A Data Protection Impact Assessment Methodology for Cloud

Rehab Alnemr[1], Erdal Cayirci[2], Lorenzo Dalla Corte[3], Alexandr Garaga[4], Ronald Leenes[3], Rodney Mhungu[3], Siani Pearson[1], Chris Reed[5], Anderson Santana de Oliveira[4], Dimitra Stefanatou[3], Katerina Tetrimida[3] and Asma Vranaki[5]

1 HP Labs, UK
2 Stavanger University, Norway
3 Tilburg University, Netherlands
4 SAP Labs, France
5 Queen Mary University of London, UK

Abstract. We propose a data protection impact assessment (DPIA) method based on successive questionnaires for an initial screening and for a full screening for a given project. These were tailored to satisfy the needs of Small and Medium Enterprises (SMEs) that intend to process personal data in the cloud. The approach is based on legal and socio-economic analysis of privacy issues for cloud deployments and takes into consideration the new requirements for DPIAs within the European Union (EU) as put forward by the proposed General Data Protection Regulation (GDPR). The resultant features have been implemented within a tool.

Keywords: Data Protection Impact Assessment, EU GDPR, Cloud, Privacy

1 Introduction

A Data Protection Impact Assessment (DPIA) method aims to identify the main risks of a project with respect to the rights of data subjects concerning their personal data. It is a systematic process to elicit threats to the privacy of individuals, identify the procedures and practices in place to mitigate these threats, and document how the risks were addressed in order to minimise harm to data subjects [26, 14]. DPIAs have been recognised as a key topic for data protection governance in Europe, as they will become mandatory according to the ongoing data protection legal framework reform, in the form of the proposed General Data Protection Regulation (GDPR) [15]. The version of the European Parliament's first reading also incorporates the concept of risk into the DPIA process (cf. Article 32a), in the scope of the DPIA mechanism by mandating data controllers to carry out a DPIA in those cases likely to present specific risks to the rights and freedoms of data subjects. Hence, the concept of risk is embedded in the DPIA process as a pre-assessment stage and a risk analysis would be able to function as an awareness methodology in order for a DPIA to be carried out. Note that in the context

adfa, p. 1, 2011.

of the present analysis the terms DPIA and Privacy Impact Assessment (PIA) are being used interchangeably [17].

A DPIA seems to perform a dual function. On the one hand, it can serve as an accountability mechanism, especially where data breaches or losses occur – in the sense that it allows organisations acting as data controllers or data processors to demonstrate their awareness about the risks concerning privacy and data protection and their commitment in ensuring an effective level of protection of personal data [54]. On the other hand, it can foster the safeguard of privacy and data protection rights [41] in the case of potentially privacy intrusive projects and services, because it requires the controller to systematically consider the intended data processing, the associated privacy risks and the measures to be taken to mitigate these risks from the very outset of its activities [54]. Accountable organisations should embrace DPIAs as part of their overall risk management practices. Unfortunately, today there is a lack of tool support for organisations to perform DPIAs of cloud services.

In this paper, we present the design of a Data Protection Privacy Impact Assessment Tool (DPIAT) developed as part of the EU funded Cloud Accountability (A4Cloud) project[1]. The tool considers a number of information sources from which cloud specific risks and existing countermeasures can be collected and evaluated, in the process of supporting impact assessments for projects considering processing personal data in the cloud. We also propose updated DPIA questionnaires with respect to existing standards and recommendations, building on the expertise of experts from different disciplines from legal research to information security and risk management and to user experience design.

The remainder of the paper is organised as follows: we discuss related work in Section 2. We describe the rationale and approach to construct the proposed DPIA based on legal and socio-economic considerations in Section 3. Our approach consists of three steps: 1) conduct a pre-assessment to determine the need for a fully-fledged DPIA (see Section 3.2); 2) conduct the full DPIA if warranted by the previous step (see Sections 3.3 and 3.4); and 3) perform a risk-based comparison of potential cloud service providers (CSPs) (see Section 4). The DPIA takes a form of a dynamic questionnaire, which aims to collect information from the user about the project under evaluation and its organisational practices. The risk evaluation of potential cloud solutions takes into account some information collected in DPIA and the implementation status of security controls by the CSP. Section 5 presents the DPIA tool design and its dynamic questionnaire to collect information about the project under evaluation and organisational practices, and its automation of steps 1-3 above. The tool produces a report containing several privacy indicators and risks based on the filled questionnaires and the selected CSP.

2 Related Work

Privacy impact assessments are already being rolled out as part of a process to encourage privacy by design [26]: in November 2007 the UK Data Protection Authority, the

[1] A4Cloud www.a4cloud.eu

Information Commissioner's Office (ICO) launched a PIA process (incorporating privacy by design) to help organisations assess the impact of their operations on personal privacy. This process assesses the privacy requirements of new and existing systems; it is primarily intended for use in public sector risk management, but is increasingly seen to be of value to private sector businesses that process personal data. Similar methodologies exist and can have legal status in Australia, Canada and the US [45]. The methodology aims to combat the slow take-up to design in privacy protections from first principles at the enterprise level. Usage is increasingly being encouraged and even mandated in certain circumstances by regulators, as considered further in the following section.

The role of a risk-based approach in data protection has been considered by a number of parties, including: as an assessment of the relative values of such an approach [5]; modifying the original OECD data protection principles to take this into account [34]; analysing the relationship with accountability [21] and recent regulatory analysis [1, 9].

In terms of automation within the privacy impact assessment process, there are a few systems that have attempted this in various contexts, which we shall consider further below.

In Canada, the Treasury Board Secretariat provided in 2003 an e-learning tool for government employees interested in learning more about privacy and PIAs and how to complete them [35]. Furthermore, a new self-assessment tool, aimed at Small and Medium Enterprises (SMEs), was launched in Canada in May 2011. It was developed jointly by the Federal, Alberta and British Columbia privacy commissioners' and is a detailed online questionnaire that helps organisations gauge how well they are protecting personal information and meeting compliance standards under Canada's private-sector privacy law on both federal and provincial levels.

The US Department of Homeland Security (DHS) employs a PIA tool called the Privacy Threshold Analysis that helps users determine whether a PIA is required under the E-Government Act of 2002 and the Homeland Security Act 2002 [49]. In the UK, the PIA Guidelines provide a number of screening questions to help users decide whether a Full-Scale PIA or a Small-Scale PIA is warranted. The Guidelines also include a number of questions for a privacy law compliance check, and a Data Protection Act (1998) compliance check. Templates are also included within the Guidelines for Data Protection compliance and the Privacy and Electronic Communications Regulations (PECR) [26].

Most of these PIA tools are based upon a simple "decision-tree" approach and are mainly procedure-based with coarse-grained granularity, offered as Web applications that do not take into account the cloud or any of its characteristics. The following are PIA automated systems that are worthy of particular mention:

- A prototype decision support tool developed by the PRAIS project [24]. This tool enables personnel working with personal information to assess the privacy implications of information sharing actions dynamically and to share information and manage users' consent and other participant needs.
- HP Privacy Advisor (HP PA). It assesses risk and degree of compliance for projects that handle personal data and guides employees in their decisions on how to handle

different types of data. HP PA uses a rule-based system to capture global privacy knowledge that is too complex to be easily captured via decision trees and to dynamically only present the relevant question to elicit privacy-relevant information about a project to the user [37, 38, 39].

- A privacy impact assessment tool prototype based upon ICO guidelines related to UK Data protection Act, allowing appropriate stakeholder views and input and using confidences within the knowledge representation to allow assessment of the value of the input as well as customisation of risk indicator values [44].
- Avepoint Privacy Impact Assessment System [4] and TRUSTe Assessment Manager [48] help to automate the impact assessment workflow and to track the tasks involved in the question answering process by the multiple organisational roles. However they do not focus on cloud services, which intrinsically involve third parties and data transfers.

Decision support systems for PIAs in cloud computing are a new field and there are few systems available, although there is some work targeted at the areas of clinical decision applications, and life science enterprise solutions [7]. Prior work includes tools for cloud assessment: the Microsoft “Security Assessment Tool” designed to help find weaknesses in an IT security environment, privacy impact assessment of cloud environments [46] and decision support tools for cloud service provisioning [40]. In addition, several standards propose cloud security guidance: European Network and Information Security Agency (ENISA) [19], National Institute of Standards and Technology (NIST) [32], ICO [28] and Commission Nationale de L'informatique et des Libertés (CNIL) [13], CSA Governance Risk and Compliance (GRC) stack [12].

In the next sections we explain how our DPIAT builds on the body of knowledge and recommended practices mentioned above, adjusting the DPIA process and questionnaire to make it informative, user-centric and synthetic. It differs from previous work by focusing on a profile of SMEs wishing to move to the cloud. Additionally, our approach for assessing cloud risks is founded on information disclosed voluntarily by CSPs in the CSA Security, Trust & Assurance Registry (STAR).[2]

3 Multidisciplinary Approach to DPIAs

The proposed GPDR provides for a series of accountability measures that aim to strengthen protection of personal data. DPIAs fall under the scope of those measures, aiming at mitigating risks resulting from certain processing operations. In practice, a DPIA screening consists of a set of questions allowing for multiple choice or free text answers, which help to assess the risks for personal data involved in the intended processing. Taking this into account, as well as the various examples of existing PIAs, this section proposes a DPIA questionnaire that is tailored to particular data protection risks associated with cloud computing services.

