A Tool-assisted Methodology for the 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

Keywords: General Data Protection Regulation (GDPR), Data Processing, Risk Analysis, Compliance, Likelihood, Impact.

Abstract: We propose a pragmatic methodology to the Data Protection Impact Assessment (DPIA) based on a tool capable of assisting users during crucial activities such as data processing specification and risk analysis. Previous work on compliance checking and our experience in developing a DPIA methodology for the Public Administration of the province of Trento in Italy are the basis of this work.

1 INTRODUCTION

According to the Working Party 29,[1] a Data Protection Impact Assessment (DPIA) "*is a process designed to describe the processing, assess its necessity and proportionality and help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data by assessing them and determining the measures to address them*." DPIA is one of the most important activity for an organization to demonstrate compliance with the General Data Protection Regulation (GDPR). Unfortunately, it is complex, time-consuming, and requires expertise in several domains. It is rarely the case that organizations—especially small or middle size ones—can afford the burden of developing an interdisciplinary portfolio of competencies, including cybersecurity and privacy. For larger organizations, another issue is to maintain uniformity of the DPIA across different departments.

To alleviate these problems, we propose a pragmatic methodology based on our previous experience in designing a methodology for the DPIA of the public administration of the province of Trento in Italy[2] and previous academic work on compliance of security policy (Guarda et al., 2017, Ranise and Siswantoro, 2017).

Our methodology is based on a tool that is capable of assisting users with the three main activities of DPIA, namely (1) the analysis of the data processing activities, (2) the assessment of the risks, and (3) the run-time monitoring. In this paper, we discuss the first two steps and leave the description of the third to a future paper as it is still on-going work. Since the methodology and the tool have been designed to tightly cooperate, it is difficult to consider one of them in isolation; instead, we regard their cooperation as the main strength of our work. Despite this, below, we will use the word "methodology" or "tool" when we want to emphasize one of the two aspects.

For activity (1), the tool helps users in carrying out crucial activities such as the functional specification of the data processing activities, the identification of the entities involved, their legal roles, and the access control policies that they must satisfy. For activity (2), it checks whether access control policies are compliant with the provisions of the GDPR and computes the risk level of a data processing activity in terms of the likelihood and impact of a data breach. The ultimate goal of our tool-based methodology is to assist organizations to master the complexity and the interdisciplinary competencies needed for the correct implementation of the DPIA. Indeed, the effective use of the tool must be complemented by adequate training on the key notions of the GDPR and the DPIA. In the rest of the paper, we omit technical details—e.g., the fact that the tool contains also a database of processing activities—and focus on the capabilities of the tool to assist users.

Plan of the paper. Section 2 describes the first two steps of our methodology. Section 3 discusses related work and highlights the differences or similari-

[1] https://ec.europa.eu/newsroom/document.cfm?doc_id=44137

[2] See resolution n. 450 of March 23, 2018 available at http://www.delibere.provincia.tn.it/.

Dashti, S. and Ranise, S.
A Tool-assisted Methodology for the Data Protection Impact Assessment.
DOI: 10.5220/0007932202760283
In *Proceedings of the 16th International Joint Conference on e-Business and Telecommunications (ICETE 2019)*, pages 276-283
ISBN: 978-989-758-378-0

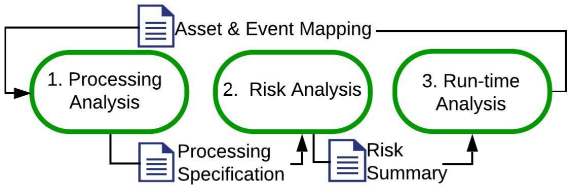

Figure 1: An overview of our methodology.

ties with our approach. Finally, Section 4 summarizes our findings and highlights future work.

2 OUR METHODOLOGY

In each one of the three steps in our methodology (shown in Figure 1), the user is assisted by a tool in gathering the necessary data to produce three documents containing the description of the DPIA activities. Such documents can be used to satisfy the accountability requirements of the GDPR (art. 5. 2) and to support the auditing process by, e.g., a (national) privacy authority.

