DEFeND DSM: A Data Scope Management Service for Model-Based Privacy by Design GDPR Compliance

Luca Piras[1], Mohammed Ghazi Al-Obeidallah[1], Michalis Pavlidis[1], Haralambos Mouratidis[1], Aggeliki Tsohou[2], Emmanouil Magkos[2], Andrea Praitano[3], Annarita Iodice[3], and Beatriz Gallego-Nicasio Crespo[4]

[1] Centre for Secure, Intelligent and Usable Systems,
University of Brighton, Brighton, United Kingdom
{l.piras,m.al-obeidallah2,m.pavlidis,h.mouratidis}@brighton.ac.uk
[2] Ionian University, Corfu, Greece
{atsohou,emagos}@ionio.gr
[3] Maticmind SpA, Rome, Italy
{andrea.praitano,annarita.iodice}@maticmind.it
[4] Atos, Madrid, Spain
{beatriz.gallego-nicasio}@atos.net

Abstract. The introduction of the European General Data Protection Regulation (GDPR) has brought significant benefits to citizens, but it has also created challenges for organisations, which are facing with difficulties interpreting it and properly applying it. An important challenge is compliance with the Privacy by Design and by default (PbD) principles, which require that data protection is integrated into processing activities and business practices from the design stage. Recently, the European Data Protection Board (EDPB) released an official document with PbD guidelines, and there are various efforts to provide approaches to support these. However, organizations are still facing difficulties in identifying a flow for executing, in a coherent, linear and effective way, these activities, and a complete toolkit for supporting this. In this paper, we: **(i)** identify the most important PbD activities and strategies, **(ii)** design a coherent, linear and effective flow for them, and **(iii)** describe our comprehensive supporting toolkit, as part of the DEFeND EU Project platform. Specifically, within DEFeND, we identified candidate tools, fulfilling specific GDPR aspects, and integrated them in a comprehensive toolkit: the DEFeND Data Scope Management service (DSM). The aim of DSM is to support organizations for continuous GDPR compliance through Model-Based Privacy by Design analysis. Here, we present important PbD activities and strategies individuated, then describe DSM, its design, flow, and a preliminary case study and evaluation performed with pilots from the healthcare, banking, public administration and energy sectors.

Keywords: Privacy by Design · Privacy Engineering · Security Engineering · Data Protection · GDPR · Data Scope Management · Privacy

1 Introduction

The European General Data Protection Regulation (GDPR) was introduced to enforce citizen data protection and privacy rights. Despite the clear benefits for citizens, GDPR is posing a major challenge for organisations, as they need to comply with a large number of areas including data classification, tracking of data processing activities with reporting and registers, data monitoring, breach detection, fast intervention and fast data deletion. Organisations failing to comply are liable to huge financial fines from relevant authorities [12]. A major problem is that GDPR is abstract and lacks detailed and clear information on how the various articles can be implemented in practice.

One of the most challenging and difficult principles to adhere with is Data Protection by Design and by Default; hereafter, for the sake of simplicity, we refer to these priciples as Privacy by Design (PbD). Although GDPR defines PbD and makes it clear that it should be followed, it does not provide details on how it can be implemented. This is problematic because organisations do not have a structured way to ensure that PbD is followed when developing new systems and services. Recently, in order to try to cover this important lack of practical guidance, the European Data Protection Board (EDPB), released an official document for providing PbD guidelines[1]. However, those guidelines, even helping in reducing such gap, are still at high-level, and offer few practical indications. What is still missing is a clear structured approach that will enable organisations to implement PbD and a set of tools that would support the automation of such structured approach.

This paper provides a novel structured framework and a toolkit that fulfils this gap of the current state of the art. The Data Scope Management (DSM) solution presented is part of the DEFeND EU Project[2] platform [11], and builds on previous work presented at TrustBus-19 [11]. In particular, this paper addresses the following Research Questions (RQs):

RQ1: What are the analysis and implementation activities required by PbD and how these can be carried out in a structured and methodological way?

RQ2: Can PbD activities being automated and supported by software tools?

RQ1 is the main RQ that this paper tries to answer while **RQ2** is a supportive question. To answer the first question we elicited information from Data Protection Officers (DPOs), experts and end-users [15, 16] of organizations from different GDPR relevant sectors (e.g., banking, public administration, healthcare, energy). We analysed the outcome of these activities and derived a set of activities, strategies and factors that are important for the implementation of PbD. We then, based on those factors and activities developed a novel service, DSM, to support those. We also individuated a number of tools, and extended them, to make them to provide automated support to DSM.

[1] `https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201904_dataprotection_by_design_and_by_default.pdf`

[2] `https://www.defendproject.eu/`

The rest of the paper is organized as follows. Section 2 summarizes the requirements we elicited in previous works [15,16], and answers to **RQ1** providing the activities and strategies for PbD we derived for the DSM flow and toolkit. Section 3 addresses **RQ2** and describes the DSM flow, toolkit, data models, our case study and preliminary evaluation within DEFeND. Section 4 compares our work with the industry and the literature. Section 5 concludes this paper.