[2] https://cloudsecurityalliance.org/star/

The DPIA tool incorporates two questionnaires. The first questionnaire (See Table 1) is a pre-screening (risk) assessment, which must be carried out to assess whether a full-scale DPIA is mandatory. The main questionnaire (See Table 2) is an extensive set of questions that comprises the full-scale DPIA [22].[3] The user of the DPIA tool will probably not be an expert in privacy and data protection. Therefore, the questions are formulated in a form and language understandable for lay users, in order to facilitate them in providing the right information [42]. We have targeted the DPIA tool to SMEs that typically lack in-house data protection experts and the resources to hire experts.[4] The tool thus should guide the user through the process as much as possible and provide meaningful feedback that helps the user to improve the privacy characteristics of their project and facilitate legal compliance with the data protection regulation.

3.1 Methodology

Cloud computing has several characteristics [30] that may adversely impact the privacy of personal data, including distributed nature, multitenancy, third-party hosting, potentially long supply chains. A cloud can be spread across multiple jurisdictions with different degrees of data protection and no transparency about this [19]. The multitenancy leads to risks of isolation failure and insecure data deletion which can compromise personal data. Third-party hosting can cause the cloud consumer to lose control over personal data, especially when the CSPs are not transparent about the data processing performed, the data protection measures used and the data security breaches that occurred [19]. This becomes even more apparent in the case of complex supply chains formed from different CSPs. When developing the DPIA questionnaire (see section 3.3) and the cloud adoption risk assessment model (see section 5) we considered these cloud characteristics and their impact on data protection.

Given that the current data protection framework within the EU is under review and that the proposed GPDR still is under extensive negotiation at the time of writing[5], we had to decide whether the questionnaires would take into account the new DPIA framework proposed within the GDPR. Following discussions within the A4Cloud consortium, all partners agreed that the DPIA tool should be as future proof[6] as possible, and

[3] Note that even if the full-scale DPIA is not required, taking it nevertheless is beneficial because the questionnaire, guiding responses and assessment may help in raising the privacy bar of any project or service.

[4] A secondary user group consists of concerned individuals who consider taking their data to the cloud. The tool will help them make considered choices regarding requirements for cloud service providers. A sister tool in the A4Cloud project, the Cloud Offerings Assistance Tool (COAT) can take these requirements to filter relevant cloud offerings for the user to choose from.

[5] Both the European Parliament and the Council have agreed on their texts amending Commission's initial proposal on a GDPR. Although, there is broad agreement between the institutions on core issues, the exact wording is to be decided –probably by the end of 2015- following a series of Trilogue Meetings.

[6] For more on the concept of “future-proof” see under section 3.5: Discussion.

therefore we took into account both the Data Protection Directive (DPD) [16], as it is still the main legal instrument within the EU, and the drafts of the upcoming GDPR[7], rather than focusing exclusively on the legislation currently in force. The aim we set was to develop a tool that could be used effectively under both regimes.

The DPD provided us with the basic concepts and principles defining the current general data protection framework, while the GDPR provided additional concepts and concrete procedural guidelines for a practical DPIA questionnaire. In particular, the principles relating to processing of personal data, such as purpose limitation and data minimisation, derived from the DPD. Articles 6 and 7 of the current DPD, which deal with the legitimacy of data processing, gave grounds for an extensive set of questions aimed at mapping the user's intention to the legal terms incorporated in the DPD[8]. Furthermore, ICO's "Code of Practice: [27], in conjunction with the PIA Guide of the Office of the Australian Information Commissioner (OAIC) [3] also proved to be useful tools in phrasing particular questions[9]. The ICO's PIA Handbook [26] constituted the key inspirational instrument in drafting the questions related to the grounds of processing.

The GDPR (in the form of the European Parliament's first reading), was used as the starting point for both questionnaires. Articles 32a and 33 provide the conditions under which a DPIA would be mandatory.

The analysis of the DPD, GDPR, and various DPIA and PIA [52, 10, 51, 54] models are reflected in the construction of the questionnaire's framework[10]: the legal norms and the PIA/DPIA models utilised[11] allowed us to develop the "Question" field (for the related "Explanation" one see Tables 1 and 2), while the sources for risks in cloud environments [11, 12, 19] were used to give a logical structure to the questionnaire and to weigh the answers provided by users. The "Answer" fields were developed to steer the user throughout the questionnaire according to a logic order that was formulated mainly through the examination of the DPD and the GDPR, while assessing the impact and the likelihood of an unwarranted event happening.

Many PIAs work on the assumption that the user is aware of certain basic data protection notions, such as 'personal data' and directly ask the user whether they process

[7] Which will arguably embody the current state of the art in data protection legislation, as well as the result of the doctrinal elaboration the concept had in the last two decades.

[8] For instance, Question 10 in Table 2 ("Are all the information and its subsets you handle necessary to fulfill the purposes of your project?") or Question 17 ("Does your project involve the use of existing personal information for new purposes?") were drafted by taking into consideration the already existing legal requirements.

[9] For instance, Question 11 in Table 2 ("Is it possible for the individual to restrict the purposes for which you process the information?").

[10] The table we developed is composed by the following categories: question, explanation of the question, question type (which frames the possible answers to be given by the users, e.g. in the form of radio buttons, checkboxes, or yes/no binary answers), responses to be given to the users in order to educate them while they go through the questionnaire, actions to be performed by the tool as a consequence of the users' answers (e.g. go to the next question). A weighing of the users' activities' impact on data subjects' privacy and data protection was originally embedded in the table as well.

[11] See *supra* note 4.

personal data and for which purposes and on what ground and so forth. Our DPIA starts from the premise that the user does not know these concepts and it therefore tries to, within limits, do a legal qualification of the user's responses to simple terms. Based on the kind of information the user intends to process, the tool will 'decide' that it constitutes personal data, rather than having the user specify so in advance. The tool does provide feedback incorporating proper legal terminology where applicable.

The risk assessment, which provides the basis for probing the user about mitigation measures, is based on a series of documents (see section 3.5 below) regarding the most commonly occurring incidents in cloud ecosystems; from a data protection viewpoint, these incidents provided valuable insights on the cloud's potential threats to informational self-determination, on their likelihood and on their foreseeable impact. We conceived risk as the by-product of the interplay between the likelihood of an event and of the impact that event would have. We based the construction of the questionnaire on that conception, which is to say we used literature and reports to investigate, on the one hand, the most harmful privacy-related incidents, and on the other the most likely ones, all in order to develop a better understanding of what to ask when assessing the impact of an undertaking's activities on data subjects' privacy and data protection rights. Since the questionnaire aims to assess, *grosso modo*, how and how much a cloud user's undertaking deviates or could deviate from the physiology dictated by data protection norms (as embodied currently in the DPD and for the future in the GDPR), and the impact of its activities on data subjects, it seemed proper to consider, amongst other prominent factors, the most likely and/or the most harmful incidents in cloud environments. Based on these considerations, we formulated questions embracing the notions of risk and likelihood in an intelligible manner for the tool user; for instance, the incorporation of the question: "How severe do you deem the consequences, in case you process outdated information for the individuals it refers to?" forms a clear example on tool's underlying perception on the notion of impact, while a question such as "For how long do you store the information you are dealing with?" captures the related perception on the notion of likelihood[12]. The situations that are most likely to threaten individuals in the cloud or that, if they occur, would harm individuals the most, provided a useful list of the risks to be incorporated in the tool. Determining their impact and likelihood turned out not to be straightforward, though. Due to the lack of available and sufficiently targeted metrics, the likelihood parameter was inferred through the review of several documents issued by public bodies tasked with the safeguard of the rights to privacy and data protection or dealing with information security, for instance [13, 14, 20, 31] among others. The impact parameter, on the other hand, is historically hard to define when correlated to the notions of privacy and, albeit to a minor extent, to the one of data protection: as it has been noted by prominent doctrinal sources, they appear "*to be about everything, and therefore [...] to be nothing*" [43]. Moreover, harms deriving

[12] Based on the intuition that the longer data is stored, the higher the likelihood that something happens to the data. Of course this is not necessarily, or always, the case, but as a heuristic it may suffice to make the user think about data retention.

from privacy and data protection violations are hardly quantifiable in that they are inherently linked to other rights, whose infringement causes the starkest impact on data subjects [43] – "*a cluster of related activities that impinge upon people in related ways*"[13]. Hence, an ontological definition of the impact deriving from a data privacy violation appears to be hardly feasible in the tool's context[14], aside, of course, from what can be directly inferred from the relevant regulations. We have therefore made reasoned assumptions about potential impacts.

It is important to stress here that this process could not capture the whole of the relevant law, which is far too complex, lengthy and granular to be represented in the tool. Qualitative decisions had to be made about which legal norms should be included, and at what level of detail. In addition, framing the questions and devising explanations of their meaning lost even more detail and richness of meaning. The version of the legal norms embodied in the tool is thus only a partial summary of the law's requirements in this area, shaped to the needs of the tool. This means that the tool cannot be relied on to identify all potentially applicable legal obligations, and that its risk assessment outputs are by definition not fully comprehensive.[15]

Despite the existence of several PIA/DPIA models which deal with traditional cases of processing, there is hardly a sufficient number of cloud-tailored DPIA models, especially when considering the growing importance and pervasiveness of the cloud computing model in the market and the differences that run between traditional IT environments and the cloud. ENISA's recommendations [19] constituted, though, a helpful methodological tool in identifying and evaluating risks on the data protection rights. Also, ENISA's framework for Cloud Security Incident Reporting [20] formed the key element for the development of the evaluation scheme we propose. Several other scholarly publications [31] have been consulted for targeted guidance on particular topics in order to articulate cloud-relevant questions[16].