1. The step **Processing Analysis** outputs a document, called *Processing Specification*, that contains a precise description of the data processing activities, including the collected data, their classes, the data subject categories involved, the purpose, etc.
2. The step **Risk Analysis** outputs a document, called *Risk summary*, that reports the compliance check of access control policies against the GDPR (this is crucial to ensure that data subjects can control the sharing of their personal data in compliance with legal provisions) and the risk levels associated to each defined data processing.
3. The step **Run-time Analysis** outputs a document, called *Asset and Event Mapping*, that contains the associations between the assets (identified in the previous step of the methodology) and the actual entities in the system together with the events that are relevant for data protection so that an Inventory Management system and a Security Information and Event Management can use the associations to detect, at run-time, possible deviations from the protection profiles previously specified or violations of compliance.

Running Example. To illustrate the main concepts of our methodology, we consider the following scenario: an Italian research organization named ITOrg provides complimentary health insurance to its Employees. To this end, ITOrg has chosen the insurance company ACME as a sub-contractor. Employees wishing to opt into the complementary health insurance service shall provide ITOrg with profile data, including first and last name, taxpayer identification number, type of contract (e.g., permanent or fixed-term), work e-mail address, and summary of the medical history of the past 3 to 5 years. ITOrg forwards the information to ACME that processes it to produce an appropriate health insurance contract that, in turn, is returned to the Employee via ITOrg.

Below, we discuss only the first two steps of the methodology and leave the third as future work.

2.1 Processing Analysis

For each processing activity, the first step of our methodology requires to specify the data collected, their classes, how data are grouped in objects,[3] the data subject categories involved (e.g., patients or minors), the purpose (e.g., advertising or billing), which entities are playing which legal roles (e.g., data controller or data subject), when and how the consent is obtained, and the adequacy (necessity, proportionality, and legal basis) of the collected data.

The data categories involved in the complementary insurance scenario are those typically associated to a research organization as ITOrg, namely Employees (researchers and administrative staff), (external) Collaborators (for consulting and cooperation in research projects), and Students (for training periods). Since it is not always easy to identify the data classes and data subject categories involved in a data processing and satisfy the requirements of necessity and proportionality stipulated in art. 35.7.b of the GDPR, the tool supporting our methodology provides hints by using a schema based on economic sectors. Table 1 shows an excerpt of such a schema where the first column reports the sectors taken from the standard classification of economic activities in the European Community (NACE);[4] the second column shows the associated data subject categories; and the third column reports the associated data classes. The legend of the table explains the abbreviations for data subject categories whereas the acronyms used for data classes are the following: PD stands for Personal Data, PD-HR for Personal Data with High Risk, and SPD for Special category of Personal Data. While PD and SPD are taken from the GDPR, we have introduced PD-HR to

[3] A data object can be seen as a record of fields, i.e. a collection of values with given types. Examples of data objects are the rows of a relational database and certain data structures of programming languages.

[4] https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NACE_REV2&StrLanguageCode=EN&StrLayoutCode=HIERARCHIC

Table 1: An excerpt from economic sectors with data classes and data subject categories.

Sectors	Data Subject Categories	Data Classes
Information and communication	E, C, Ci,	PD, PD-HR: location
Telecommunication	N, L, M, P	
Professional, scientific & technical activities	E, C,	PD
legal and accounting activities	N, L, Ci	PD-HR: Financial status
Scientific research & development	S, Pr	PD-HR, SPD
Education	E, C, Pr, S, D, Ci	PD, PD-HR: Financial status SPD:Health Data
Pre-primary, Primary, sport & culture	M, P	

Legend: E(mployee), M(inor), Ci(tizen), P(arent), Pa(tient), C(ollaborator), N(atural Person), L(egal Person), S(tudent), No(minee), De(tainee), P(olice)/M(ilitary) forces, D(isable), Pr(ofessor/Teacher/Trainer)

make the classification more fine-grained to simplify the second step of our methodology (i.e. Risk Analysis). In the first column, the main sectors of NACE are highlighted in light gray (e.g., Information and communication), while those in white are sub-sectors (e.g., Telecommunication). In the complementary insurance scenario, the user selects the sector 'Professional, scientific & technical activities' and the sub-sector 'Scientific research & development.'