2 PbD Activities and Strategies for GDPR Compliance

As indicated above, an important aspect of our work was to identify a set of analysis and implementation activities related to PbD. in doing so, we employed a Human-Centered Design (HCD) approach [8], where questionnaires and interviews were used as the basic tool to capture the main stakeholders' requirements with regards to PbD and also to understand the main characteristics that an automated toolkit should possess to support PbD [15, 16]. Our approach consisted of 3 main stages [15, 16] describe in the next.

Questionnaire Preparation. After an initial phase where the internal and external key stakeholders were identified, e.g., DPOs, IT managers, citizens etc., a questionnaire was prepared, for each user category, in a systematic way [16], aiming to capture the legal, functional, security, privacy and technology acceptance needs [10]. Specifically, we followed the approach of [1] for customer development, including steps such as Customer Segmentation, Problem Discovery and Validation, Product Discovery and Validation [16]. Two online questionnaires were prepared: 1 for end-users[3] and 1 for citizens[4];

Questionnaire Validation and Distribution. A validation phase were organized, where: **(i)** 10 DPOs from all project partners commentated on the questionnaires, and **(ii)** a focus group with internal stakeholders from the banking sector were set to revise and discuss final questionnaires. Questionnaires were then distributed to both end-users (i.e., organisations from 4 different sectors: banking, energy, health, public administration) and citizens from 7 European countries (i.e. Italy, Greece, Spain, Bulgaria, France, Portugal, UK), and were filled using semi-structured interviews and online surveys [16];

Data Analysis. During a data collection phase, we collected information from 10 DPOs via interviews and 31 DPOs via online survey, representing the energy, education, banking, health, public administration and information technology consultancy sectors. We also collected data from 174 citizens. The captured needs were analyzed, using qualitative data techniques and value analysis, and translated into software development requirements.

2.1 Identified Activities and Strategies for PbD

Our analysis of the above interviews, and questionnaires, identified Activities and Strategies (AS), which are important for PbD. We discuss them below.

[3] https://ec.europa.eu/eusurvey/runner/DEFeNDEndUser
[4] https://ec.europa.eu/eusurvey/runner/DEFeNDCitizens

AS1: *Organization Situation and Context.* It is fundamental to execute deep analyses and data collection on the organization, for having an important baseline on which to perform PbD activities identified in the next AS. Thus, from the very early stages of the analysis, for achieving GDPR compliance in a PbD way, it is needed to start the data collection by working on the GDPR self-assessment of the organization. This will help to produce, later, according to the other AS, a GDPR action plan identifying current gaps of compliance of an organization, on which to perform further PbD analyses.
AS2: *Organization and* 3^{rd} *Parties Profiles.* On the basis of the high-level contextual information identified in **AS1**, it is needed to further analyse and collect more details for creating complete profiles of the organization and 3^{rd} parties, including economic, financial and legal aspects.
AS3: *Data Processing Activities and Data Categories.* It is also needed to conduct a deep analysis on data processing activities performed by the organization itself, and in collaboration with 3^{rd} parties. This should include also the identification of data categories and assets involved.
AS4: *GDPR Data Syntheses, Graphical Representations and Model-Based, Visual Support.* At support of all the AS, in particular for the analyses, it is beneficial to provide further support and guidance with graphical representations and synthesis of GDPR information analysed and collected. These should be provided to business analysts, privacy/security experts and other end-users involved, based on the completion of the GDPR Self-Assessment, and at support to other activities (e.g., Data Protection Impact Assessment, data minimization analysis, creation of GDPR action plans). While, privacy/security analysiss, threat analysis, continuous risk assessment configurations, and other critical activities and analyses, could be performed supported by visual model-based techniques enhanced and adapted for GDPR purposes.
AS5: *Data Protection Impact Assessment (DPIA), Preventive/Reacting Analyses and GDPR Action Plan.* On the basis of the elements identified by the other AS, it is important to analyse, in a preliminary way, GDPR lacks, vulnerabilities and assets that can be affected by data issues/breaches, and which preliminary mitigation mechanisms to adopt, and if preventive/reactive actions are in place (e.g., data breach plans). These analyses should be performed for producing a DPIA and a GDPR Action Plan, for identifying current gaps of compliance of an organization, on which to perform further PbD analysis.
AS6: *Privacy/Security Model-Based, and Pattern-Based, Analysis.* The GDPR Action Plan of **AS5** identifies the gaps, but it is usually at high-level, thus, needs to be enacted by further critical analysis, performed by privacy/security analysts, supported by visual model-based techniques enhanced and adapted for GDPR purposes (**AS4**). This concerns analysis of the organization context, data/assets/accountability mapping with also analysis of risks, analysis of threats and measures in place, privacy/security requirements constraints and conflict resolution, supported through libraries of patterns and modeling techniques specifically designed for GDPR.
AS7: *Continuous Model-Based GDPR Compliance.* On the basis of the analyses performed for the previous AS, it is needed to support the organization:

(i) in having software systems able to put in place GDPR compliance solutions individuated, **(ii)** to receive automated support for configuring such systems, **(iii)** to monitor continuously the compliance, according to the GDPR plan, for identifying new potential lacks with GDPR and data breaches, **(iv)** to enable the organization to react to such problems, and **(v)** to make this process iterative, for a continuous Model-Based GDPR compliance, by enabling the analysts to analyse in a visual, model-based way the new GDPR lacks, and to perform again AS analysis, in a continuous way, for updating/re-configuring the system for being again GDPR compliant.

