6. Article 29 Working Party: Opinion 5/2010 on the industry proposal for a privacy and data protection impact assessment framework for RFID applications. `https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp175_en.pdf`, accessed: 2019-19-06
7. Bieker, F., Friedewald, M., Hansen, M., Obersteller, H., Rost, M.: A process for data protection impact assessment under the European general data protection regulation. In: Annual Privacy Forum. pp. 21–37. Springer (2016)

8. Clarke, R.: Privacy impact assessments. http://www.xamax.com.au/DV/PIA.html/ (1999), accessed: 2019-22-10
9. Clarke, R.: Privacy impact assessment: Its origins and development. Computer law & security review **25**(2), 123–135 (2009)
10. CNIL (Commission Nationale de l'Informatique et des Libertés): Methodology for privacy risk management. https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf (2012)
11. CNIL (Commission Nationale de l'Informatique et des Libertés): How to carry out a PIA. https://www.cnil.fr/sites/default/files/typo/document/CNIL-PIA-1-Methodology.pdf (2015)
12. CNIL (Commission Nationale de l'Informatique et des Libertés): Privacy risk assessment (PIA). https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf (2018)
13. Coles, J., Faily, S., Ki-Aries, D.: Tool-supporting data protection impact assessments with CAIRIS. In: 5th International Workshop on Evolving Security & Privacy Requirements Engineering. pp. 21–27. IEEE (2018)
14. Conference of the independent data protection authorities of the Federal and State Governments of Germany: The standard data protection model, v.1.0 EN1. (2017)
15. Cuijpers, C., Koops, B.J.: Smart metering and privacy in europe: lessons from the Dutch case. In: European data protection: coming of age, pp. 269–293. Springer (2013)
16. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.H., Metayer, D.L., Tirtea, R., Schiffner, S.: Privacy and data protection by design-from policy to engineering. arXiv preprint arXiv:1501.03726 (2015)
17. Dashti, S., Ranise, S.: A tool-assisted methodology for the data protection impact assessment. in proceedings of the international conference on security and cryptography (2019)
18. De, S., Le Métayer, D.: Priam: a privacy risk analysis methodology. In: Data Privacy Management and Security Assurance, pp. 221–229. Springer (2016)
19. 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 **16**(1), 3–32 (2011)
20. FIPS (Federal Information Processing Standard Publication) 200: Minimum security requirements for al information and information systems. https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.200.pdf (2006)
21. Guarda, P., Ranise, S., Siswantoro, H.: Security analysis and legal compliance checking for the design of privacy-friendly information systems. In: Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies. pp. 247–254 (2017)
22. Hansen, M.: Top 10 mistakes in system design from a privacy perspective and privacy protection goals. In: IFIP PrimeLife International Summer School on Privacy and Identity Management for Life. pp. 14–31. Springer (2011)
23. Hansen, M., Jensen, M., Rost, M.: Protection goals for privacy engineering. In: IEEE Security and Privacy Workshops. pp. 159–166 (2015)
24. ICO (Information Commission's Office): Conducting privacy impact assessments code of practice. https://www.pdpjournals.com/docs/88317.pdf (2014), accessed: 2019-19-06
25. ICO (Information Commission's Office): Data protection impact assessments. https://ico.org.uk/media/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/

data-protection-impact-assessments-dpias-1-0.pdf (2018), accessed: 2019-19-06
26. NIST (National Institute of Standard and Technology): Security and privacy controls for federal information systems and organization. NIST special publication 800-53. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf (2013)
27. NISTIR (National Institute of Standard and Technology Internal Report): Nist privacy risk assessment methodology (PRAM). https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/resources
28. Oetzel, M.C., Spiekermann, S.: A systematic methodology for privacy impact assessments: a design science approach. European Journal of Information Systems **23**(2), 126–150 (2014)
29. Ranise, S., Siswantoro, H.: Automated legal compliance checking by security policy analysis. In: International Conference on Computer Safety, Reliability, and Security. pp. 361–372. Springer (2017)
30. Rost, M., Bock, K.: Privacy by design and the new protection goals. datenschutz und datensicherheit 35, 30–35 (2011)
31. Rost, M., Pfitzmann, A.: Datenschutz-schutzziele—revisited. Datenschutz und Datensicherheit **33**(6), 353–358 (2009)
32. Schulz, W., Wittner, F., Bavendiek, K., Schupp, S.: Modeling and verification in GDPR's data protection impact assessment. https://www.cpdpconferences.org/archive (2019)
33. Siena, A., Morandini, M., Susi, A.: Modelling risks in open source software component selection. In: International Conference on Conceptual Modeling. pp. 335–348. Springer (2014)
34. Solove, D.J.: A taxonomy of privacy. University of Pennsylvania Law Review **154**, 477 (2005)
35. Spiekermann, S.: The RFID PIA–developed by industry, endorsed by regulators. In: Wright, D., de Hert, P. (eds.) Privacy impact assessment, chap. 15, pp. 323–346. Springer (2012)
36. Spiekermann, S., Cranor, L.F.: Engineering privacy. Transactions on software engineering **35**(1), 67–82 (2008)
37. Spiekermann, S., Oetzel, M.C.: A systematic methodology for privacy impact assessments: a design science approach. European Journal of Information Systems **23**(2), 128–150 (2014)
38. Van Dijk, N., Gellert, R., Rommetveit, K.: A risk to a right? Beyond data protection risk assessments. Computer Law & Security Review **32**(2), 286–306 (2016)
39. Vemou, K., Karyda, M.: An evaluation framework for privacy impact assessment methods. 12th Mediterranean Conference on Information Systems (2018)
40. Wright, D.: The state of the art in privacy impact assessment. Computer law & security review **28**(1), 54–61 (2012)
41. Wright, D., Finn, R., Rodrigues, R.: A comparative analysis of privacy impact assessment in six countries. Journal of Contemporary European Research **9**(1) (2013)
42. Wright, D., de Hert, P.: Privacy Impact Assessment, vol. 6. Springer Science & Business Media (2012)