The table shows only the sub-sectors which have either more data class or data subject category than its main sectors; as sub-sectors inherit from its main sector. For instance, the complementary insurance scenario which falls under sub-sector 'Scientific research & development,' can access SPD and PD-HR (according to art. 9.2.j) as well as PD from the main sector; and to S(tudent) and Pr(ofessor/Teacher/Trainer) with E(mployee) and C(ollaborator) as data subject categories.

The three data classes are ordered as follows: SPD is a strict sub-set of PD-HR that in turn is a strict subset of PD. Indeed, this means that PD, PD-HR, and SPD have an increasing level of sensitivity. We also consider sub-classes (such as Location or Political Opinion) of the main data classes that are mentioned in GDPR or recitals (e.g., recital 75) for a more precise specification of the data processing activities and—similarly to what has been done above for data subject categories—to ease the Risk Analysis step.

We exploit the information contained in Table 1 as follows. Users indicate only the economic sector to which their organization belongs and the tool returns a series of suggestions for the data classes and data subject categories that are more likely to be relevant to the data processing activities of the organization. In this way, the user is presented with a restricted subset of choices: this, not only, alleviates the burden of considering multiple options but also helps in satisfying the necessity and proportionality requirements of art. 35.7.b of the GDPR. This process is fully transparent to users, who are free to add more data classes and data subject categories as needed, attaching short justifications. The identification of the economic sector and related data subject categories and data classes is done once and for all data processing of an organization.

We use automated reasoning techniques to mechanize the suggestion generation activity. We encode the schema in Table 1 into formulae of propositional logic and use the capabilities of a constraint solver to generate the hints.

While we have validated part of the schematization in Table 1 in our experience of applying the proposed methodology in the public administration of the province of Trento in Italy, we believe that there is room for improvements and refinements. The feedback of users of the tool will be key to improve the precision of the relationships in the table.

We now consider the other information needed to describe a processing activity. Consider again the complementary insurance scenario, although it is tempting to define the purpose of the processing simply as the 'production of the contract for complementary insurance,' this is not enough for verifying that a certain processing step is performed for the declared purpose. What is missing, is the definition of the plan (i.e. the context) to which the processing step belongs, to achieve the declared purpose; see, e.g., (De Masellis et al., 2015) for a discussion on this issue. For this reason, we define the purpose of a processing activity as a plan (i.e. the sequence of steps) that must be executed to achieve a certain goal. The tool supports well-known standards for the specification of plans, such as Message Sequence Charts and Business Process Modeling Notation. To avoid technicalities, here we propose a natural language specification of the plan for the complementary insurance scenario:

1. Employees provide some information, such as full name, taxpayer identification number, type of their contract and the medical history, via a data object insurance complementary form and pass it

to ITOrg;

2. ITOrg informs Employees that the insurance complementary form will be shared with ACME to produce a complementary health insurance contract; and ask them for their consent;
3. ITOrg passes the form to ACME;
4. ACME reads the data insurance complementary form, produces a contract, and sends it to ITOrg;
5. ITOrg forwards the contract to the Employees.

