Journal of Data Intelligence, Vol. 2, No. 2 (2021) 136--165

A DATA SCOPE MANAGEMENT SERVICE TO SUPPORT PRIVACY BY DESIGN AND GDPR COMPLIANCE[a]

LUCA PIRAS
School of Computing, Robert Gordon University, Aberdeen, United Kingdom
l.piras@rgu.ac.uk

MOHAMMED GHAZI AL-OBEIDALLAH
Faculty of Computer Science and Informatics, Amman Arab University, Amman, Jordan
m.obeidallah@aau.edu.jo

MICHALIS PAVLIDIS
Centre for Secure, Intelligent and Usable Systems, University of Brighton, Brighton, United Kingdom
m.pavlidis@brighton.ac.uk

HARALAMBOS MOURATIDIS
Centre for Secure, Intelligent and Usable Systems, University of Brighton, Brighton, United Kingdom
h.mouratidis@brighton.ac.uk

AGGELIKI TSOHOU
Faculty of Information Science and Informatics, Corfu, Greece
atsohou@ionio.gr

EMMANOUIL MAGKOS
Faculty of Information Science and Informatics, Corfu, Greece
emagos@ionio.gr

ANDREA PRAITANO
Maticmind SpA, Rome, Italy
andrea.praitano@maticmind.it

In order to empower user data protection and user rights, the European General Data Protection Regulation (GDPR) has been enforced. On the positive side, the user is obtaining advantages from GDPR. However, organisations are facing many difficulties in interpreting GDPR, and to properly applying it, and, in the meanwhile, due to their lack of compliance, many organisations are receiving huge fines from authorities. 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 propose the design of such flow, and our comprehensive supporting toolkit, as part of the DEFeND EU Project platform. 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 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

[a]This paper is an extended version of the work published in the TrustBus 2020 International Conference [27]

1. Introduction

The European General Data Protection Regulation (GDPR) is important because it improves the protection of European data subjects' rights and clarifies what companies that process personal data must do to safeguard these rights. GDPR is posing a major challenge for organisations [16], as they need to comply with a large number of requirements including data classification, recording of data processing activities with reporting and registers, data monitoring, breach detection and notification, fast intervention and fast data deletion. Organisations failing to comply with GDPR are liable to financial fines [31]. One of the most challenging and difficult principles to adhere with is Data Protection by Design and by Default [12]; hereafter, for the sake of simplicity, we refer to these principles 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. When developing new systems and services, organisations do not have a structured way to ensure that PbD is followed [13]. 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.[b] However, even those guidelines helping in reducing such gap, they are still at high-level. 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.

In this article, a novel structured framework and a toolkit that fulfils this gap of the current state of the art will be presented. The Data Scope Management (DSM) solution presented is part of the DEFeND EU Project[c] platform [30], and builds on previous work presented at TrustBus-19 [30] and TrustBus-20 [27]. In particular, this paper addresses the following Research Questions (RQs):

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

RQ2: Can PbD analysis and implementation activities (**RQ1**) being automated and supported by software tools?

RQ1 is the main RQ that this paper tries to answer while **RQ2** is a supportive question. Information from Data Protection Officers (DPOs), experts and end-users [35, 36] of organizations from different GDPR relevant sectors (e.g., banking, public administration, healthcare, energy) has been elicited to answer the first RQ. In addition, 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 and extended a number of tools to provide automated support to DSM.

The European General Data Protection Regulation (GDPR) has been enforced for many important reasons. The main reason has been to empower the data subjects' data protection and rights. Before the advent of GDPR, there were many problems related to having distinct and different national regulations among the EU Countries [7], giving uncertainty on guaranteeing coherent EU citizen rights all over the Europe [31]. Such differences led to significant problems

[b] https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201904_dataprotection_by_design_and_by_default.pdf
[c] https://cordis.europa.eu/project/id/787068

such as uncontrolled exchange, manipulation and exploitation of user data with important drawbacks and impact for the EU citizens. The lack of a ''standard'' in Europe, for these crucial aspects, created problems also to organizations interested in enlarging their market [7]. For instance, companies having businesses in a EU country had to be compliant with the specific national data protection regulation of that Country, and if they wanted to extend services to more EU Countries, they would have had to cope with different regulations for selling the same products, by dealing differently with personal data of their customers [10]. This had significant impact, for organizations, concerning the difficulty on tackling a complex reality, requiring additional high costs, such as different expertise involved, and different software systems needed for dealing with the heterogeneous situations [31].

GDPR overcomes most of these problems (and others not mentioned above, for the sake of space), by offering a single data protection regulation across Europe. However, although the citizen (i.e. the data subject) is obtaining many advantages from GDPR, such as enriching, guaranteeing and enforcing her rights, organisations are facing many difficulties, and increasing costs, on interpreting GDPR and understanding exactly how to properly apply it. Accordingly, many organisations, due to their lack of compliance with GDPR, are receiving huge fines from authorities [31]. In fact, GDPR demands organizations to guarantee compliance regarding many aspects, for example: data classification, recording of data processing activities with reporting and registers, data monitoring, breach detection and notification, fast intervention and fast data deletion. The problem is that GDPR in many requirements does not provide clear and detailed information on how to implement them in practice, and does not suggest what specific tools to use, remaining at a very abstract and interpretable level, as it happened also in other important cases related to the Law and IT fields which needs to interoperate [5,19,34]. Accordingly, achieving GDPR compliance is a big challenge [16] that requires to translate such principles [11] in many complex activities, for instance: conducting self-assessments, identifying data processing activities, and identifying the involved data controllers, data processors and 3rd parties, categorising data to understand how to properly manage the different typologies, performing data minimization analysis, guaranteeing a continuous risk assessment, responding to data breaches. Furthermore, one of the most challenging and difficult principles/approaches to adhere with is: Privacy by Design [12] and by default (PbD). Indeed, all the activities implied by GDPR are required to be performed with a PbD approach. Also in this case, unfortunately, GDPR does not provide enough and specific guidance. Therefore, this further complicates the work of organizations, which have to deal, in a PbD way, with unclear, interpretable GDPR concepts [11].