3.2 The Pre-Assessment Stage

The pre-assessment stage includes a set of seven questions, fully presented in Table 1. It aims to identify whether the processing operations to be undertaken can be perceived as potentially risky to the protection of personal data of the individuals and as such

[13] A gross negligence in an anonymization process giving ability to unduly infer a data subject's identity, for instance, which is usually a data protection violation per se, can lead to a diverse array of consequences (such as identity theft, physical harm – e.g. domestic violence victims tracked down by their assailants) depending on the concrete circumstances of the case.

[14] Our consideration of the impact deriving from privacy and data protection violations, however, was largely shaped according to Solove's classification (*Ibid.*), which taxonomizes privacy violations according to four macro-categories (Information collection, information processing, information dissemination, intrusion), each of which can be subdivided into more specific subcategories.

[15] The user may notice while going through the tool that their situation is not satisfactory covered by the questions. This may be a clear indicator to seek professional help to supplement the tool's assessment.

[16] Questions 48-50 in Table 2 refer to the service models in a cloud environment.

trigger the full-scale DPIA when this is the case. It initially assesses whether the information s/he deals with constitutes personal data or not, and then evaluates the kind of information processed, its sensitivity, the purposes of the processing, the actors involved and the extent to which the information is likely to be diffused. Our purpose was mainly to provide the user with a very short and incisive quick-scan to assess the presence or the absence of some general factors that indicate the use of personal information, e.g. the very qualification as personal data of the information dealt with by the tool's user, or the presence of sensitive data amongst it.

3.3 The Assessment Stage

The (conditionally) following full-scale DPIA includes 50 questions (see Table 2 for an excerpt and [22] for the full version including explanation of implication of each answer option). The questions are grouped into to five (5) topical areas (the key inspirational document which enabled the taxonomy of these topical areas was [33]), which refer to: 1) the type of project, 2) the collection and use of data, 3) the project's storage and security policies, 4) data transfers, and 5) cloud specific issues. The aim of this set of questions is to assess how the interactions between the subjects that perform the DPIA and CSPs affects data subjects' rights to privacy and data protection.

Each question has several possible suggested answers (single selection or multi-choice), avoiding open questions, which are hard to process automatically. While answering some questions the user can get guidance from the DPIAT (see section 5) on how to address the privacy issues related to the specific answers. In particular, questions 35 and 39 cover respectively a set of privacy and security controls supporting data protection; this helps the user document existing controls and to understand which others could be implemented.

3.4 Evaluation of the Results

Each question has a formula for computing the privacy impact score based on its answer and a weight prioritising the importance relative to other questions. For example, the Question 4 in Table 2 "*Are you relying exclusively on consent in order to process information of individuals?*" has the following possible answers:

a) Consent is given directly by the individual by a statement (e.g. by a consent form)

b) Consent is given directly by the individual by an affirmative action (e.g. by ticking a box)

c) Consent has been obtained implicitly by the individual (e.g. by merely use of the service or inactivity)

We assign the value for the *privacy impact score* for the answer to this question using the following formula: *If option 'a' then the score is 0, Else if option 'b' then the score is 1/4, Else if option 'c' then the score is ¾.*

Intuitively, the option 'c' would have a bigger impact on privacy than option 'b' and 'a' so the score is chosen to be proportional to the perceived impact. We compute the *final privacy impact score* (FI) taking into account the answers to all the questions:

$$FI = \frac{\sum_i^N s_i \alpha_i}{\sum_i^N \alpha_i} \ (1)$$

Here *N* is the number of questions in the DPIA questionnaire, s_i is the score for the answer to the question *i*; and $\alpha_i = 1$ if the question *i* is answered and $\alpha_i = 0$ otherwise.

In addition, we associate the questions with several *privacy indicators*, capturing different privacy aspects: *data sensitivity*, *compliance*, *trans-border data flow*, *transparency*, *data control*, *security*, and *data sharing*. For example, the answer to the question above influences the *data control* and *transparency* indicators. Some of the indicators can enhance privacy (*compliance, transparency*, *data control* and *security*), while the others diminish it (*data sensitivity, trans-border data flow and data sharing*). Therefore, the privacy indicator scores will be either proportional to the privacy impact scores of individual answers or inverse. So in the example above a higher score for the answer (option 'c') implies less data control and transparency.

We compute the *final privacy indicator score* for the indicator *j* (FI_j):

$$FI_j = \frac{\sum_i s'_{ij} \alpha_i \beta_{ij}}{\sum_i \alpha_i \, \beta_{ij}} \ (2)$$

Here $s'_{ij} = s_i$ if the indicator j negatively affects privacy and $s'_{ij} = 1 - s_i$ otherwise; β_{ij}= 1 if the answer to question i impacts indicator j and β_{ij}= 0 otherwise. The ratio $\sum_i^N \alpha_i / N$ represents the coverage of the questionnaire and indicates the reliability of the indicators.

Finally, we define the overall *privacy impact level* and *privacy indicator levels* for the assessment by translating correspondingly *FI* and $FI_j$to a uniform qualitative scale: *Low* < *Medium* < *High* and use color-coding to facilitate the presentation: *Low* → Green, *Medium* → Yellow and *High* → Red.

In order to provide users with actionable guidelines, the DPIAT final report contains an additional section that delivers textual guidance generated according to the user's answers. Far from being considerable as legal advice – as the tool specifically disclaims – the section is still able to make the tool's user focus on specific privacy and data protection-related issues s/he might have overlooked. For instance, when a user indicates that data protection is not considered from the outset of the assessed project's development, the section highlights the importance of the concepts of Data Protection by Design and Data Protection by Default.

3.5 Discussion

Under the GDPR, as amended by the outcome of the European Parliament's first reading, there is a trend to make DPIAs compulsory when the processing operations of controllers are likely to present specific risks for rights and freedoms of data subjects (Article 32a of the Parliament's text Respect to Risk). This approach seems to confirm the importance of DPIAs to protect data subjects' rights and freedoms: this meant for us embedding in the DPIA process the concept of risk analysis introduced in the earlier stated Article 32a of the European Parliament's amended text.

As to the first area of questions relating to the type of project undertaken by the tool's user, our aim was to frame both the kind of activity performed by the CSP's client and the aim of that activity. We considered the fact that a controller could handle personal data (for instance, the controller may obtain information such as the name and e-mail address of users through online subscription forms) for a number of different reasons and aims (e.g. commercial purposes) Therefore, we decided to include two separate inquiries: one regarding the activities through which data is processed, and another regarding the purpose of the processing.

The second area of questions regards the collection of the information, the usage that processors make of that information and the means with which personal data is handled. This section draws heavily from the basic principles of both the DPD and the GDPR. For instance, it attempts to discover whether there appear to be solid, legitimate grounds for processing, identify the main risks of non-compliance with the data protection principles and assess the tool user's plans for compliance with the rights of the data subject sanctioned by law.

Storage and Security (deletion included[17]), moreover, is considered a third area, which deserves specific consideration, especially in relation to the traits of Cloud Computing.

The investigation we propose was developed according to an "individual-centric approach", which tried to deepen the level of protection accorded to data subjects, irrespective of who (either CSPs or their customers) exerts concrete control over the particular aspect considered: that is to say, we considered it more useful to ask SME users (and individuals using the tool) questions pertaining to the CSPs' areas of control[18], accepting the chance they might not know the answer to our inquiry, in which case the user simply refrains from answering. Leaving questions open provides a less 'accurate' assessment, but still provides guidance. Users can also return to the questionnaire after obtaining answers to questions they cannot answer from others to provide a more complete picture. The tool thus is not a one-way street, but can be used iteratively.

A major concern we had related to the "updatedness" of the information dealt with by the tool user. The questionnaire includes two questions regarding the foreseen negative consequences of the outdated information processed by the tool user's undertaking; specifically the questionnaire addresses the consequences of outdated information about individuals[19] and how such outdated information can lead to regulatory liability[20]. Whether or not outdated information may result in civil or criminal liability, however, is outside the scope of the DPIA. An individual-centric approach has also been adopted

[17] Note that deletion assumes particular importance in the cloud: the remoteness of the physical machines and the lack of control cloud users have over them, considered in relation to the fact that several different layers of deletion exist (from a mere drag-and-drop in the OS' virtual rubbish bin to the physical destruction of the hardware in which the virtual machine of the user lies), make deletion a focal point when assessing the risks a data subject is prone to.

[18] E.g. Question 47 in Table 2

[19] See question 28 in Table 2

[20] See question 29 in Table 2

for the fourth set of questions, which relates to the transfer of information. This is because transferring information is controlled by the law to attempt to limit the risks that the data subjects are subject to by prescribing conditions for data transfer. Furthermore, due to the target of the DPIA tool, this class of inquiries caters for the possibility that the tool's user does not possess an adequate level of knowledge to answer all questions. Much like with the third set of questions, we considered the possibility of a lack of answer appropriate.

The final set of questions refers exclusively to cloud computing services. Given the complexities of cloud computing technology, it was a challenge to formulate those questions in an understandable language for an ordinary user. Each deployment model has various ramifications which are not necessarily known in the first place to the user of the DPIA tool who is to decide whether to opt for a particular cloud computing service or not.