3 DEFeND Data Scope Management (DSM) Service, Case Study and Evaluation

Based on the above set of Activities and Strategies (AS), we have designed a flow for such AS, and developed a novel service, the Data Scope Management service (DSM), for the DEFeND platform to support PbD. According to our AS, DSM supports organizations in performing GDPR self-assessments by collecting organizational information (**AS1**), also related to 3^{rd} parties (**AS1**), data processing activities (**AS3**), and creating a profile of the organization regarding multiples perspectives such as legal, economic and financial aspects (**AS2**). Furthermore, it also enables organizations in executing DPIA (**AS5**) by collecting/revising and refining organizational assets (**AS3**), and elaborating the other information collected for supporting the organizations with data synthesis and graphical representations (**AS4**) through a set of DSM tools. Moreover, DSM helps organizations in performing threats analysis (**AS4**, **AS6**), data minimization analysis (**AS4**), privacy/security analysis and design with tool-supported modelling techniques (**AS4**, **AS6**), continuous risk assessment (**AS4**, **AS6**), and configuration for executing a continuous model-based GDPR compliance (**AS7**).

In the next subsections, we start giving an overview of DSM, its components, the tools we selected, extended and integrated for creating DSM, and the data models used by the tools for exchanging PbD information needed by our AS. Then, we outline our case study, performed by involving pilots from the healthcare, banking, public administration and energy sectors. Together with the case study, we describe the DSM PbD flow through a healthcare storyline. In the last subsection, we discuss our preliminary evaluation.

3.1 DSM Components, Integrated Tools and Data Models

In order to design and develop DSM, we individuated candidate tools, supporting specific features, and extended and integrated them, according to AS and the DSM flow, for creating a service supporting the entire set of features required for a PbD approach. Specifically, DSM involves the following tools: the MM-Assess (MaticMind-Assess) tool, which supports the business analyst to conduct a self-assessment for the organization; MM-REPA (MaticMind Record of Processing

Activities), which is a tool that creates a list of all data processing activities in the organization based on a guided questionnaire; MM-PIA, a Risk Assessment Management (RAM) tool, which provides a centralized system to identify risks, evaluate their impact, probability, and the vulnerability they pose to organizational assets, linking them to mitigating controls and managing their resolution; the SecTro tool, which is a CASE tool guiding and supporting analysts in the construction of appropriate models, based on the Secure Tropos method [9]; the Risk Assessment Engine (RAE), which is an ATOS tool supporting organizations in the assessment of cyber-risks.

Interactions of the DSM tools are made through the exchange of information stored in data models as shown in Fig. 1. Therefore, data models involved

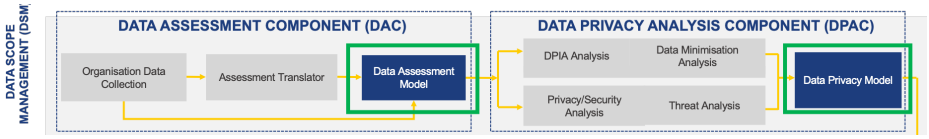

Fig. 1. DSM components, modules and DSM Data Models (green rectangles) [11,15,16]

in DSM are the Data Assessment Model (DAM) and the Data Privacy Model (DPM). DAM is produced in the Data Assessment Component (DAC), then read in the Data Privacy Analysis Component (DPAC) that in turn produces the DPM model. The DPM model is then used by other services of the DEFeND Platform, for instance from the GDPR Reporting Service [11]. Concerning DSM components and modules (Fig. 1), DAC is constituted by the Organization Data Collection (ODC) module and the Assessment Translator (ATr) module. While, DPAC is composed of the DPIA Analysis module, Data Minimization Analysis module, Privacy/Security Analysis module and the Threats Analysis module.

3.2 Case Study, Storyline and DSM PbD Flow

Our case study used a storyline, we devised, for touching the most important PbD activities of DSM, and we used such storyline for demonstrating and discussing DSM, and our approach, with pilots from the banking, healthcare, public administration and health sectors, within the DEFeND Project[2]. In the following, we start introducing our storyline, then describe DSM and its flow, phase by phase, by using the storyline, for demonstrating DSM in a way compliant with the case study performed with the pilots. Fig. 2 represents the DSM flow as an activity diagram: **(i)** the phase number is indicated in the top, left corners of rectangles; **(ii)** some phases include more than one rectangle; **(iii)** each activity has a label in the top, right corner indicating the name of the tool fulfilling it.