Besides defining the purpose of the processing, the plan specifies: the data objects used, here is insurance complementary form; their content (for the sake of brevity, here we omit the details); the source of data (step 1), here are the data subjects themselves; and when and how the consent is obtained from the data subject (step 2). The plan also helps in clarifying the roles played by the involved entities and why these are entitled to perform some processing step on the data objects. In the complementary insurance scenario, ITOrg—playing the role of controller—mandates (via an appropriate contract) the third party organization ACME—playing the role of data processor—to perform the necessary computations on the data object insurance complementary form provided by Employees—playing the role of data subjects. Some care is needed when considering the capabilities of the data subject. While the GDPR states that data subjects have unrestricted access to their data, an IT system is designed to empower them with only a subset of their rights, i.e. those that are needed to guarantee that the processing achieves its purpose. In practice, data subjects retain all rights but exercise them in ways which are not immediately available in the data processing under consideration. In the complementary insurance scenario, ITOrg empowers Employees with read, write, and update rights on their profile data; but not delete right. While the latter is indeed a data subject's right, it is reasonable that such a "delicate" operation is supported in other ways, and it should not be considered as a lack of compliance the fact that the data subject is not granted all permissions at any time and in any context. The permissions that are granted to the various entities involved in the data processing are specified by an access control policy. The tool supports the specification of such policies written in the Attribute Based Access Control (ABAC) framework (Hu et al., 2013) for its well-known capability to express and combine a wide range of different policy idioms (Jin et al., 2012). The access control policy is specified at the level of organization and it holds for each processing activity specified in the DPIA; this is aligned with the desideratum of organizations to have a single point of administration of policies that must be enforced uniformly across (possibly distributed) operational units. To avoid technicalities, we report a rule of a policy for the complementary insurance scenario in natural language: ACME can "modify" the content of insurance complementary form. We will see below that this rule causes a compliance violation.

2.2 Risk Analysis

The second step of our methodology is organized in two sub-tasks: (i) verifying compliance of access control policies and (ii) evaluating the likelihood and impact. For the former, we adapt to the GDPR the approach in (Guarda et al., 2017, Ranise and Siswantoro, 2017), developed to check compliance for its precursor, the European Data Protection Directive. For the second, there are several available standards to adopt during a DPIA. To allow users of our methodology to choose their favorite approach, we make the common assumption that risk is the product of two factors: the likelihood of an adverse event (such as a data breach) and the impact of the event on the fundamental rights and freedoms of the data subject (e.g., reputational damage). Before describing in details each sub-task, we make the following observation.

While sub-task (ii) can be easily seen to fit in the context of risk analysis as soon as we observe that it is standard to decompose risk into likelihood and impact, it may be less clear why we consider the compliance checking of access control policies—i.e. sub-task (i)—as part of risk analysis. To understand why this is the case, we observe that modern approaches decompose access control in three main components: policy language, model, and enforcement—see, e.g., (De Capitani di Vimercati et al., 2007). The first two define in abstract mathematical terms the syntax and semantics, respectively, of access control policies whereas the third specifies the mechanisms put in place to enforce the policies. One of the main advantage of this approach is to separate the analysis of high-level security properties from the correctness of the enforcement mechanisms. In fact, it is well-known that writing and maintaining policies is an error-prone activity because of the possibility of inserting redundancies, conflicts, and other logical problems—see, again (De Capitani di Vimercati et al., 2007) for an introduction to this and related problems. Similarly, guaranteeing the compliance of access control policies expressed in mathematical terms against legal provisions is difficult, even disregarding the risks deriving from vulnerable implementations of the enforcement mechanisms, that are evaluated in sub-task (ii). Logical errors in access control policies prevent compliance already at the design level and thus constitute an important source of risk to identify and eliminate; sub-task (i) is designed to achieve this objective.

2.2.1 Verifying Compliance of Access Control Policies

The literature on the security analysis of access control policies offers a wide range of techniques to automatically solve several types of policy analysis problems; see, e.g., (Turkmen et al., 2017). One of such problems is that of refinement; a policy p refines a policy p' iff every authorization requests permitted or negated by p is also so by p'. Our tool reduces the compliance of the access control policy p of the organization with respect to the GDPR to check that p refines the policy obtained by instantiating a selected subset of the articles of the GDPR that are relevant to control access to the data processing activity. In this way, every authorization requests permitted or negated by the policy of the organization is also so by the policy derived from the GDPR. The instantiation of the GDPR consists of mapping the legal roles to the entities involved in the data processing as specified in the first step of the methodology as well as taking into consideration the relationships of "mandate" and "empower" between the data controller and data processor or data subject, respectively.