Recently, in order to try to cover this important lack of practical guidance on how to effectively apply GDPR with a PbD approach, the European Data Protection Board (EDPB), released an official document for providing PbD guidelines. However, those guidelines, even helping in reducing such gap, are still at high-level, offer a few practical indications, and miss completely to supply organization with a technical toolkit for implementing GDPR. Specifically, organizations are still facing difficulties in identifying a flow for executing in a coherent, linear and effective way the heterogeneous activities mentioned before (and others not indicated, in fact the ones indicated above are only a subset of all the activities needed), and, above all, a complete supporting toolkit. In fact, available tools, from the literature and the industry, are individually able to fulfil only specific GDPR aspects. This calls for the need

of a complete guiding toolkit for GDPR compliance, covering all these aspects in a coherent, supportive flow.

In this paper, we propose the design of such flow, and our comprehensive supporting toolkit called DEFeND Data Scope Management service (DSM). We evaluate our proposed method, toolkit, and flow, through an empirical investigation with three workshops and a qualitative user survey. DSM is part of the DEFeND EU Project[d]platform [30]. Specifically, this work is a follow-up of the papers [30] [27].

The rest of the paper is organized as follows. Section 2 summarizes the requirements we elicited in previous works [35, 36], 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

DEFeND is an Innovation Action project, and as such its main focus will be on improving existing software tools and frameworks and developing new 'integration software', driven by market needs, to deliver a unique organizational data privacy governance platform. The architecture of DEFeND is composed of 5 services. These services are Data Scope Management (DSM) service, Data Process Management (DPM) service, Data Breach Management (DBM) service, GDPR Planning service, GDPR Reporting service. The first service of the DEFeND Platform is the DSM service. DSM assists the organizations to collect, analyze and operationalize different aspects and articles of the GDPR. Furthermore, DSM service supports organizations in performing GDPR Self-Assessments by collecting organizational information (also related to 3^{rd} parties), data processing activities, and creating a profile of the organization. This profile involves different organization aspects such as legal, economic and financial aspects. DSM service enables organizations in executing Data Protection Impact Assessment (DPIA) by collecting/revising and refining organizational assets and by elaborating other information collected for supporting the organizations with data synthesis and graphical representations through DSM tools. In addition, The DSM service provides appropriate reporting capabilities. DSM helps organizations to perform Threats Analysis, Data Minimization Analysis, Privacy-by-Design based analysis, design tool-supported modelling techniques, and continuous risk assessment. The users of the DEFeND platform, based on their roles in the organization and their expertise in privacy policies, can be classified into three types: Business Analyst, Security and Privacy Analyst, and Platform End-user.
We employed a Human-Centered Design (HCD) approach [18], where questionnaires and interviews were used as the basic tool to capture the main stakeholders' requirements with regards to PbD [35, 36]. Our approach consisted of 3 main stages [35, 36]. These stages are presented below.

Questionnaire Preparation. The internal and external key stakeholders, such as DPOs, IT managers, and citizens, were identified. For each user category, a questionnaire was prepared in a systematic way [35, 36], aiming to capture the legal, functional, security, privacy

[d] https://cordis.europa.eu/project/id/787068

and technology acceptance needs [28, 29]. Specifically, we followed the approach of [4] for customer development, including steps such as Customer Segmentation, Problem Discovery and Validation, Product Discovery and Validation [35, 36]. Two online questionnaires were prepared: 1 for end-users[e] and 1 for citizens[f].

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 [35, 36].

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

2.1. *Identified Activities and Strategies for PbD*

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

AS1: *Organization Situation and Context.* 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.Therefore, in the very first phases, it is needed to create a high-level, big picture of the organization current situation, by identifying also its context and relations with other potential organizations, 3^{rd} parties, involved in personal data management; accordingly, the first activities should led to the collection of preliminary information on the organization, its context and 3^{rd} parties.

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.* Conducting a deep analysis on data processing activities performed by the organization itself, and in collaboration with 3^{rd} parties is also needed. In addition, the identification of data categories and assets involved should be included.

AS4: *GDPR Data Syntheses, Graphical Representations and Model-Based, Visual Support.* It is beneficial to provide support and guidance with graphical representations and synthesis

[e] https://ec.europa.eu/eusurvey/runner/DEFeNDEndUser
[f] https://ec.europa.eu/eusurvey/runner/DEFeNDCitizens

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, such as Data Protection Impact Assessment, data minimization analysis, and creation of GDPR action plans. Other critical activities, such as Privacy/security analysiss, threat analysis, and continuous risk assessment configurations, could be performed and 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.* Needs to be enacted by further critical analysis supported by visual model-based techniques enhanced and adapted for GDPR. This concerns analysis of the organization context, data/assets/accountability mapping with also analysis of risks, threats and measures in place, privacy/security requirements constraints and conflict resolution, supported via libraries of patterns and modeling techniques specifically designed for GDPR.