Storyline Introduction. A Hospital wants to improve its GDPR compliance by using the DEFeND DSM service. It is important to note that, even though for this example we are considering the healthcare sector, the DSM service has been designed and delivered to be as much flexible as possible to support organizations from heterogeneous sectors. One of the most critical aspects for a hospital is to manage the patient medical record and to have verifications, from a supervisor,

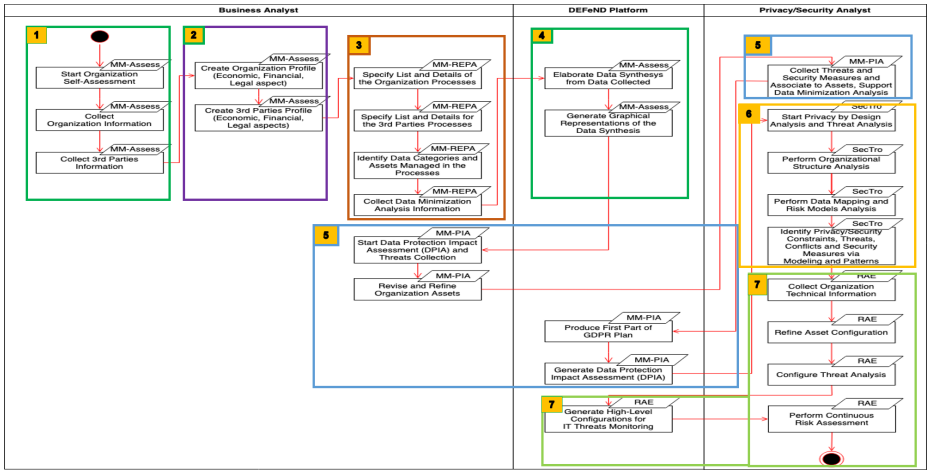

Fig. 2. Activity Diagram of the DSM Flow

for any changes happening to it (for instance adding a new medical exam result, etc.), and to establish retention periods for this data. Furthermore, this data has not to be stolen or to be compromised; for instance, in relation to potential threats and data breaches; therefore, the Hospital needs to analyse, design and put in place monitoring of those potential problems; in the organizational processes are involved also 3^{rd} parties (external laboratories for medical exams), therefore it is needed to consider also this for improving GDPR compliance.

DSM Flow: Phase 1 (DAC: Initial Organization Data Collection). Phase 1 covers mainly **AS1** and partially **AS2**. Its activities are represented in Fig. 2. Main objectives of this phase are to support the organization in: performing GDPR self-assessment (**AS1**), collecting high-level organization information (**AS1**) and 3^{rd} parties information (**AS2**). This phase is associated to the MM-Assess tool (Fig. 2) within the DAC component and the ODC module (Fig. 1). The user of the organization for this phase is typically a business analyst (Fig. 2). Most of the activities performed during this phase are related to collection of information through questionnaires compilation. Information collected are saved in the DAM model. This phase is illustrated by the following part of the storyline:

> "*The Hospital starts using the DSM service and inputs in the system relevant Organizational and* 3^{rd} *Parties information by compiling initial questionnaires for giving an overview of the organizational context.*"

For instance, the business analyst of the hospital can collect, by using MM-Assess questionnaires, the laboratory information, i.e. the lab in charge of executing medical exams to patients for the hospital, and related information will be populated in the DAM data category called "3^{rd} Parties".

DSM Flow: Phase 2 (DAC: Organization Data Collection for Profiles Creation). This phase covers **AS2**, and its activities are represented in Fig. 2. Here, the organization is able to create complete profiles, both for the organization and 3^{rd} parties, concerning economic, financial and legal aspects (**AS2**). This phase is performed by a business analyst of the organization, in the context of the DAC component and the ODC module (Fig. 1), using the MM-Assess (Fig. 2)

by being guided in compilation of questionnaires, which will populate the DAM model. This phase is illustrated by the following part of the storyline:

> ''*Afterwards, the system proposes to the user to compile more detailed questionnaires able to create a complete organizational profile and* 3^{rd} *parties profile regarding economic, financial and legal aspects.*''

For example, the business analyst can input information on the organization business, legal and economic situation that could be related to organization debts of the hospital (data category ''Organization General Information'' of DAM).

DSM Flow: Phase 3 (DAC: Organization Data Collection of Data Processing Activities). This phase covers mainly **AS3** and partially **AS4**. Its activities are represented in Fig. 2. The main objectives of this phase are to complete the self-assessment by identifying the data processing activities of the organization (**AS3**), including also the ones occurring with 3^{rd} parties, the data categories and assets involved and managed (**AS3**), and to collect data minimization analysis information in relation to how it has been conducted so far by the organization (**AS4**). This phase is performed with the MM-REPA tool (Fig. 2) within the DAC component and the ODC module (Fig. 1). To execute these activities, a business analyst of the organization inputs this information via questionnaires compilation. Information collected are saved in the DAM model. This phase is illustrated by the following part of the storyline:

> ''*Subsequently, categories of data managed within data processing activities are inserted in the system. Them are mainly related to medical exams results managed by the hospital. Also, the full list, and details, of data processing activities of the hospital, and* 3^{rd} *parties, is collected.*''

For instance, the business analyst of the hospital collects, by using MM-REPA, the processing activities related to the interaction of the lab and the hospital concerning performing medical exams and sending the results to the hospital; related information will be populated in the DAM data categories such as ''Processing List"'' and ''Processing Description, Scope, Purpose and Legal Basis''.