To illustrate, consider again the complementary insurance scenario and recall the rule of the access control policy specified at the end of Section 2.1, i.e. ACME can "modify" the content of insurance complementary form. A rule instantiated from the GDPR to the use case scenario stipulates that ACME (data processor) can "read" (process) the data object insurance complementary form (which contains health data) since the ITOrg (data controller) has mandated ACME to do so. By taking into consideration these two rules, an automated tool for policy analysis is able to detect a violation—because the mandate relation specifies that ACME can only read such an object—and return one or more authorization requests that expose such a violation to help users understanding (and hopefully eliminating) it. Details are in (Guarda et al., 2017, Ranise and Siswantoro, 2017); the use of these policy analysis techniques is fully transparent to the user.

2.2.2 Evaluating Likelihood and Impact

We describe how our methodology and tool support users to evaluate likelihood and impact to determine the risk level of each data processing.

Likelihood. There are several methods and standards (e.g., ISO 27001:2013)[5] for likelihood evaluation that can be re-used almost off-the-shelf for a DPIA. Since users may prefer certain approaches over others, our tool allows importing results from external tools or by manual entry. For instance, a possible (coarse-grained) way to evaluate likelihood is to first identify the functionalities and assets used to implement a certain data processing activity, define indicators to measure the adoption of security mechanisms (such as authentication and access control) and privacy enhancing technologies (e.g., pseudo-anonymization), complement them to estimate the likelihood of a violation, and finally take the maximum of the values as a first approximation of the likelihood. More sophisticated approaches can be used only when needed, e.g., when the first estimate of the likelihood is above an "acceptable" threshold.[6] Our tool supports this refine-and-check approach to likelihood evaluation by permitting the specification of alert conditions (e.g., is the likelihood value above a certain threshold?) to signal that a more precise technique should be used. The alert conditions are predicates, defined on several indicators that measure the likelihood of a security incident. Examples of such indicators are the likelihood that standard security (e.g., confidentiality, integrity, and availability) or privacy (e.g., linkability, transparency, and intervenability) properties are violated. Once the various indicators are calculated, the tool supports the definition of a function to compute a single value for the likelihood (the default function is to take the maximum value of the indicators following a cautious approach). We observe that tracking several indicators is important for conducting a good DPIA for two reasons. First, the values of the indicators suggest which data protection mechanisms to improve to reduce the likelihood of a data breach. Second, the availability of several indicators improves accountability and simplifies the task of impact evaluation—as we will see below.

There are two (easily satisfied) requirements for integrating an external method of likelihood evaluation in our methodology. The first one is to normalize all values on a scale between 1 (unlikely) and 5 (almost certain). The second one (derived from art. 35.7.a of the GDPR) is to complement the values of indicators with a specification of the technological system used to implement the data processing activities that contain at least a description of its functionalities and the assets—including hardware, software, networks, and people—used to store, manipulate, and transfer personal data. In some cases, such informa-

[5] https://www.iso.org/standard/54534.html

[6] For the time being, we leave the task of setting the appropriate threshold value to the user. Such a value may depend on considerations that are difficult to quantify; e.g., the fact that a processing activity supports (or not) one of the main business goals of the organization.

tion is already available as a pre-requisite for applying many of the available methodologies and can be readily imported.

Impact. It is difficult to re-use pre-GDPR methodologies (e.g., ISO 27001) for impact evaluation in the context of a DPIA since they consider the impact on the business goals of the organization (i.e. with respect to the data controller or the data processor) rather than on the fundamental rights and freedoms of data subjects. After the introduction of the GDPR, there has been work to align some methodologies (e.g., ISO 27005) with the GDPR or to provide new ones (e.g., that proposed by the French privacy authority CNiL). While it is possible to integrate the recent methodologies in our tool, we describe our own approach in the rest of this section with the hope of providing further insights in the challenging process of impact evaluation.