AS7: *Continuous Model-Based GDPR Compliance.* On the basis of analyses performed for the previous AS, it is needed to support the organization to: **(i)** have software systems able to put in place GDPR compliance solutions individuated, **(ii)** receive automated support for configuring such systems, **(iii)** monitor continuously the compliance, according to the GDPR plan, for identifying new potential lacks with GDPR and data breaches, **(iv)** enable the organization to react to such problems, and **(v)** 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.

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. In the following we introduce the main needs, objectives we identified for our example and the related context with the different actors, roles, processes and data involved. One of the most critical aspects for a hospital is to manage the patient medical record and to have verifications, from a supervisor, 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.

Storyline Execution (Usage of DSM within the DEFeND Platform). In the following, we describe the scenario related to the usage of DSM within the DEFeND platform. 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. 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. 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. Then, on the basis of the answers, the platform produces a self-assessment of the organization, data synthesis and graphical representations.

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.

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.

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 medical exam results received by 3^{rd} parties as external labs) and obtaining a validation for them from a Supervisor. The system helps also in modelling the data mapping with organizational assets, identifying the different data categories managed by the organization, and assigning data retention periods to them. 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. 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.

2.2. *DSM Candidate Tools*

MM-Assess. The MM-Assess (MaticMind-Assess) tool supports the business analyst to conduct a Self-Assessment for the organization. This self-assessment is based on a well-structured questionnaire to evaluate the status of the organization in relation to the relevant parts of the

privacy laws of data protection and information security. The Data Protection Assessment is designed to evaluate the organization at a periodic level. It is a closed questionnaire where the result of the assessment could be compared with previous result(s) to evaluate the progress of the organization related to the data protection. The Data Protection Assessment questionnaire is structured in different sections and subsections and each subsection is composed by one or more questions. The questions are based on the European Data Protection Framework and other connected best practices and standards. The MM-ASSESS tool, as a first service tool of the Data Scope Management (DSM), can assist organizations to collect, analyze and operationalize different aspects and articles of the GDPR. It also provides appropriate reporting capabilities.

MM-REPA. MM-REPA (MaticMind Record of Processing Activities) is a tool that creates a list of all data processing activities in the organization based on a guided questionnaire. This questionnaire provides, from a template list, all common and generic processing activities in an organization. The tool basically provides a register of processing activities of the organization. The MM-REPA tool supports the Data Assessment Component (DAC), as a first component of the DSM, to collect data processing activities, listing all these processes, and collecting the detailed information for the organization, departments and third parties. During the collection of data processing activities it is possible to identify and detail the type of personal data managed/processed by the organization, departments and third parties. Asset information and data category, used and processed in the processing activity, are also collected. Moreover, the MM-REPA tool contributes, in relation to planning level, to the data minimization analysis as a part of the Data Scope Management of the DEFeND Platform.

MM-PIA. The MM-PIA is a Risk Assessment Management (RAM) tool that 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 main functionality of this tool is to enable organizations to measure and review their privacy level and when necessary propose design changes. Moreover, MM-PIA includes the safeguards and privacy/security measures for mitigating potential risks. The MM-PIA tool performs a DPIA by using and analyzing the risk patterns of the Data Scope Management and can contribute to conduct privacy risk assessments. The MM-PIA tool enables organizations to measure and review their privacy level and when necessary propose design changes.

SecTro. SecTro tool is a CASE tool that guides and supports the developers in the construction of the appropriate models of Secure Tropos. Secure Tropos is a security requirements engineering methodology that considers security throughout the whole development process. The approach identifies, models and analyses the security issues from the early stages of software development within the organization. The main functionalities of the SecTro are to support the developer in the modelling activities of Secure Tropos. Therefore, the tool enables the developer to perform security reference modelling, security constraint modelling, secure entities modelling, and secure capability modelling. The last steady version of SecTro before the commence of DEFeND included 5 views: Organisational, Security Requirement, Attack, Cloud and Trust [7, 20--26]. The steady version has been extended to fulfil the requirements

of DEFeND project at the DSM service level [1, 27, 30].

RAE. Risk Assessment Engine (RAE) is an Atos software tool that supports organizations in the assessment of cyber-risk and was originally developed in the context of H2020 project WISER. The design and architecture of the RAE is described in [37]. The tool combines different types of inputs, known as indicators, and executes risk model-based algorithms based on risk patterns derived from well-known and widely used libraries. In DEFeND, RAE is a tool that supports organizations in continuously monitoring the risk exposure of critical assets, i.e. those involved in the processing of personal data. Through the analysis of the risk patterns in the context of the organization infrastructure, RAE helps organizations understanding how threats affect confidentiality, integrity and availability of the personal data processed by the organization; and instruct security administrators on which specific mechanisms and configurations must be deployed in the infrastructure in order to monitor for potential attacks and privacy incidents.

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

We have designed a flow for Activities and Strategies (AS), and developed a novel service, the Data Scope Management service (DSM), for the DEFeND platform to support PbD. 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-supporhttps://www.overleaf.com/project/604669168101c0dd6e82544eted 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.

3.1. *DSM Components, Integrated Tools and Data Models*

The candidate tools have been individuated, extended and integrated, 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, MM-REPA (MaticMind Record of Processing Activities), MM-PIA, a Risk Assessment Management (RAM) tool, the SecTro tool, and the Risk Assessment Engine (RAE) tool.

3.2. *Position of DSM in the DEFeND Architecture*

Figure 1 shows the DEFeND Platform Architecture [30].