DSM Flow: Phase 4 (DAC: Assessment Translation and Data Synthesis). This phase covers mainly **AS4**. Its activities are represented in Fig. 2. On the basis of all the data collected in the previous steps, in this phase the aim is to translate this data for creating data synthesis and data graphical representations of them, to facilitate the organization in understanding the current situation (self-assessment) both textually and graphically (**AS4**). This information will be also the baseline for important activities in the next phases. This phase is associated to the MM-Assess tool (Fig. 2) within the DAC component and the ATr module (Fig. 1). This phase does not require user intervention, it is completely automated by MM-Assess. However, business analysts will be able to see, and to use in the next steps, results produced here. This phase is illustrated by the following part of the storyline:

> ''*Then, on the basis of the answers, the platform produces a self-assessment of the organization, data synthesis and graphical representations.*''

For example, data synthesis elaborated and saved are hospital percentage of readiness and index of complexity (DAM data category ''GDPR Self-Assessment'').

DSM Flow: Phase 5 (DPAC: Data Protection Impact Assessment, Preliminary Threat Analysis and Data Minimization Analysis). This phase covers **AS5** and partially **AS6** and **AS4**. Its activities are represented in Fig. 2. The main objectives of this phase are to support the organization in performing DPIA (**AS5**), generating the GDPR Plan (**AS5**), conducting a preliminary Threat Analysis by collecting threats, security measures and revising/refining assets (**AS6**) involved and collected previously. Finally, in this phase the organization is supported also concerning data minimization analysis, through visual data synthesis and graphical representations (**AS4**). This phase is fulfilled by the MM-PIA tool (Fig. 2) within the DPAC component and the DPIA Analysis, Threats Analysis, and Data Minimization Analysis modules (Fig. 1). The user of the organization for this phase is typically a privacy/security analyst (Fig. 2), which could collaborate with the business analysts that used the DEFeND platform in the previous steps. Most of the activities performed here, are related to collection of information through questionnaires compilation for collecting information related to the goals outlined above, and automated activities for producing related results. Some results are shown in visual/graphical ways. Baseline information, collected in previous phases, are read by MM-PIA from the DAM Model, and information collected and generated here saved in the DPM model. This phase is illustrated by the following part of the storyline:

> ''*On the basis of data collected so far, and new data collected also in this step with further questionnaires, the system generates a DPIA and proposes a GDPR plan.*''

Regarding DAM and DPM models, for instance MM-PIA, reading from DAM, shows to hospital privacy/security analysts information regarding assets collected before (DAM data category ''Processing Assets''), and asks to revise/refine them by adding also other relevant information (saved in DPM by MM-PIA), via questionnaires, for collecting GDPR risks and vulnerabilities (data category ''Vulnerabilities'' in DPM) related to assets, privacy/security requirements to guarantee (e.g., confidentiality, integrity and availability of patient medical records, data category ''Privacy and Security Requirements'' in DPM) and potential threats that could attack them (e.g., illegitimate access to patient medical records, and malwares that could perform attacks affecting hospital computers, data category ''Threats'' in DPM) and security measures to apply (e.g., antivirus and firewalls, data category ''Security Mechanisms'' in DPM).

DSM Flow: Phase 6 (DPAC: Privacy/Security and Threat Analysis Based on Modelling and Privacy Patterns). This phase covers mainly **AS6** and partially **AS4** and **AS7**. Its activities are represented in Fig. 2. High-level goals of this phase concern to support the organization in performing GDPR Privacy/Security Analysis and Threat Analysis (**AS6**) based on Modelling (**AS6**, **AS4**, **AS7**) and Privacy Patterns (**AS6**). In detail, in DSM, this is performed via Organizational Structure Analysis, Data Mapping and Risk Models Analysis, Privacy/Security Requirements Analysis, Requirements Conflicts Analysis and Resolution based on Patterns, Threat Analysis, Attacks Analysis and Security Measures Identification based on Patterns. This phase is associated to the Secure

Tropos (SecTro) tool (Fig. 2), and its method, extended in DEFeND, within the DPAC component and the Privacy/Security and Threat Analysis modules (Fig. 1). Users of this phase are privacy/security analysts (Fig. 2). Activities performed during this phase concern modelling by using graphical editors showing models, where it is possible to add concepts and relationships from a palette to editors, according to semantic and syntactic constraints related to the modelling language and method behind, and being supported having the possibility to leverage on ready-to-use libraries of patterns. The SecTro method supports the analyst via modelling in different steps by focusing on different perspectives of the problem. Such perspectives are called views in SecTro, and are the: Organizational View, Data Mapping View, Privacy/Security View and Attack View. This phase is partially illustrated by the following extract of the storyline:

> ''*The platform, on the basis of the info collected, the assessment and the GDPR plan elaborated, shows graphical models of the Organizational Structure of the Hospital, with the main actors and interactions.*''

In fact, SecTro reads some of the information mentioned above by DAM and DPM models, and generates the organizational model in the Organizational View, where it is possible to perform organizational structure analysis. For instance, identifying main actors involved such as hospital departments, doctors, supervisors, the lab - as 3^{rd} party -, high-level interactions among them, processing activities, organization assets, initial privacy/security requirements occurring in the interactions, etc. Also the next storyline extract illustrates part of this phase:

> ''*On the basis of this, DEFeND users are able to identify the importance of fulfilling the confidentiality and integrity of patient medical record, through also validation processes, and to perform data mapping with organizational assets. Specifically, the hospital privacy/security analyst improves the graphical representation by modelling how a Doctor can change the patient medical record (for instance by adding exam results received by 3^{rd} parties as external labs) and obtaining a validation for them from a Supervisor.*''

This means that initial privacy/security requirements occurring in the interactions can be refined (e.g., confidentiality and integrity) by modeling validation processes related to data processing activities, and mapping organizational data assets involved (e.g., patient medical record and medical result) in the Data Mapping View. Also next storyline extract illustrates part of this phase:

> ''*Furthermore, the modelling helps also in identifying further important privacy/security requirements (e.g., accountability, anonymity, etc.) relevant also for performing threat analysis. Accordingly, the system helps a hospital privacy/security analyst in modelling potential threats that could affect confidentiality, integrity and availability of this important kind of data, and privacy and security measures that could mitigate/solve those potential problems. For instance, concerning threat analysis, a threat is modelled and considered regarding the possibility that the computer and web applications, used by the Doctor for changing the medical record, are affected by a malware, for example a Trojan.*''

Accordingly, in the Privacy/Security View the focus is deeply oriented on privacy/security requirements, potential requirements conflicts, threats and security mechanisms. In fact, the analyst can individuate vulnerabilities in the system, and by doing this deeper analysis, can identify even more privacy/security requirements to be satisfied. Here, the analyst can model at high-level potential threats affecting vulnerabilities, and use libraries of patterns, provided by SecTro, including security mechanisms for threats mitigation. Threats Analysis is done iteratively, at different levels of abstraction, by switching from the Privacy/Security View to the Attack View of each of the threats individuated.

Concerning DAM and DPM models, for example SecTro for generating the model of the organizational view can read from DAM the actors (''Organization Departments'', ''Employees, Roles and Responsibilities'' and ''3^{rd} Parties'' data categories in DAM), which in the storyline are the doctor, the supervisor and the lab. While, output of analysis activities regarding threats and attacks is saved in the DPM model. For instance, assets that could be involved in threats such as the computer, the patient medical record and the exam results are saved in DPM in the data category ''Privacy related Resources and Assets''.

DSM Flow: Phase 7 (DPAC: Threat Analysis for Continuous GDPR Risk Assessment and Compliance). This phase covers mainly **AS7** and partially **AS4** and **AS6**. Its activities are represented in Fig. 2. Goals of this phase are to support the organization in collecting organization technical information, refining IT assets configuration and configuring threats analysis (**AS7**, **AS4**, **AS6**), generating high-level configurations for IT threats monitoring (**AS7**, **AS4**), and creating the conditions for performing continuous GDPR risk assessment and compliance (**AS7**). This phase is satisfied by the RAE tool (Fig. 2) within the DPAC component and the Threat Analysis module (Fig. 1). Users of this phase are privacy/security analysts (Fig. 2). Activities performed regard collection of technical information via technical questionnaires compilation, automatic generation of high-level configurations for IT threats monitoring, verification and revision of them by analysts, and starting continuous GDPR risk assessment and compliance monitoring based on those configurations. Some information is read by DAM and DPM models, while information collected, generated and revised is saved in DPM. This phase is illustrated by the following storyline part:

> ''*The system, on the basis of the GDPR Self-Assessment, DPIA, Risk Assessment, Processes modelled for changing data and validating changes, Threats modelled, and additional technical information asked through technical questionnaires, generates monitoring configurations. A hospital privacy/security analyst read such configurations, and optionally improve them by adding further specific information. After all these complex analyses, the system is able to perform monitoring of threats for Continuous Model-Based GDPR risk assessment and Compliance.*''

Regarding DAM and DPM models, RAE can read from them some information. For instance, 3^{rd} parties information such as the lab for the hospital (data category ''3^{rd} Parties'' of DAM), and privacy/security requirements the hospital

should fulfil such as confidentiality, integrity, availability, accountability and anonymity (DPM data category ''Privacy and Security Requirements''). Such information is used by RAE, together with other technical information collected in this phase, for executing automated activities, and to support the manual activities of the privacy/security analyst of the hospital. For example, RAE collects, through IT technical questionnaires, IT monitoring configuration such as IT assets and their IP addresses. Thus, in this step the analyst is guided, and supported, in refining IT assets , and to configure threat analysis monitoring (DPM data categories ''Risk Information'' and ''Risk Mapping''). RAE, on the basis of all this information, generates high-level configurations for IT threat monitoring, and asks the analyst to verify and potentially refine such configuration These steps create the conditions for performing Continuous Model-Based GDPR Compliance, by reiterating the previous phases in a systematic way.