It is important for the users of a cloud service to know how to secure the information they process within the cloud environment. Taking that into account, the cloud relevant questions aim at ascertaining the level of exposure to risk that the user may have by virtue of using a specific type of cloud service. Two major aspects are important to establish in this regard. Firstly, it is important to know whether the cloud service used by the user of the DPIA tool is public, and thus shared with third parties, or private, and thus solely used by the user. Secondly, it is important to establish what the user utilises the cloud service for[21].

The inclusion of a specific part of the questionnaire targeted only to the cloud environment serves as an enabler for the applicability of the DPIA tool to a non-cloud setting as well, in an attempt to ensure that the DPIA Questionnaire remains future proof so far as technological change is concerned. This technology neutral approach enables the application of the tool to future Internet services. If the cloud-relevant questions are removed, the questionnaire can potentially be used to assist in achieving compliance with the legal framework irrespective of whether the assessed undertaking operates in the cloud or not.

Future proofing the tool in terms of its legal content is more problematic. Even once the GDPR has been agreed and becomes law, the content of the law will not be static because laws are regularly amended. More challenging is that the *meaning* of legal provisions develops and changes over time, in response to court decisions about specific sets of facts and policy decisions and guidance issued by regulators. For this reason a mechanism will need to be developed to review and update the legal content of the tool at appropriate intervals to ensure that it does not become dangerously inaccurate.

4 Cloud Adoption Risk Assessment Model

We employ the Cloud Adoption Risk Assessment Model (CARAM) to evaluate the risks resulting from adoption of cloud services (see [8] for full details). CARAM is

[21] See Questions 48-50 in Table 2

designed to assist (potential) cloud customers assess all kinds of risks—not only privacy-related—that they face by selecting a specific CSP. The results of CARAM risk assessment constitute a part of the DPIA report (see section 5).

CARAM is a qualitative deductive risk assessment model based on ENISA's cloud risk assessment model [19] and the Cloud Security Alliance's (CSA) Cloud Assessment Initiative Questionnaire (CAIQ)[22]. Like in [19] we conceived risk as the by-product of the interplay between the likelihood of an event and of the impact that event would have. CARAM complements ENISA's approach to take into account cloud customers' assets (modelled based on the list of assets from the ENISA report) and the implementation status of security controls in CSA STAR public registry to perform a relative risk assessment of (potential) cloud solutions. This can help cloud consumers to determine which CSPs have acceptable risk profiles for security, privacy, and quality of service.

Most of the entries in STAR use a template that provides 148 questions grouped into several control areas covering the state of implementation of various security controls. We have categorised the answers of more than 50% of the CSPs from the STAR—including several big players—into the following categories:

- *Implemented*: the control is in place
- *Conditionally Implemented*: the control can be implemented under some conditions
- *Not Implemented*: the control is not in place
- *Not Applicable*: the control is not applicable to the provided service

Since the answers were given in a verbose free text form instead of simple Yes/No and the number of answers was big (circa 9000) we used supervised machine learning algorithms provided by the WEKA tool [25] to automate this classification.

We used these answers together with other information from the ENISA report to calculate the vulnerability index for different risk scenarios (see Table 3 for the list of risk scenarios). The vulnerability index is defined to be proportional to the number of implemented security controls that mitigate vulnerabilities involved in the risk scenario. It is later used to adjust the probability of the risk scenario using the values provided by the experts from the ENISA report as a baseline. Eventually, the risks are grouped into three categories: risks for security, privacy and service: to provide a high level risk profile which is easier to interpret. Based on these results the customers can compare different cloud solutions and select those satisfying their risk tolerance.

Fig. 1 displays the level of exposure (vulnerability index) for privacy risks among the analysed CSPs (similarly, the vulnerability index can be computed for security and service risks). According to these results, the lowest vulnerability index for a cloud solution is 0.011 while the vulnerability index for the highest risk cloud solution is 0.491. Although the later index is more than 44 times higher than the former, it is still less than 0.5. This means that the likelihood value for even the highest risk cloud solution in STAR will be reduced significantly, and become "LOW" according to the risk matrix from [8]. This is expected since all analysed CSPs report that they have implemented at least 70% of the controls from CAIQ.

[22] https://cloudsecurityalliance.org/research/cai/

In this approach, we rely on the self-assessment provided by the CSPs since it is not possible to verify independently the status of each control: only three of the analysed CSPs had a third party certification from CSA when we performed the data collection. Certification report details are not available to the public.

5 DPIA Tool and Report

DPIAT's web interface enables an easy and user friendly experience of a questionnaire about a perceived complex issue. Screenshots are shown in Fig. 2 and Fig. 3. The landing page asks the user whether they would like to start with pre-screening questions to determine if they need to answer the full-scale questionnaire (screening questions). The full-scale assessment questionnaire (see section 3.3) contains a set of a bit more than 50 questions displayed in five stages categorising them. The stages are Type of Project, Collection and use of information, Storage and security, Transfer of information, and Cloud specific questions. During the completion of the questionnaire, the user is provided feedback on the answers and choices they make. This includes, for instance, pointing out that the chosen option increases the privacy risk, thus subtly suggesting the user to reconsider their choice. The tool does not judge, but is rather aimed at stimulating the user to think about their project from the perspective of privacy and data protection.

The output is a report including the data protection risk profile, assistance in deciding whether to proceed or not, and suggested mitigations. The report contains three sections. The first, "*Risk Related to Your Proposed Application*", is based on the answers to the questionnaire and contains the overall data protection impact score and several privacy indicator scores (see section 3.4) namely, risks related to Sensitivity, Compliance, Trans-border Data Flow, Transparency, Data Control, Security, and Data Sharing (see Fig. 4). The second part, "*Risk Related to the selected Cloud Provider*", displays the risks based on the security controls used by the CSP (see section 4). It contains the 35 ENISA [19] risk scenarios with their associated scores. The last section contains additional information related to the GDPR article 33. It also explains to the user that DPIA is meant to be an ongoing process and guides the user on the general phases of the assessment. The final decision of whether to proceed with the desired transaction (which triggered the DPIA in the first place) is up to the user or his manager (i.e. an approver in case the result of the DPIA is *high risks*).

The implementation of the *server-side* application and web-service (Questionnaire Provider) is written in Java. This application provides access to the Questionnaire data and also provides a rules-engine that helps determine the flow of the questionnaire for the client as well as providing further details and information based on the user's responses to the questions offered. The *rules engine* is based on the Drools [23] library. The

[23] Drools Business Rules Management System Solution: http://www.drools.org/

client-side application is implemented using HTML5 and JavaScript and utilises a number of open-source libraries to simplify the underlying business logic layer. We use RESTful[24]API as a transport layer and JSON [25]as the data-interchange format.

During the development of the tool, testing on how the user experience should look like was conducted. The tool was presented to several users including partners in HP Privacy Office and the feedback received was incorporated in the final implementation of the tool. Positive feedback was given on the amount of guidance provided for the user in terms of information text for both the questions and the answers. Also, dividing the 50 questions into five stages was considered a good impact on the tool's usability. Additional testing was carried out with privacy researchers from a variety of interdisciplinary backgrounds, and further changes are planned to the tool in respect of this feedback. In particular, there was a strong perceived need for more explanation about both how the tool derives its recommendations and about how these recommendations should be interpreted and acted upon.

6 Conclusions

We have presented a contemporary Data Protection Impact Assessment methodology focusing on the use of cloud services, supported by a tool that aims at helping users to understand privacy risks of their intended project and help them consider means to mitigate these concerns. The DPIAT is based on existing PIAs, legal sources and specific cloud risk scenarios. It is aimed specifically at SME users that typically have limited knowledge about privacy and data protection and have restricted resources to consult experts in the field, yet will have a legal obligation (once the GDPR comes into effect) to conduct a DPIA. Although the tool does not incorporate advanced intelligence to help the user, we believe that the way we have structured the issues, framed the questions and provide situation specific feedback and a crude likelihood/impact score, actually will help the target audience understand the importance of privacy and data protection in their context and help improve legal compliance.

7 Acknowledgement

This work is conducted as part of the EU-funded FP7 project grant number 317550 titled as "Accountability for Cloud and Other Future Internet Services" (A4Cloud - http://www.a4cloud.eu/).