The DEFeND architecture is composed of, starting from the top to the bottom of the

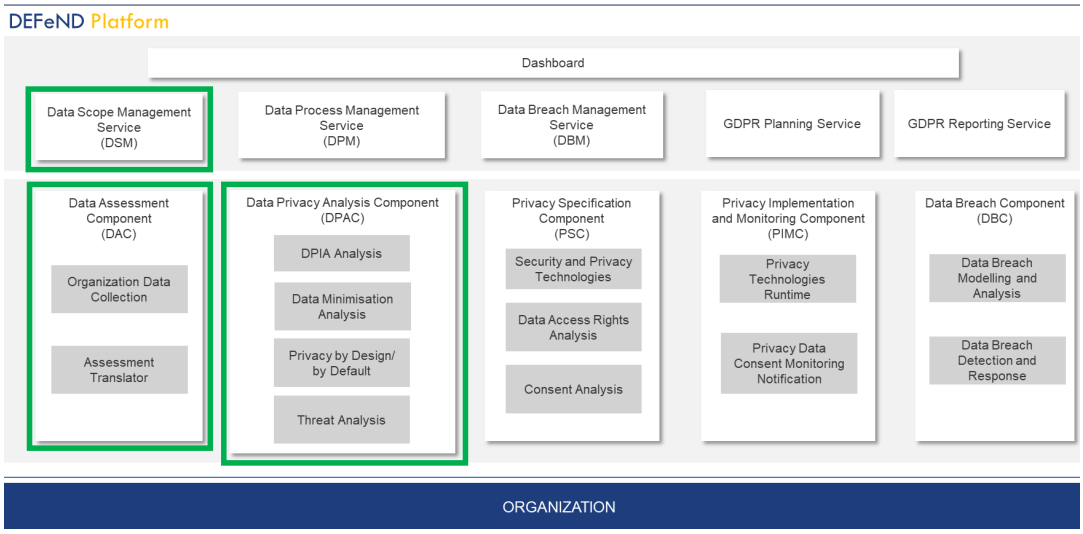

Fig. 1. DSM Position in the DEFeND Platform [27,30] (DSM components and modules surrounded by green rectangles)

figure, the DEFeND Dashboard, the five DEFeND services, i.e. the DSM service, Data Process Management (DPM) service, Data Breach Management (DBM) service, GDPR Planning Service, GDPR Reporting Service and all the components included in the DEFeND service. Often GDPR Planning Service and GDPR Reporting Service offer functionalities and fulfil requirements that are strictly related to the Dashboard. Therefore, hereafter, when we mention the dashboard, we are referring also to functionalities supported by these 2 services. According to Figure 1, DSM is represented by the elements surrounded by green rectangles. Therefore, DSM is composed of the Data Assessment Component (DAC) and the Data Privacy Analysis Component (DPAC) [27,30]. In turn, 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.

Figure 2 indicates the tools that have been assigned to the different DSM components and that have been extended for satisfying DSM objectives.

Specifically, DAC is fulfilled by MM-Tools such as MM-Assess and MM-REPA. DAC is constituted also of RAE. DPAC is composed of MM-PIA, SecTro and RAE. Figure 2 shows also, on the right, the transversal services for the DEFeND platform, providing transversal functionalities (e.g., Authentication, Authorization, logging, etc.) for multiple services of the DEFeND Platform such as DSM in this case. It is worth to mention that DSM, as other services of the DEFeND platform, benefits from these transversal services, however them are not covered in this document, being not exclusively related to DSM. Specifically, DAC [27,30] supports the organization, via the ODC module, for conducting GDPR Self-Assessments by collecting organizational information (also related to 3^{rd} parties), data processing activities with data categories and assets involved, and creating a profile of the organization regarding multiples perspectives such as legal, economic and financial aspects. Furthermore, it elaborates, via the

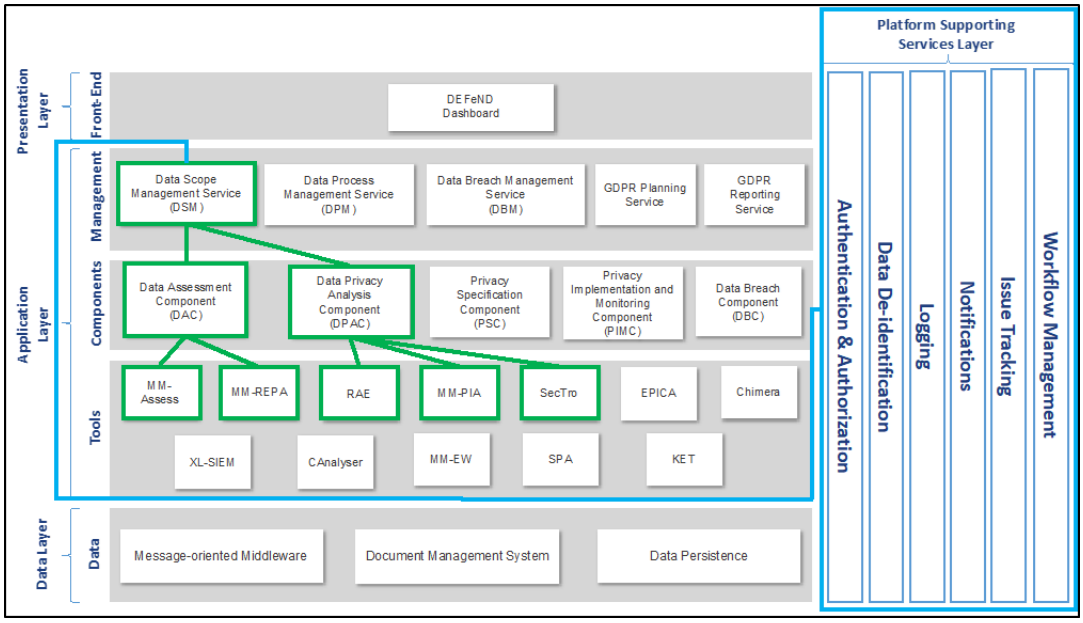

Fig. 2. DSM Position in the DEFeND Platform: DSM Components and Tools [27, 30] (items surrounded by green rectangles) and transversal services [27, 30] (indicated with light blue)