3.3 Evaluation

Having described the DSM service in the previous sections, here we present our preliminary evaluation. First, we present our evaluation strategy towards the evaluation of the DSM service, and then the obtained results.

Evaluation Strategy. PbD activities and strategies presented in this paper are inherently human-centred activities. From collecting organisational and 3^{rd} parties information, identifying assets and processing activities through data minimisation, DPIA, threat analysis, and continuous risk assessment, the inputs, processes, and outputs are primarily created, performed, and evaluated by humans. For this reason, we used humans for our preliminary evaluation of our research claims, and in particular members of the pilot organisations that participate in the DEFeND project. Therefore, we had users that work in the healthcare, banking, energy, and public administration sectors. Our user evaluation was descriptive, artificial, and qualitative. Descriptive, because it involved asking participants questions about their experiences, artificial because we created artefacts and context for the purposes of the user evaluation, and qualitative because it was aimed at establishing how well the methods and tools fit the needs and culture of organisations. In particular, we created a storyline that covered all the features of the methods, and the toolkit that were demonstrated, and created some artificial data for demonstration purposes. The user evaluation was carried out in three iterations. Three physical workshops were held, where the methods and tools were demonstrated, in order to receive feedback from the participating users, and incorporate the feedback in the subsequent versions of the method and toolkit.

Evaluation Results. Inline with our RQs, Participants in our evaluation were asked whether the method and toolkit, demonstrated to them, would likely be appropriate to support them concerning the execution of complex PbD activities for GDPR compliance. In the next, we summarize some of the descriptive questions made to participants: **(i)** To what extent do the proposed AS are the ones required and relevant for PbD GDPR compliance? **(ii)** To what extent do the proposed flow, demonstrated with the toolkit, offers a structured method for PbD GDPR compliance? **(iii)** To what extent do the automation and guidance,

provided by the toolkit, is appropriate, clarify how to perform PbD GDPR compliance, and provide support for this? The three iterations of user evaluation, which we performed, enabled us to gain insights which we may not otherwise have had. In general, the results of the user evaluation exercises were favourable. In each physical workshop the participating users expressed their confidence that their needs are satisfied by the features of the method and of the toolkit. However, they expressed concerns and criticisms about the usability and look and feel of the toolkit. This can be explained as the service was not fully integrated to the whole DEFeND platform and was lacking the full final user interface.

4 Related Work

4.1 Industry Comparison

The EC-funded H2020 project cyberwatching.eu has launched the GDPR Temperature Tool, to help European SMEs understand just how at risk they are to sanctions or fines [4]. By answering a set of questions on data protection, the Tool provides an indication of a company's risk to sanctions. In addition, a free customised set of recommendations is provided. However, the provided recommendations are too generic and not specific to the company. According to the 2019 Privacy Tech Vendor Report from IAAP [12], the number of vendors providing privacy management tools is constantly increasing, although as the report highlights ''there is no single vendor that will automatically make an organization GDPR compliant'' [12]. The IAAP's report classifies the solutions into 2 key categories: Privacy Program Management and Enterprise Privacy Management. The first are grouped into 6 subcategories: assessment managers, consent managers, data mapping, incident response, privacy information managers and website scanning. The second are grouped in 4 subcategories: activity monitoring, data discovery, de-identification/pseudonymity and enterprise communications. None of the listed vendors is able to provide solutions that cover all sub-categories. Differently than the tools presented in the report, DEFeND and DSM cover a much wider set of subcategories. Forrester [3] released a report evaluating the 12 most significant providers in the market of EU GDPR compliance and privacy management. Platforms are evaluated against 10 criteria. One important conclusion of the report is that a functionality such as data discovery across systems, is a key feature to avoid bad consequences of doing such task manually (i.e. inaccuracies, guesswork), and increases assurance in terms of accountability. DSM supports this functionality via the Organization Data Collection module, where organizational data is collected and automatically transformed to a Data Assessment Model.

4.2 Research Novelty

This section briefly discusses literature and research challenges in areas associated with the Data Scope Management, and describes how DSM addresses them.

Privacy by Design (PbD). PbD is an important principle of GDPR (Data Protection by Design and by Default), but only few efforts exist to support practical implementation of PbD [2,6,7,14]. The Data Scope Management service facilitates the implementation of PbD principles using methods and techniques from privacy requirements engineering, and privacy design.

Privacy Impact Assessment (PIA). Systematic assessment of privacy-related risks, in the form of PIA, is requested by GDPR (art. 35). PIA shall be embedded in the early phases of software design and development. PIA adoption in most industry sectors is considered at an early stage [13], while state of the art methodologies and tools to implement PIA are very few (e.g., [5]. The DEFeND DSM service advances the current state of the art in PIA by providing an in-depth processing analysis based on a recognized methodology and international standards. DSM integrates PbD approaches with PIA and threat analysis at planning level, to provide organisations with the abilities to check GDPR compliance, measure and review their privacy level, analyse safeguards and security measures for mitigating potential risks, but also with the capability to develop new services and systems in accordance with GDPR.