[24] RESTful is a standard for web APIs and transport protocol

[25] JSON Data Interchange Format: http://www.json.org/

8 References

1. Article 29 Data Protection Working Party: Statement on the role of a risk-based approach in data protection legal frameworks (WP218), May. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf. (2014)
2. Article 29 Data Protection Working Party, Opinion 05/2012 on Cloud Computing. EU; 2012 p. 1–27. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf (2012)
3. Australian Government, Office of the Australian Information Commissioner: Privacy Impact Assessment Guide (OAIC) (2010)
4. Avepoint: Avepoint Privacy Impact Assessment (APIA) System. https://privacyassociation.org/resources/apia (2015)
5. Bennett, C.J. and Raab, C.D.: The Governance of Privacy: Policy Instruments in Global Perspective. MIT Press, Cambridge, Massachusetts (2006)
6. Bernsmed, K., Felici, M., Santana De Oliveira, A., Sendor, J., Brede Moe, N., Rübsamen, T., Hasnain, B.:. Use Case Descriptions (Deliverable No. D:B-3.1) p. 68 A4Cloud. (2013)
7. CambridgeSoft: ChemBioOffice Cloud–An Integrated Decision Support System for CHDI. http://chembionews.cambridgesoft.com/WhitePapers/Default.aspx?whitePaperID=43 (2010)
8. Cayirci, E.; Garaga, A.; Santana de Oliveira, A.; Roudier, Y.: A Cloud Adoption Risk Assessment Model. Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference, pp.908-913 (2014)
9. Centre for Information Policy Leadership (CIPL): A Risk-based Approach to Privacy: Improving Effectiveness in Practice. http://www.hunton.com/files/upload/Post-Paris_Risk_Paper_June_2014.pdf (2014)
10. Clarke, R.: Privacy impact assessment: Its origins and development. Computer law & security review25.2,: 123-135 (2009)
11. Cloud Security Alliance (CSA): Security guidance for critical areas of focus in cloud computing, v3.0. http://www.cloudsecurityalliance.org/guidance/ (2011)
12. Cloud Security Alliance (CSA): The notorious nine: Cloud computing top threats in 2013, v.1.0. http://cloudsecurityalliance.org/research/top-threats/ (2013)
13. Commission Nationale de L'informatique et des Libertés (CNIL): Recommendations for Companies Planning to Use Cloud Computing Services. http://www.cnil.fr/fileadmin/documents/en/Recommendations_for_companies_planning_to_use_Cloud_computing_services.pdf (2012)
14. Commission Nationale de L'informatique et des Libertés (CNIL): Methodology for Privacy Risk Management (2012)
15. COM 11 final 2012/0011 (COD) European Commission: Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Brussels, 25.1.2012 p. 1. (2012)
16. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data OJ L281/31 (DPD) (1995)
17. De Hert, P.: A human rights perspective on Privacy and Data Protection Impact Assessment. Privacy Impact Assessment. Law, Governance and Technology Series. Volume 6, pp 33-76, Springer (2012)

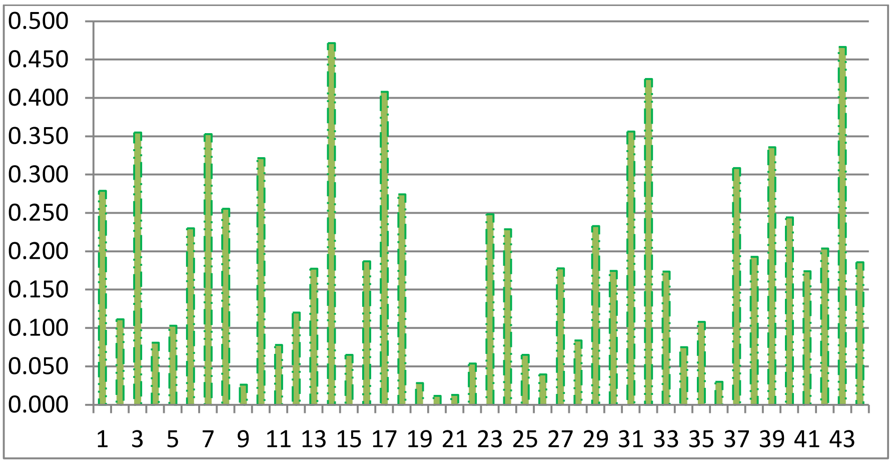
18. De Oliveira A.S, Garaga A., Martucci L. A. , Felici M., Alnemr R., Stefanatou D. , Niezen M., Fernandez C., Nuñez D., Hasnain B. , Vranaki A. and Cayirci E.: D:C-6.1: Risk and trust models for accountability in the cloud. Deliverable. A4Cloud project (2013)
19. European Union Agency for Network and Information Security - European Network and Information Security Agency. Cloud Computing - Benefits, risks and recommendations for information security (2009)
20. European Network and Information Security Agency: Cloud Security Incident Reporting: Framework for reporting about major cloud security incidents, ENISA (2013)
21. Felici, M. and Pearson, S.: Accountability, Risk, and Trust in Cloud Services: Towards an Accountability-Based Approach to Risk and Trust Governance. Proc. SERVICES, IEEE, pp. 105-112 (2014)
22. Garaga A., Santana de Oliveira A., Cayirci E., Dalla Corte L., Leenes R., Mhungu R., Stefanatou D., Tetrimida K., Alnemr R., Felici M., Pearson S., Vranaki A..: D:C-6.2 Prototype for the data protection impact assessment tool. A4Cloud Deliverable D36.2. http://www.a4cloud.eu/sites/default/files/D36.2%20Prototype%20for%20the%20data%20protection%20impact%20assessment%20tool.pdf (2014)
23. Habib, S.: A Trust-aware Framework for Evaluating Security Controls of Service Providers in Cloud Marketplaces. CCSW, 459–468. doi:10.1109/TrustCom.2013.58 (2013)
24. Harbird, R., Ahmed, M., Finkelstein, A., McKinney, E., Burroughs, A.: Privacy Impact Assessment with PRAIS. http://www.cs.ucl.ac.uk/staff/A.Finkelstein/papers/hotpets.pdf (2007)
25. Hall, M. *et al*: The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue (2009)
26. Information Commissioner’s Office: Privacy Impact Assessment Handbook, http://ico.org.uk/pia_handbook_html_v2/files/PIAhandbookV2.pdf (2011)
27. Information Commissioner’s Office: Conducting privacy impact assessments code of practice. https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf (2014)
28. Information Commissioner's Office: Guidance for Companies on the Use of Cloud Computing, v1.1 http://ico.org.uk/for_organisations/data_protection/topic_guides/online/cloud_computing (2012)
29. Marnau, N.: D1.2.4 Cloud Computing – Data Protection Impact Assessment, Deliverable, TClouds project (2010)
30. Mell P, Grance T. The NIST definition of cloud computing, NIST Special Publication 800 (2011)
31. Millard, C. J., ed.: Cloud Computing Law, Oxford University Press (2013)
32. National Institute of Standards and Technology NIST: Guidelines on Security and Privacy in Public Cloud Computing, SP 800-144. http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf (2011)
33. NOREA: Privacy Impact Assessment: Introductie, handreiking en vragenlijst. beroepsorganisatie van IT-auditors available at http://www.norea.nl/readfile.aspx?ContentID=36650&ObjectID=343968&Type=1&File=0000040117_NOREA%20A4%20Privacy%20Impact%20Assessment%2003%20WEB.pdf (2013)
34. Organisation for Economic Co-operation and Development OECD: Guidelines Concerning the Protection of Privacy and Transborder Flows of Personal Data http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf (2013)

35. Office of the Privacy Commissioner of Canada: Securing Personal Information: A Self-Assessment Tool for Organisations. http://www.priv.gc.ca/resource/tool-outil/security-securite/english/AssessRisks.asp?x=1 (2011)
36. Office of the Privacy Commissioner of Canada (OPCD): Privacy Impact Assessments. http://www.priv.gc.ca/resource/fs-fi/02 05 d 33 e.asp (2011)
37. Pearson, S: Simple Mode: Addressing Knowledge Engineering Complexity in a Privacy Expert System, HP Labs External Technical Report, HPL-2010-75, June. Available via http://www.hpl.hp.com/techreports/2010/HPL-2010-75.html (2010)
38. Pearson, S. and Sander, T.: A Decision Support System for Privacy Compliance. In: Threats, Countermeasures, and Advances in Applied Information Security, Manish Gupta, John Walp, and Raj Sharman (eds.), Information Science Reference, IGI Global, New York, pp. 158-180. (2012)
39. Pearson, S., Rao, P., Sander, T., Parry, A., Paull, A., Patruni, S., Dandamudi-Ratnakar, V., Sharma, P.: Scalable, accountable privacy management for large organizations. Enterprise Distributed Object Computing Conference Workshops, EDOCW 2009. 13th , vol., no., pp.168-175 (2009)
40. Sander, T. and Pearson, S.: Decision Support for Selection of Cloud Service Providers. International Journal on Computing (JoC), GTSF, vol.1, no. 1, pp. 106-113, August. (2010)
41. SEC 72 final, Commission Staff Working Paper: Impact Assessment Accompanying the document Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) and Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data. Brussels, 25.1.2012, p. 81 available at http://ec.europa.eu/justice/data-protection/document/review2012/sec_2012_72_en.pdf. (2012)
42. Svantesson, D., and Clarke, R: Privacy and consumer risks in cloud computing, Computer Law & Security Review 26.4 p.392 (2010)
43. Solove, D.: A taxonomy of privacy, University of Pennsylvania law review. p 479. (2006)
44. Tancock, D., Pearson S., Charlesworth. A.: The emergence of privacy impact assessments. Internet: http://www.hpl.hp.com/techreports/2010/HPL-2010-63.pdf (2010)
45. Tancock, D., Pearson, S. and Charlesworth, A.: Analysis of Privacy Impact Assessments within Major Jurisdictions. Proc. PST 2010, Ottawa, Canada, IEEE, pp. 118-125 (2010)
46. Tancock, D., Pearson, S. and Charlesworth, A.: A Privacy Impact Assessment Tool for Cloud Computing. Privacy and Security for Cloud Computing, S. Pearson and G. Yee (eds.), Computer Communications and Networks, Springer, pp. 73-123 (2013)
47. Trilateral Research and Consulting: Privacy Impact Assessment and Risk Management. http://trilateralresearch.com/wp-content/uploads/2013/08/pia-and-risk-management-full-report-for-the-ico-1-released-6-Aug-20131.pdf (2013)
48. Truste: TRUSTe Assessment Manager. https://www.truste.com/resources?doc=516
49. United States Department of Homeland Security: Privacy Threshold Analysis (PTA). http://www.dhs.gov/xlibrary/assets/privacy/DHS_PTA_Template.pdf (2007)
50. Warren A., Bayley R., Bennett C., Charlesworth A., Clarke R., Oppenheim C.: Privacy Impact Assessments: International experience as a basis for UK Guidance. Computer Law & Security Review 24, no. 3: 233-242 (2008)
51. Wright, D.: The state of the art in privacy impact assessment. Computer law & security review 28.1: 54-61 (2012)