ATr module, information collected and produces data synthesis and graphical representations of them. While, DPAC [27,30] enables organizations, via the DPIA Analysis module, in executing DPIA by revising and refining organizational assets collected in DAC, and by producing related data synthesis and graphical representations. Moreover, DPAC helps organizations also in performing Threats Analysis and continuous Risk Assessment, through the Threats Analysis module, and, with the Privacy/Security Analysis module, supports the organization in using analysis and design tool-supported Modelling techniques for privacy/security analysis, and analysis on Data Minimization through the Data Minimization Analysis module.

3.2.1. *DSM Service Capabilities*

The following one is the list of capabilities that DSM offers, as a service, within the DEFeND platform architecture:

- Get Organization Information. Return the organizational information, collected by the DSM service, including a complete organization profile with economic, financial and legal aspects, data processing activities, data categories, assets and data minimization analysis information.
- Get 3^{rd} Parties Information. Return the 3^{rd} parties' information of the organization, collected by the DSM service, including complete 3^{rd} parties' profiles with economic, financial and legal aspects, data processing activities, data categories, assets and data minimization analysis information.
- Get GDPR Self-Assessment. Return the GDPR Self-Assessment of the organization

including data synthesis and graphical representations.

- Get Data Protection Impact Assessment (DPIA). Return the DPIA of the organization.
- Get GDPR Plan. Return the GDPR Plan.
- Get Data Minimization Analysis Results. Return the information collected on data minimization analysis performed by the organization before using the DEFeND platform, and the results of the data minimization analysis performed by using DSM.
- Get Privacy/Security Analysis and Threat Analysis Results. Return results concerning organizational structure, data mapping, risk models, privacy/security requirements, threat, and attacks analyses performed by using DSM.
- Get Continuous Risk Assessment Configurations. Return organization technical information collected, IT assets monitoring configuration, IT threats monitoring, and continuous risk assessment configuration.

3.3. *DSM, Data Models, Case Study and PbD Flow*

Here, we describe more in detail how the DSM service tools interact. Specifically, we illustrate the DSM Flow, related to the usage of DSM by an organization, by showing data collected, exchanged, analysed and elaborated by the different tools in different phases of the flow. Within the phases we use also the storyline for providing concrete examples for the healthcare domain. Therefore, the storyline is used only as a motivational example for giving practical examples to the complex DSM tools interactions. Therefore, the DSM Flows and phases that follow describe the full details of DSM at the service level. The storyline enables the reader to understand better, through explanatory examples, complex aspects of DSM even it is related to the domain of the healthcare. Furthermore, the storyline has been used also during our evaluation involving DEFeND pilots from the healthcare, banking, public administration and energy sectors.

Interactions of the DSM tools are made through the exchange of information stored in data models as shown in Figure 3. Therefore, data models involved 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 [27, 30]. Concerning DSM components and modules (Figure 3), 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.

Figure 4 gives the high-level overview of how these models are used, however we have design DSM in a flexible way that could allow other services to leverage on DAM and DPM models. These models, as the others of the DEFeND Platform, are stored in the DEFeND DB and can be reached by the tools through REST API Services. Other components of the DEFeND platform, which are used by the DSM tools, for data exchange and for fulfilling

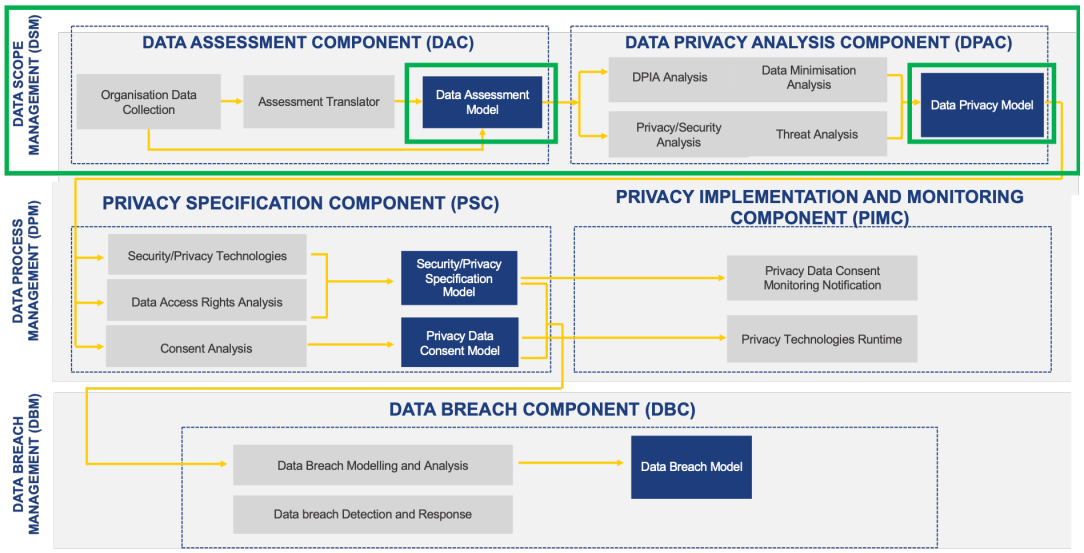

Fig. 3. DSM components, modules and DSM Data Models [27, 30, 35, 36] (green rectangles)

some of the related workflows, are a Document Management System (a version of Alfresco[g] that we customized for DEFeND) and a Workflow Management System (a version of Liferay [h] that we customized for DEFeND in relation to workflow management and for the Dashboard development). Figure 4, DAM and DSM are shown with a high-level representation of some of the most important data categories managed by them.