Our approach is divided in two phases: first, we ask a “subjective” evaluation of the impact to the user and then we weight the value by using a combination of an “objective” compliance indicators related to the data processing. In the first phase, the tool asks the user to provide a first estimate of the impact level of a security incident for the data processing under consideration, in a scale from 1 to 5 (where 1 indicates the lowest possible severity level and 5 the highest one). To assist users in this phase, the tool provides three sets of information: (a) the defined data classes and the data subject categories, (b) a list of privacy concerns, and (c) a list of potential damages on data subjects that may result from data breaches. The tool also explains the relevance of the sets of information to evaluate the impact.

For (a), the idea is that the more sensitive the data processed, the higher the impact. The tool associates the impact based on the selected data classes, which are normal, high, very high to personal data, personal data with high risk and special category of data, respectively. In case, vulnerable data subjects (e.g., minors) are associated with data processing, the user may increase the value. The tool encourages them not to lower the value, though. The list of potential (physical, material, or non-material) damages provided in (c) is derived from recital 85 and includes loss of control over personal data, discrimination, identity theft or fraud, financial loss, and damage to reputation. The list can be extended, customized, annotated with comments, and shared within an organization to help users share their views and reach a more uniform level in impact evaluation (in large organizations, it is often the case that several users will participate in the DPIA as they are responsible for different collections of data processing activities). In the complementary insurance scenario, it seems reasonable to evaluate the impact as medium-high (value 4 of the scale) since the data being processed is a special category but the involved data subject categories (employee and collaborator) do not seem to require particular attention.

The second phase of impact evaluation requires the user to enter values for a given set of compliance indicators related to the data processing. The indicators[7] are the following: transfer of data (e.g., personal data are kept in the EU or can reach a non-EU country),[8] necessity and proportionality of processing (e.g., purpose is specific, explicit, and legitimate; the data retention period is limited or not; there is a contract with data processors), the privacy information provided to data subjects is appropriate, and organizational measures for data protection have been adopted (e.g., personnel managing personal data has been adequately trained). Concretely, the tool shows a list of sentences about the indicators and the user is invited to evaluate—in the scale from 1 (full) to 5 (very little)—the level of compliance of the data processing under consideration for each statement. In the case of the complementary insurance scenario, a statement ‘employees managing personal data have been trained on the GDPR’ can be rated 1; this means that the personnel has undergone a serious program of training and their level of awareness has been verified.

Once the user has entered the values for all indicators, the tool computes the average and uses it to weight the value of the impact entered by the user in the first phase above. To guarantee uniformity in impact evaluation across an organization, the tool checks whether the values of the compliance indicators are in line with those of similar data processing activities (we say that two processing activities are similar when they deal with the same sets of data classes and data subjects). In case of lack of uniformity, the tool signals the anomaly and asks users to consider further scrutiny to align the value; the final decision is indeed left to the user.

3 RELATED WORK

There are several lines of works relevant to DPIA such as Privacy Enhancing Technologies (see e.g. (Bennett and Raab, 2017)), the Privacy-by-Design

[7] Derived from https://ec.europa.eu/newsroom/document.cfm?doc_id=44137.

[8] In future work, we plan to refine the case of non-EU countries in those that provide adequate safeguards and those that do not as suggested in https://ec.europa.eu/info/law/law-topic/data-protection/data-transfers-outside-eu/adequacy-protection-personal-data-non-eu-countries_en.

approach (see e.g. (Blix et al., 2017)), Privacy Impact Assessment (van Puijenbroek and Hoepman, 2017, Vemou and Karyda, 2018), and Privacy Risk Assessment (Oetzel and Spiekermann, 2014). For the sake of brevity, we focus on those that are more closely related to our approach and classify these works in two broad categories: legal (from national and European privacy authorities) and academic (in the scientific literature about Data Privacy Impact Assessments).