52. Wright, D., De Hert P.: Introduction to privacy impact assessment. Springer Netherlands,; Wright, David, et al. "Privacy, trust and policy-making: Challenges and responses", Computer law & security review p. 69-83 (2009)
53. Wright D., Wadhwa K., Lagazio M., Raab C., Charikane, E.: Integrating privacy impact assessment in risk management. International Data Privacy Law (2014)
54. Wright D.: Should Privacy Impact Assessments be Mandatory? Communications of the ACM, Vol. 54 No. 8, p. 121-131 (2012)

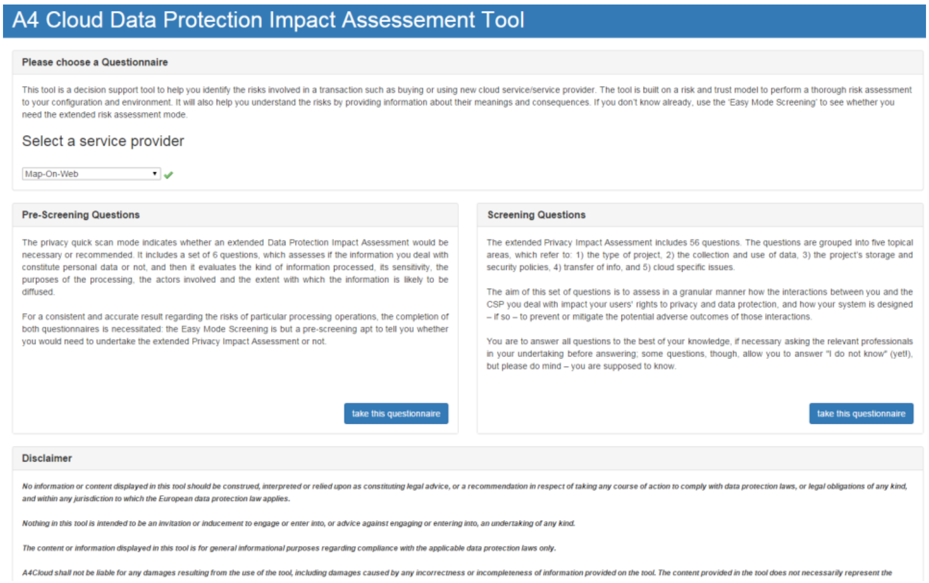
9 Appendix

Fig. 1. Privacy vulnerability index for 44 CSPs[26] in STAR



[26] The actual CSP names were omitted for confidentiality reasons

Fig. 2. DPIAT initial screen



A4 Cloud Data Protection Impact Assessement Tool

Please choose a Questionnaire

This tool is a decision support tool to help you identify the risks involved in a transaction such as buying or using new cloud service/service provider. The tool is built on a risk and trust model to perform a thorough risk assessment to your configuration and environment. It will also help you understand the risks by providing information about their meanings and consequences. If you don't know already, use the 'Easy Mode Screening' to see whether you need the extended risk assessment mode.

Select a service provider

Map-On-Web ▾ ✓

Pre-Screening Questions

The privacy quick scan mode indicates whether an extended Data Protection Impact Assessment would be necessary or recommended. It includes a set of 6 questions, which assesses if the information you deal with constitute personal data or not, and then it evaluates the kind of information processed, its sensitivity, the purposes of the processing, the actors involved and the extent with which the information is likely to be diffused.

For a consistent and accurate result regarding the risks of particular processing operations, the completion of both questionnaires is necessitated: the Easy Mode Screening is but a pre-screening apt to tell you whether you would need to undertake the extended Privacy Impact Assessment or not.

Screening Questions

The extended Privacy Impact Assessment includes 56 questions. The questions are grouped into five topical areas, which refer to: 1) the type of project, 2) the collection and use of data, 3) the project's storage and security policies, 4) transfer of info, and 5) cloud specific issues.

The aim of this set of questions is to assess in a granular manner how the interactions between you and the CSP you deal with impact your users' rights to privacy and data protection, and how your system is designed – if so – to prevent or mitigate the potential adverse outcomes of those interactions.

You are to answer all questions to the best of your knowledge, if necessary asking the relevant professionals in your undertaking before answering; some questions, though, allow you to answer "I do not know" (yet!), but please do mind – you are supposed to know.

Disclaimer

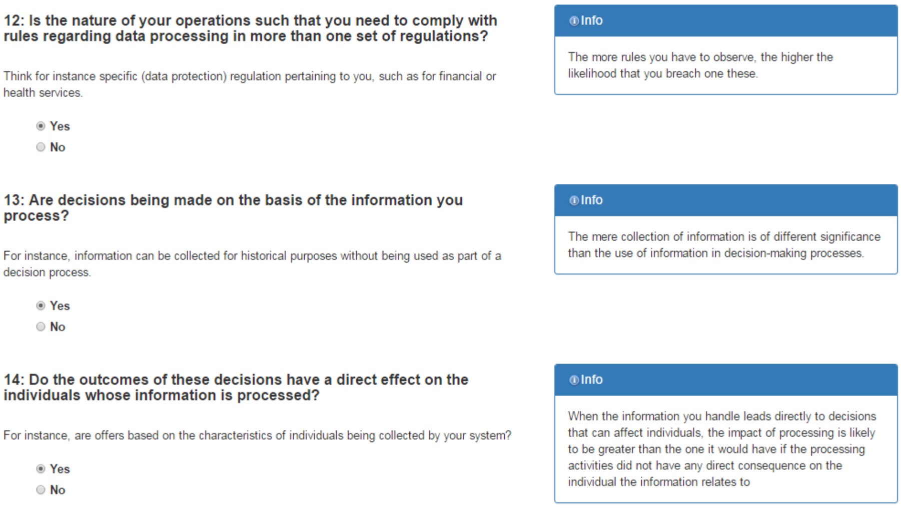
No information or content displayed in this tool should be construed, interpreted or relied upon as constituting legal advice, or a recommendation in respect of taking any course of action to comply with data protection laws, or legal obligations of any kind, and within any jurisdiction to which the European data protection law applies.

Nothing in this tool is intended to be an invitation or inducement to engage or enter into, or advice against engaging or entering into, an undertaking of any kind.

The content or information displayed in this tool is for general informational purposes regarding compliance with the applicable data protection laws only.

A4Cloud shall not be liable for any damages resulting from the use of the tool, including damages caused by any incorrectness or incompleteness of information provided on the tool. The content provided in the tool does not necessarily represent the

Fig. 3. DPIAT tooltip displaying information about the selected options



12: Is the nature of your operations such that you need to comply with rules regarding data processing in more than one set of regulations?

Think for instance specific (data protection) regulation pertaining to you, such as for financial or health services.

● Yes
○ No

ⓘ Info

The more rules you have to observe, the higher the likelihood that you breach one these.

13: Are decisions being made on the basis of the information you process?

For instance, information can be collected for historical purposes without being used as part of a decision process.

● Yes
○ No

ⓘ Info

The mere collection of information is of different significance than the use of information in decision-making processes.

14: Do the outcomes of these decisions have a direct effect on the individuals whose information is processed?

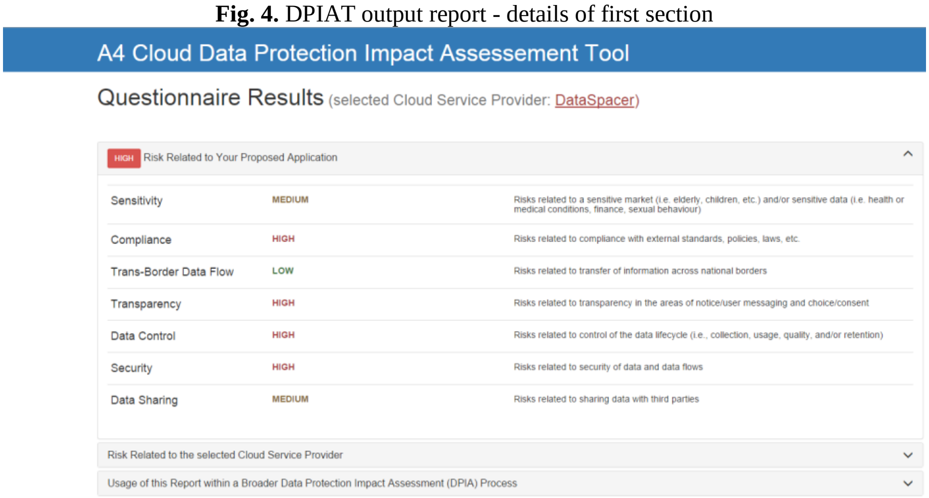
For instance, are offers based on the characteristics of individuals being collected by your system?