For example DAM supports the collection of information related to the list of data processing activities of an organization ("Processing List" in Figure 4), the departments of the organization that execute those processing ("Processing Departments" in Figure 4) and the data categories involved in these processing ("Data Categories" in Figure 4). In the next Subsection, we will give also some examples related to Data Models, together with the description of the Case Study, DSM Flow and our Storyline.

3.4. *Case Study, Storyline and DSM PbD Flow*

A storyline has been used to represent the most important PbD activities of DSM. In addition, this storyline has been used for demonstrating and discussing DSM, and our approach, with pilots from the banking, healthcare, public administration and health sectors, within the DEFeND Project[d]. 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. Figure 6 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.

In the following subsections, we outline the DSM Flow, phase by phase, with the aim of

[g] https://www.alfresco.com/ecm-software/document-management
[h] https://www.liferay.com/

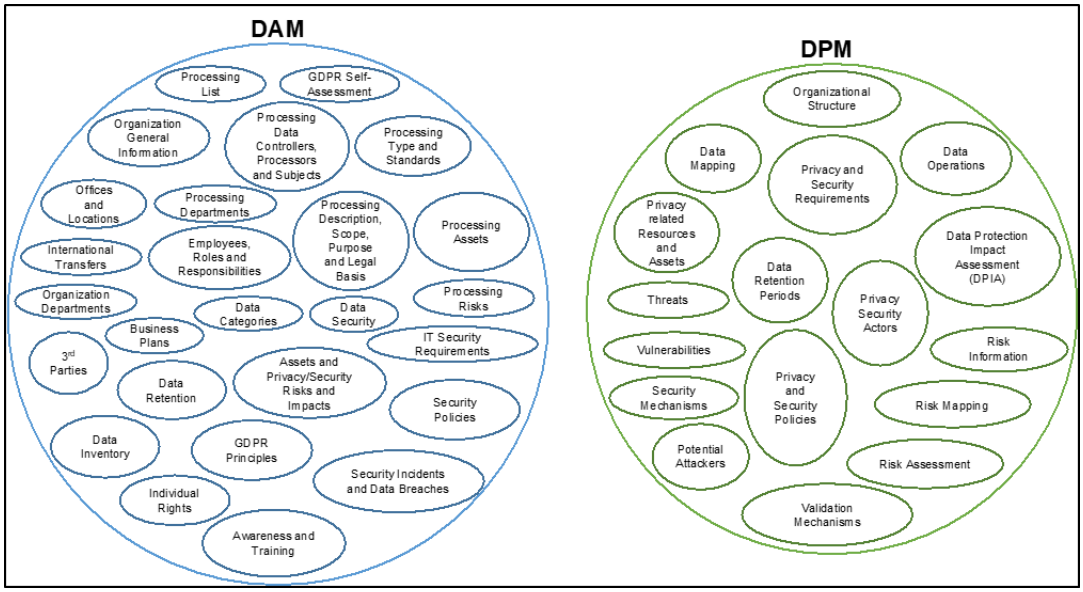

Fig. 4. The High-level Representation of Data Categories Managed by the Data Assessment Model (DAM) and the Data Privacy Model (DPM) of DSM [27, 30]

describing the interaction of the DSM tools based on information exchanged through the data models, by giving practical examples also related to the storyline.

In Figure 5 the DSM flow is represented with its phases and the tools involved, while Figure 6 represents the DSM flow as an activity diagram.

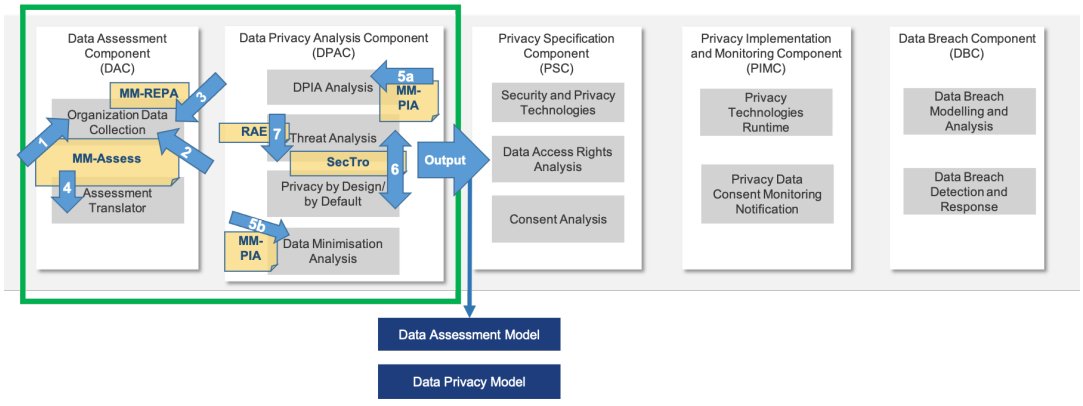

Fig. 5. The DSM Flow Phases and the Tools [27, 30]