Legal Works. Art. 35 in the GDPR introduces the notion of DPIA, specifies under which conditions the data controller is forced to perform such an assessment, lists a set of requirements that the DPIA must satisfy, and identifies some classes of data processing activities for which the DPIA is necessary (e.g., using new technology, automated processing, large scale of special category of data). The text of art. 35 does not provide detailed guidelines to perform the DPIA and this has stimulated contributions from official bodies and authorities.

For instance, the Article 29 Working Party (WP) proposes an iterative process for carrying out a DPIA 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. Our methodology covers these steps, 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 covers step (6). Along similar lines, national privacy authorities—such as CNiL and ICO—have put forward guidelines to conduct a DPIA (CNil, 2018, CNil, 2015, ICO, 2018). While many such proposals do not provide tool-support, the one by CNiL introduces a tool that requires to fill in a detailed form, containing around 350 questions and guide users through the DPIA. The tool returns a visual representation of the collected data that include the considerations of the Data Protection Officer (DPO) and the Data Subjects (or their representatives). By using the output of the tool, users can decide which data protection mechanisms are needed to reduce risks to an acceptable level with justifications to the DPO and the Data Subjects. While our work shares the same goal of providing a tool to support the implementation of a DPIA, the main difference is that the tool proposed by CNiL resembles a check-list and does not provide effective assistance as it is the case of our tool. For instance, our tool suggests the relevant data subject categories and data classes to the economic sector to which the organization belongs (see Section 2.1 and Table 1); that assists the user to comply with the necessity and proportionality requirements (art. 35. 7. b). While, the tool by CNiL asks the user to identify such information, without any further assessment.

Similar observations hold concerning impact evaluation (see Section 2.2.2)—that as we have already discussed—is a difficult activity because of the need to bridge the large gap between data breaches (technological level) and the rights and freedoms of data subjects (legal and social level). Our tool provides a guided approach to perform such an evaluation with the goal of controlling subjectivity and guaranteeing uniformity across different data processing activities; both goals are difficult to achieve by using a form-based tool such as the one proposed by CNiL.

A distinctive feature of our approach is the emphasis on access control policies and the verification of their compliance against the legal provisions of the GDPR (see Section 2.2.1). As already observed, checking compliance at the logical level of the policies, i.e. disregarding vulnerabilities in their enforcement, is already important to implement a privacy-by-design approach and avoid compliance violations that may dramatically increase the risk level of data processing activities. To further support the importance of checking policy compliance already at design time, it is important to observe that articles 5.1.f and 32.2 of the GDPR—together with recitals 39 and 49—state that personal data should be processed in a manner that ensures an appropriate level of protection against destruction and unauthorized access; indeed, access control is one of the key security mechanisms to guarantee this. To the best of our knowledge, no other work (including those from the academy, see below) considers this aspect of risk analysis.

Academic Works. Only recently, scientific papers have been published on DPIA. For instance, the authors of (Coles et al., 2018) use UML class diagrams to specify crucial requirements underlying various aspects of a DPIA such as consent and necessity. The focus of this work is to integrate security and privacy requirements engineering processes into a DPIA and understand how a previously developed tool for risk analysis of UML diagrams can be effectively used in this context. They argue in which steps of a DPIA, requirements engineering processes may be helpful and supported by a tool. However, they do not consider the risk to rights and freedoms of data subjects which is the focus of the GDPR.

The authors in (Alnemr et al., 2015) offer a tool-supported DPIA for cloud deployments. Their approach is based on two questionnaires: the former aims to assess whether the DPIA is necessary while the latter is to establish how the interactions between the subjects that perform the DPIA and the Cloud Service Providers affect data subjects' rights to data protection. The questions have some pre-defined answers; they are weighted according to the impact they have on privacy; the weights are used to calculate an impact score. Each question is also associated to multiple privacy indicators to capture different privacy aspects (such as sensitivity, compliance and data control) that can enhance or be detrimental to privacy; 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 with respect to the technology used to implement the data processing activities. Consequently, the Risk Analysis step of our methodology is parametric with respect to the particular technique used for risk evaluation.