● Yes
○ No

ⓘ Info

When the information you handle leads directly to decisions that can affect individuals, the impact of processing is likely to be greater than the one it would have if the processing activities did not have any direct consequence on the individual the information relates to

Fig. 4. DPIAT output report - details of first section



A4 Cloud Data Protection Impact Assessement Tool

Questionnaire Results (selected Cloud Service Provider: DataSpacer)

HIGH Risk Related to Your Proposed Application

Sensitivity	MEDIUM	Risks related to a sensitive market (i.e. elderly, children, etc.) and/or sensitive data (i.e. health or medical conditions, finance, sexual behaviour)
Compliance	HIGH	Risks related to compliance with external standards, policies, laws, etc.
Trans-Border Data Flow	LOW	Risks related to transfer of information across national borders
Transparency	HIGH	Risks related to transparency in the areas of notice/user messaging and choice/consent
Data Control	HIGH	Risks related to control of the data lifecycle (i.e., collection, usage, quality, and/or retention)
Security	HIGH	Risks related to security of data and data flows
Data Sharing	MEDIUM	Risks related to sharing data with third parties

Risk Related to the selected Cloud Service Provider

Usage of this Report within a Broader Data Protection Impact Assessment (DPIA) Process

Table 1. Data Protection Impact Assessment Screening Questions

ID	***Question***	***Explanation***	***Question type***
1	Based on the information that you process, can you identify one or more individuals about whom you are processing information?	Can the information used be associated to a particular customer or employee, either directly (e.g. by using names) or indirectly (e.g. by using license plates, social security number, addresses, telephone numbers or other information that you hold)?	Y/N
2	Does the information that you process reveal certain characteristics of individuals?	Can you, or will you, use the information you process to qualify your customer or employee, for instance on the basis of (online) behavior, attendance, marital or social status, salary level, work performance, or zip code? If you build 'profiles' of individuals, answer yes to this question.	Y/N
3	Do you deal with any kind of the following categories of information?	The following categories of information are of a particularly sensitive nature, and need to be dealt with.	[Checkbox] • race or ethnic origin; • political opinions; • religion or philosophical beliefs; • sexual orientation or gender identity; • trade-union membership and activities; • genetic or biometric data or data concerning health or sex life;

			• administrative sanctions, judgments, criminal or suspected offences; • data on children; • data on employees; • location data; • data that can be used for identity theft, such as social security number, credit card information, passport or driving license data.
4	What is the scale of your processing operations?	The scale includes, for instance, the number of persons to whom the information you deal with relates to, the amount and granularity of information per person or the number of people who have access to the information that you process.	• Large • Medium • Small • I don't know • Not applicable
5	Is the nature, scope and/or purpose of your business, profession or activity based on a regular and systematic monitoring either of any natural person(s) or of publicly accessible areas?	Think, for instance, of virtual public areas, such as social networks or public fora.	Y/N
6	How likely is that incidents will raise concerns amongst individuals and/or legal entities?	Think of, for instance, data breaches, inaccurate, incomplete or outdated data related to the information that you process, use of data for purposes other than the ones for which they were collected.	• Large • Medium • Small • I don't know • Not applicable
7	Are there any third parties involved in the storage, processing, use, or transfer of any information that you deal with?	The interplay with third parties exponentially increases the risks deriving from processing activities.	Y/N

Table 2. DPIA Questionnaire

ID	***Question***	***Explanation***	***Question type***
		Type of project	
1	Is the establishment of your activities in European territory?	Whether the processing of personal information of your undertaking takes place in the European Union or not is not relevant. If you are not established in European Union territory, but you offer goods or services to individuals in the EU or monitor them,	Y/N

		then you should answer Y to this question.	
2	Do you gather information that can identify other people through one or more of the following activities?	Think for instance, if you use names, identification numbers or location data. The collection of information related to individuals can be potentially intrusive to the information privacy rights of these individuals. In some types of projects information provided is more sensitive than in other ones e.g. Financial data.	[Checkbox] - Web Browsing - Account and/or Subscription Management - Authentication and Authorization - Customization - Responding to User - (Service) Delivery - Software Downloads - Sales of Products or Services - Communications Services - Banking and Financial Management - Payment and Transaction Facilitation - Charitable Donations - Government Services - Healthcare Services - Education Services - Advertising, Marketing, and/or Promotions - News and Information - Arts and Entertainment - Surveys and Questionnaires - Online Gambling - Online Gaming - Search Engines - State and Session Management
3	For which of the following purposes or legitimate interests do you process the information?	To be legitimate, the processing of information should be based on legitimate interests. Some interests carry more weight than others. For instance processing for historical, scientific statistical or research purposes is likely to be less intrusive to information privacy rights than processing for exercise of the right to freedom of expression or information.	[Checkbox] Purposes related to the commercial objective of your undertaking *Health purposes:* - for preventive or occupational medicine, medical diagnosis, the provision of care or treatment or the management of health-care services - for public interest in the area of public health, such as protecting against serious cross-border threats - for other reasons of public interest in areas such as social protection *Employment context:* - for purposes of the recruitment and job applications within the group of undertakings

			- for the performance of the contract of employment, including discharge of obligations, laid down by law and by collective agreements, - management, planning and organisation of work, health and safety at work, - for the purposes of the exercise and enjoyment of rights and benefits related to employment - for the purpose of the termination of the employment relationship Purposes within the social security context Processing for historical, scientific statistical or research purposes Enforcement of legal claims and/or compliance with law enforcement agencies Exercise of the right to freedom of expression or information (including in the media and the arts) Other (Please specify)
	Collection and Use of Information		
4	Are you relying exclusively on consent in order to process information of individuals?	Consent means ‘any freely given specific, informed and explicit indication of his or her wishes by which the individual either by a statement or by a clear affirmative action signifies agreement to information relating to them being processed.’	Y/N
5	How have you obtained the consent of individuals?	Consent requires prior information and an explicit indication of the intent to consent.	a) Consent is given directly by the individual by a statement (e.g. by a consent form) b) Consent is given directly by the individual by an affirmative action (e.g. by ticking a box) c) Consent has been obtained implicitly by the individual (e.g. by the mere use of the service or inactivity)
6	If individuals have given their consent, can they withdraw it with ease and whenever they want to?	Individuals should be able to withdraw their consent at any time and every step of the processing of their information without detriment. It should be as easy to withdraw consent as it is to give it.	Y/N
7	Are the consequences of withdrawal of consent significant for individuals?	For instance, will the service to the individual be terminated *tout court*, while the individual still depends on it?	Y/N

8	On what basis do you process the information?	In order for the processing to be lawful, at least one of these grounds must be satisfied.	[Checkbox] a) The individual has given his consent b) Processing is necessary for the performance of a contract between you and the individual whose information you process c) Processing is necessary for compliance with a legal obligation you have d) Processing is necessary in order to protect vital interests of the individuals whose information you process e) None of the above
9	Do you provide clear information about:		[Y/N Radio button] - the purposes for which you process personal information - the different types of information that you process - your identity
10	Are all the information and its subsets you handle necessary to fulfil the purposes of your project?	The information you collect/process/handle should be adequate, relevant, and limited to the minimum necessary in relation to the purposes for which they are processed. This means that you have to use the minimum information necessary for your purposes, but you are not prohibited to have multiple purposes.	Y/N
11	Is it possible for the individual to restrict the purposes for which you process the information?	For instance, are individuals given the possibility to opt-out of receiving email offers from you?	Y/N
12	Is the nature of your operations such that you need to comply with rules regarding data processing in more than one set of regulations?	Think for instance specific (data protection) regulation pertaining to you, such as for financial or health services.	Y/N
13	Are decisions being made on the basis of the information you process?	For instance, information can be collected for historical purposes without being used as part of a decision process.	Y/N
14	Do the outcomes of these decisions have a direct effect on the individuals whose information is processed?	For instance, are offers based on the characteristics of individuals being collected by your system?	Y/N

15	Does the information you process about individuals produce a full and correct image of these individuals?	The chances of taking wrong decisions increase if the information is incomplete, outdated or wrong. In such cases, the risk of setting individuals' rights at stake is higher.	Y/N/IDK
16	Does the information you process about the individual come from different sources?	Think, for instance, whether you obtain databases from other parties	Y/N
17	Are the individuals whose information you process aware of the fact that the information comes from different sources?	Consider whether you have informed the individuals about the information you process and which might come from other sources.	Y/N
18	Does your project involve the use of existing personal information for new purposes?	For instance, you may decide that you want to use the contact details you obtained for signaling the user that their order has been fulfilled for marketing purposes later on.	Y/N
19	Do your additional processing operations relate closely to the original purposes for which you first collected the information?	For instance, using a customer's home address for frequent delivery of packages after the first delivery is compatible use, whereas providing a patient list to one spouse, who runs a travel agency; so that he can offer special holiday deals to patients needing recuperation is not.	Y/N
20	Is the use of existing personal information for new purposes clearly communicated to the individual in a timely manner?	Consider whether you have informed the individuals about the specific (new) purposes for which you process the information.	Y/N
21	Is the use of existing personal information for new purposes clearly communicated to your organization's data protection officer?	Consider whether you have informed the data protection officer about the specific (new) purposes for which you process the information.	Y/N
22	Do you appropriately notify your national DPA before performing data processing operations subject to prior checking?	In some cases your processing activities are subject to prior checking by your national DPA.	Y/N