Storyline Introduction. By using the DEFeND DSM service, a hospital wants to improve its GDPR compliance. 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

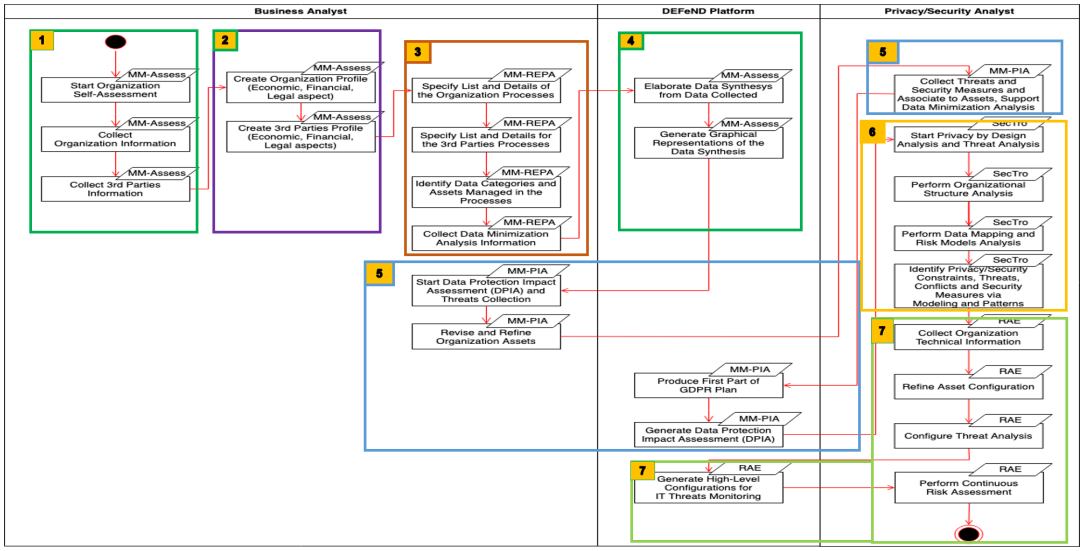

Fig. 6. Activity Diagram of the DSM Flow [27, 30]

organizations from heterogeneous sectors. In the following we introduce the main needs, objectives we identified for our example and the related context with the different actors, roles, processes and data involved. One of the most critical aspects for a hospital is to manage the patient medical record and to have verifications, from a supervisor, for any changes happening to it, such as adding a new medical exam result, 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.

DSM Flow: Phase 1 (DAC: Initial Organization Data Collection). Phase 1 covers mainly **AS1** and partially **AS2**. Its activities are represented in Figure 7. 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 (Figure 6) within the DAC component and the ODC module (Figure 3). The user of the organization for this phase is typically a business analyst (Figure 6). 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.*"

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}

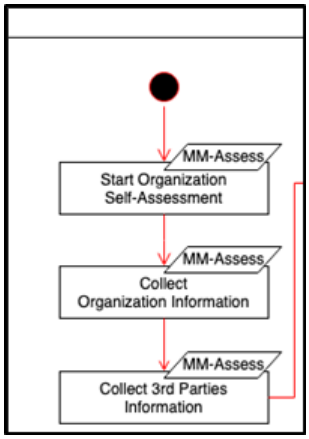

Fig. 7. DSM Flow Activities of Phase 1 [27,30]

Parties''.

Other examples of data collected in DAM during this phase, even not mentioned in the storyline, can be the organization name such as the hospital name ("Organization General Information" in Figure 4), the hospital divisions ("Processing Departments" in Figure 4), the employees such as the doctor that changes the medical record and the supervisor that validate those changes ("Employees, roles and responsibilities" in Figure 4). Therefore, in summary, within DAC and the ODC module, MM-Assess collects (through questionnaires) data for supporting the organization in carrying out a self-assessment and for collecting data needed for the purposes of the other DSM components and modules.

DSM Flow: Phase 2 (DAC: Organization Data Collection for Profiles Creation). This phase covers **AS2**, and its activities are represented in Figure 8. Here, the organization

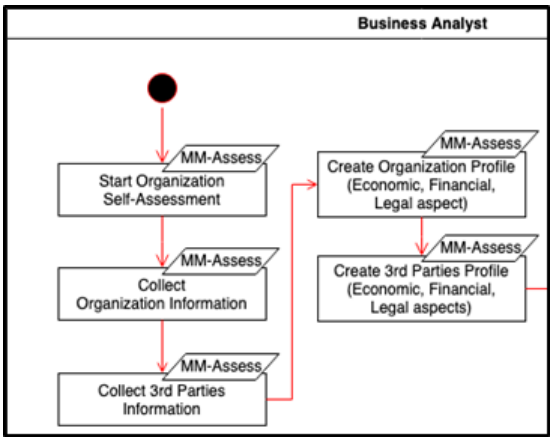

Fig. 8. DSM Flow Activities of Phase 2 [27,30] (and previous phase)

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 (Figure 3), using the MM-Assess (Figure 6) 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). Summarizing the DSM flow so far, in the first phase the self-assessment starts by collecting first organizational and 3^{rd} parties information for producing an overall situation of the organization. While in this phase, this process is continued by inserting more detailed aspects for creating a complete profile of the organization and 3^{rd} parties concerning economic, financial and legal aspects. In the next phase, this process is continued, and completed, with further details related to data categories managed by the organization and processing activities among the organization and 3^{rd} parties. DAC and ODC functionalities are fulfilled in these 2 phases by MM-Asses, while in the next phase it will be done by another tool, i.e. MM-REPA.