4 CONCLUSIONS AND FUTURE WORK

We have presented the first two steps of our tool-assisted methodology that combines automated policy analysis techniques with a flexible risk-based evaluation to implement a substantial part of a DPIA. Some ideas have been developed and implemented when contributing to the definition of a DPIA methodology for the public administration of province of Trento in Italy that comprises more than 2,000 processing activities (of which 650 handles special category of personal data) that are distributed across (almost) 100 organizational units. To permit effective use of the methodology and the tool, the training sessions were crucial.

In future work, we plan to investigate how to combine data analytic techniques with selected monitoring tools—such as those for Inventory Management (IM) and Security Information and Event Management (SIEM)—to map assets (first step) and likelihood indicators (second step) and implement the run-time analysis (third) step—cf. Figure 1—of our methodology to make the methodology continuous.

REFERENCES

Alnemr, R., Cayirci, E., Dalla Corte, L., Garaga, A., Leenes, R., Mhungu, R., Pearson, S., Reed, C., de Oliveira, A. S., Stefanatou, D., et al. (2015). A data protection impact assessment methodology for cloud. In *Annual Privacy Forum*, pages 60–92. Springer.

Bennett, C. J. and Raab, C. D. (2017). *The governance of privacy: Policy instruments in global perspective*. Routledge.

Blix, F., Elshekeil, S. A., and Laoyookhong, S. (2017). Data protection by design in systems development: From legal requirements to technical solutions. In *12th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 98–103. IEEE.

CNil (2015). How to carry out a pia. https://www.cnil.fr/sites/default/files/typo/document/CNIL-PIA-1-Methodology.pdf.

CNil (2018). Privacy risk assessment (pia). https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-1-en-methodology.pdf.

Coles, J., Faily, S., and Ki-Aries, D. (2018). Tool-supporting data protection impact assessments with cairis. In *IEEE 5th International Workshop on Evolving Security & Privacy Requirements Engineering (ESPRE)*, pages 21–27. IEEE.

De Capitani di Vimercati, S., Foresti, S., Jajodia, S., and Samarati, P. (2007). Access control policies and languages. *JCSE*, 3(2):94–102.

De Masellis, R., Ghidini, C., and Ranise, S. (2015). A declarative framework for specifying and enforcing purpose-aware policies. In *STM*, volume 9331 of *Lecture Notes in Computer Science*, pages 55–71. Springer.

Guarda, P., Ranise, S., and Siswantoro, H. (2017). Security analysis and legal compliance checking for the design of privacy-friendly information systems. In *SACMAT*, pages 247–254. ACM.

Hu, V. C., Ferraiolo, D., Kuhn, R., Friedman, A. R., Lang, A. J., Cogdell, M. M., Schnitzer, A., Sandlin, K., Miller, R., Scarfone, K., et al. (2013). Guide to attribute based access control (abac) definition and considerations (draft). *NIST special publication*, 800(162).

ICO (2018). Data protection impact assessments (dpias). https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/.

Jin, X., Krishnan, R., and Sandhu, R. (2012). A Unified Attribute-Based Access Control Model Covering DAC, MAC and RBAC. In *DBSec*, number 7371 in LNCS, pages 41–55.

Oetzel, M. C. and Spiekermann, S. (2014). A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems*, 23(2):126–150.

Ranise, S. and Siswantoro, H. (2017). Automated legal compliance checking by security policy analysis. In *SAFECOMP Workshops*, volume 10489 of *Lecture Notes in Computer Science*, pages 361–372. Springer.

Turkmen, F., den Hartog, J., Ranise, S., and Zannone, N. (2017). Formal analysis of XACML policies using SMT. *Computers & Security*, 66:185–203.

van Puijenbroek, J. and Hoepman, J.-H. (2017). Privacy impact assessments in practice: Outcome of a descriptive field research in the netherlands. *International Workshop on Privacy Engineering*.

Vemou, K. and Karyda, M. (2018). An evaluation framework for privacy impact assessment methods. *MCIS Proceedings. 5.*