23	Do you process information which could potentially be perceived as discriminatory?	Think for instance, whether you process information solely on the basis of race or ethnic origin, political opinion, religion or beliefs, trade union membership, sexual orientation or gender identity etc.	Y/N
		Storage and Security	
24	Are procedures in place to provide individuals access to information about themselves?	Consider, for instance, whether individuals can request an overview of the information about them that you have	Y/N
25	Can the information you process be corrected by the individuals, or can individuals ask for correction of the information?	An increased level of involvement by the individual decreases the likelihood of unwarranted events (e.g. incorrect information)	Y/N
26	Do you check the accuracy and completeness of information on entry?	Consider, for instance, whether you apply specific procedures (e.g. use of journalistic archives to double-check the content) in order to ensure the validity and authenticity of the information you process.	Y/N
27	How often is the personal information you process updated?	Outdated information has a negative impact on the accuracy of information you process.	[Checkbox] - Frequently - When requested by the individual - Whenever necessary to comply with technological developments - Rarely - Never
28	How severe would you deem the consequences, in case you process outdated information for the individuals it refers to?	For instance, having outdated information about individuals (e.g. wrong date of birth) may hold you liable.	- High - Medium - Low - None
29	Would the fact that the information you process is not up to date lead to sanctions provided in relevant regulations?	Think, for instance, whether the nature of your activities requires you to comply with specific sets of regulations, which provide sanctions in order to keep the information updated.	Y/N/IDK
30	Do you have a Data Security Policy?	Think of aspects such as: is it clear who is responsible for security, do you adopt security standards, is the (sensitive) nature of the information you process taken into account	Y/N

31	Do you implement any technical and organizational security measure from the outset of your activities?	Think, for instance, whether you are using signatures, hashing, encryption etc. or whether you implement Privacy by Design and/or Privacy by Default mechanisms from the very design phase of your projects.	Y/N
32	Do you differentiate your security measures according to the type of information that you process?	For instance information related to race or ethnic origin, political or sexual orientation, religion or gender identity of the individuals requires specific security measures.	Y/N
33	Is the personnel in your undertaking trained on how to process the information you deal with according to the organisational policies you implemented?	Consider if you apply specific procedures or timetables to train your employees with regard to the manner in which they should process the information.	Y/N
34	How often are your Security and Privacy Policies updated?		[Radio button] - Frequently - Whenever necessary to comply with technological developments - Rarely - Never
35	Do you adopt one or more of the following measures and/or procedures as a safeguard or security measure to ensure the protection of personal information?	The application of one or more of the following measures may prevent potential misuse of the information you handle.	[Checklist] - Personal information is kept confidential - Access control is enforced - Segregation of duty is used - Special authorization for personnel who access the information - Compliance with further regulations is ensured - Use of personal information are properly documented - Procedures to maintain personal information use up-to-date regularly - Subcontractors follow the same guidelines on documenting the use of information - Procedures to notify individuals, when necessary, are in place - Procedures to take into account the impact of the information lifecycle - Procedures to record individuals' requests for correction of information - Specific procedures to respond to Law Enforcement access or court orders - Modalities to express, withhold, or withdraw informed consent to the processing

			- Anonymization - Pseudonymisation - Encryption - Aggregation - Separation - Limitation of usage - Data segregation - Sticky Policies - All of the above - None of the above
36	If you use encryption methods, are you responsible for encrypting and decrypting the information that you process?	If you are the only one responsible for encrypting and decrypting the information you process, you are subsequently the only one who has control over this information. Instead, if you have given such a competence to a cloud service provider you do not have the same level of control over the information.	Y/N
37	Do the protection measures you have in place, in case of unwarranted incidents, specifically target the particular type of incident that might happen?	For instance, in case of unauthorized access/disclosure/modification, intentional or reckless destruction of or damage to your equipment, loss or theft of your assets etc. Such incidents threaten the protection of personal information	Y/N
38	Do you take action in order to notify individuals in case of (security) incidents?	E.g. by sending emails.	Y/N
39	What do you do to minimize the damages of physical, technical and/or security incidents?		[Checklist] - Segregation of data bases - Limitation of use/transfer functionalities on system layer - Separation on system layer - Multi-tenancy limitations - Physical separation of infrastructure - None of the above - Others (please indicate)
40	Does the project(s) include the possibility by individuals to set retention periods on their own?	Setting retention periods allows you to ensure that the information that you process about individuals is kept for no longer than is necessary for your operations.	Y/N
41	For how long do you store the information you are dealing with?		[Checklist] a) Only for the completion of the project's purposes b) Information is retained for a certain time after the project has been completed c) Information is retained for the possibility of future uses or new purposes d) Until individual requests for erasure

Transfer of information			
42	Do you normally transfer the information you deal with to third parties during your normal processing operations?	Do you, for instance, outsource the processing of the information you deal with to third parties?	Y/N
43	Is the third parties' use compatible with the one you set for your undertaking?	If you transfer information to third parties, do they use the information in a manner consistent with your original purpose(s) and their mandate?	Y/N
44	Do you sell, rent or by any means disseminate information to third parties?		Y/N
45	Are you transferring and/or simply disclosing personal information exclusively to countries or territories outside the EEA?	The EEA consists of the following countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxemburg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom.	Y/N
46	Are you transferring personal information exclusively to one or more of the following non-EEA countries?	Each of these countries are deemed to have adequate privacy protection in terms of the EU data protection regulations	[Checklist] - Andorra - Argentina - Australia - Canada - Switzerland - Faeroe Islands - Guernsey - Israel - Isle of Man - Jersey - New Zealand - Uruguay - U.S.
47	Are measures in place to ensure an adequate level of security when the information is transferred outside of the EEA?	Not all countries have the same level of protection as regards to the processing of personal information. Transferring personal information towards countries without an adequate level of protection is a breach of EU data protection laws.	Y/N/IDK
Cloud Specific Questions			

48	The cloud infrastructure I use is:	The potential threats to privacy and protection of personal information are influenced by the deployment model of the CSP. This means that the risk is higher if the number of the subjects who operate in the system is also high.	a) owned by or operated for only me (private cloud) b) is owned by or operated for a specific group of users with common interests in a shared manner (community cloud) c) is shared amongst multiple users (public cloud)
49	Does the service provider that you use provide you just with raw computing resources, such as processing capacity or storage, for the information that you process?	Think for instance of Amazon AWS or Microsoft Azure	Y/N
50	Does the service provider you use provide you with an environment or platform in which you can develop and deploy software?	Think for instance of Google App Encine or Force.com	Y/N
51	Does the service that you use consist of the provision of end user applications run by the cloud service provider?	Think for instance of SalesForce CRM or Wuala.	Y/N
52	Are specific arrangements in place with regards to your information in case you want to terminate or transfer the cloud service?	The application of such rules/procedures gives you the ability to have control over the information you process. For instance, you can transfer the information you process to another provider if necessary (e.g. in case of bankruptcy, force majeure etc).	Y/N/IDK
53	Does the CSP apply specific procedures in order to secure the information you handle and/or process in case your business is discontinued?	Think, for instance, if the information that you process are preserved in case of merger, acquisition, bankruptcy, etc.	Y/N/IDK
54	Does the CSP have an insurance policy against the possible loss or compromise of the information you process in a cloud environment?	Think for instance if the provider is able to redress you in case of unwarranted incidents concerning the information that relates to them through an insurance scheme or similar ones.	Y/N/IDK

55	Does the CSP use resource isolation mechanisms in order to secure the information you entrust it?	Think, for instance, about how the CSP ensures the isolation of your information from the information of other customers potentially located in the same physical machine, albeit of course in a different virtual one.	Y/N/IDK
56	Are the CSP's activities certified by any kind of supervisory organisation or body?	Think for instance, if the CSP has obtained a certification by a supervisory body or organization, which can guarantee the quality of his services and his compliance with the law.	Y/N/IDK

Table 3. ENISA's list of risk scenarios and their categories

Risk Category	**Risk name**	
Policy & Organisational	P1.	Lock-in
	P2.	Loss of governance
	P3.	Compliance challenges
	P4.	Loss of business reputation due to co-tenant activities
	P5.	Cloud service termination or failure
	P6.	Cloud provider acquisition
	P7.	Supply chain failure
Technical	T1.	Resource exhaustion (under or over provisioning)
	T2.	Isolation failure
	T3.	Cloud provider malicious insider - abuse of high privilege roles
	T4.	Management interface compromise (manipulation, availability of infrastructure)
	T5.	Intercepting data in transit
	T6.	Data leakage on up/download, intra-cloud
	T7.	Insecure or ineffective deletion of data
	T8.	Distributed denial of service (DDoS)
	T9.	Economic denial of service (EDOS)
	T10.	Loss of encryption keys
	T11.	Undertaking malicious probes or scans
	T12.	Compromise service engine
	T13.	Conflicts between customer hardening procedures and cloud environment
Legal	L1.	Subpoena and e-discovery
	L2.	Risk from changes of jurisdiction
	L3.	Data protection risks
	L4.	Licensing risks
Not Specific to the Cloud	N1.	Network breaks
	N2.	Network management (i.e., network congestion / mis-connection / non-optimal use)
	N3.	Modifying network traffic
	N4.	Privilege escalation
	N5.	Social engineering attacks (i.e., impersonation)
	N6.	Loss or compromise of operational logs
	N7.	Loss or compromise of security logs (manipulation of forensic investigation)
	N8.	Backups lost, stolen
	N9.	Unauthorised access to premises (including physical access to machines and other facilities)
	N10.	Theft of computer equipment
	N11.	Natural disasters