5 Conclusions

In this paper, we presented a set of Activities and Strategies (AS) for Privacy by Design (PbD), a flow and a toolkit, DSM (the Data Scope Management service of the DEFeND EU project platform), supporting them, for carrying out major activities for PbD GDPR compliance. The need to individuate the most relevant AS, and designing a PbD flow for them, derives from the fact that organizations are facing many difficulties regarding interpreting GDPR and properly understanding how to apply it. Our DSM flow, presented here, provides organizations with a clear, coherent, linear flow of activities, and method, for performing GDPR compliance in a PbD fashion. Furthermore, it is missing, from the literature and the industry, a complete toolkit supporting the organization in performing, in automated ways, such complex PbD activities. We individuated candidate tools, fulfilling isolated GDPR aspects, extended and integrated them for developing the DSM service, as a comprehensive toolkit, compliant with the DSM flow we designed, and able to automate PbD activities to support organizations for continuous model-based GDPR compliance. To evaluate our proposed method, toolkit, and flow, we organised 3 workshops and performed a qualitative user survey evaluation. During the workshops the DSM service was demonstrated to pilots from the healthcare, banking, public administration and energy sectors, and feedback was collected. The feedback was favourable, as the organisations' responses were that the features of the method, toolkit, and flow satisfy their needs and have the potential to support them for a systematic and structured PbD GDPR compliance. As future work, we plan to deploy the whole DEFeND platform at the pilots' infrastructures, and assess the effectiveness of DSM by carrying out quantitative and qualitative case study evaluations.

Acknowledgments. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 787068.

References

1. Blank, S.: The four steps to the epiphany: successful strategies for products that win. John Wiley & Sons (2007)
2. Deng, M., Wuyts, K., Scandariato, R., Preneel, B., Joosen, W.: A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requirements Engineering Journal (2011)
3. The forrester new wave™, https://www.forrester.com/report/The%20Forrester%20New%20Wave%20GDPR%20And%20Privacy%20Management%20Software%20Q4%202018/-/E-RES142698
4. The gdpr temperature tool, http://gdprtool.cyberwatching.eu/Pages/Home.aspx
5. Horák, M., Stupka, V., Husák, M.: Gdpr compliance in cybersecurity software: A case study of dpia in information sharing platform. In: Proceedings of the 14th International Conference on Availability, Reliability and Security (2019)
6. Kalloniatis, C., Belsis, P., Gritzalis, S.: A soft computing approach for privacy requirements engineering: The pris framework. Applied Soft Computing (2011)
7. Kurtz, C., Semmann, M., et al.: Privacy by design to comply with gdpr: a review on third-party data processors. In: Americas Confer. on Information Systems (2018)
8. Maguire, M.: Methods to support human-centred design. International journal of human-computer studies (2001)
9. Mouratidis, H., Argyropoulos, N., Shei, S.: Security requirements engineering for cloud computing: the secure tropos approach. Springer (2016)
10. Piras, L., Dellagiacoma, D., Perini, A., Susi, A., Giorgini, P., Mylopoulos, J.: Design Thinking and Acceptance Requirements for Designing Gamified Software. In: 13th Intern. Confer. on Research Challenges in Information Science (RCIS). IEEE (2019)
11. Piras, L., Al-Obeidallah, M.G., Praitano, A., Tsohou, A., Mouratidis, H., Crespo, B.G.N., Bernard, J.B., Fiorani, M., Magkos, E., Sanz, A.C., et al.: Defend architecture: A privacy by design platform for gdpr compliance. In: International Conference on Trust and Privacy in Digital Business (TrustBus). Springer (2019)
12. Privacy tech vendor report, https://iapp.org/resources/article/2019-privacy-tech-vendor-report/
13. Rantos, K., Drosatos, G., Demertzis, K., Ilioudis, C., Papanikolaou, A., Kritsas, A.: ADvoCATE: a consent management platform for personal data processing in the IoT using blockchain technology. In: International Conference on Security for Information Technology and Communications. Springer (2018)
14. Romanou, A.: The necessity of the implementation of privacy by design in sectors where data protection concerns arise. Computer law & security review (2018)
15. Tsohou, A., Magkos, E., Mouratidis, H., Chrysoloras, G., Piras, L., Pavlidis, M., Debussche, J., Rotoloni, M., Crespo, B.G.N.: Privacy, Security, Legal and Technology Acceptance Elicited and Consolidated Requirements for a GDPR Compliance Platform. Information and Computer Security Journal (2020)
16. Tsohou, A., Magkos, E., Mouratidis, H., Chrysoloras, G., Piras, L., Pavlidis, M., Debussche, J., Rotoloni, M., Crespo, B.G.N.: Privacy, Security, Legal and Technology Acceptance Requirements for a GDPR Compliance Platform. In: Int. Workshop on Security and Privacy Requirements Eng. (SECPRE). Springer (2019)