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 Figure 9. Partially **AS4** because of **AS4** covers only part of the data minimization analysis, as explained later, for collecting info on how so far the organization performed data minimization analysis The main objectives of this phase are to complete the self-assessment

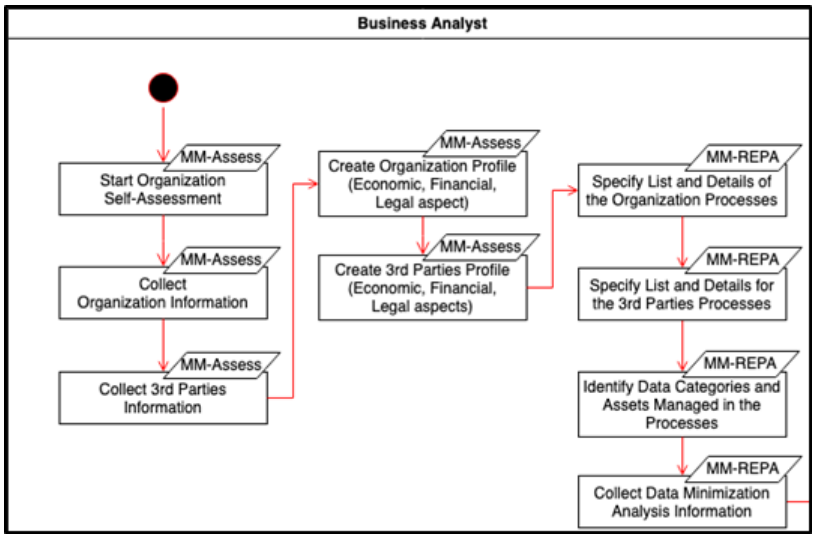

Fig. 9. DSM Flow Activities of Phase 3 [27, 30] (and previous phases)

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 (Figure 6) within the DAC component and the ODC module (Figure 3). 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''.

Other examples of data collected in DAM during this phase, also according to storyline, can be the data processing for patient record modification performed by the doctor, and the validation executed by her supervisor ("Processing List" and "Processing Description, Scope, Purpose and Legal Basis" in Figure 4), the assets that can be involved in these processes such as the computer, the medical record, medical exams ("Processing Assets" in Figure 4), and categories of data used within these processes such as "Confidential" and "Personal" ("Data Categories" in Figure 4).

Therefore, in summary, within DAC and the ODC module, MM-REPA completes the data collection, for the self-assessment of the organization, by supporting the business analyst in collecting (through questionnaires) information regarding data processing activities, of the organization and 3^{rd} parties, and typologies of data and assets involved. Moreover, some of this information is useful also for the purposes of the other DSM components and modules. For instance, later this is used by RAE (also in relation to 3^{rd} parties) for risk assessment purposes and high-level configuration generation for threats monitoring.

DSM Flow: Phase 4 (DAC: Assessment Translation and Data Synthesis). This phase covers mainly **AS4**. Its activities are represented in Figure 10. 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. In the next phase, for producing the DPIA and identifying crucial points where the organization could work on for improving GDPR compliance. This phase is associated to the MM-Assess tool (Figure 6) within the DAC component and the ATr module (Figure 3). No user is in charge of this phase, as instead happened for the previous ones, because this particular phase is completely automated by the DEFeND platform, by using MM-Assess. 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.

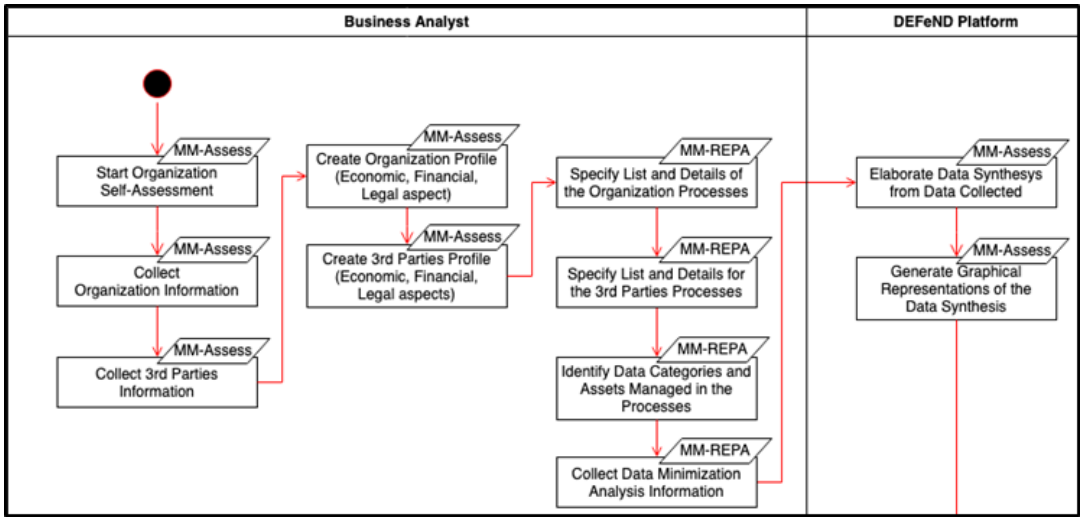

Fig. 10. DSM Flow Activities of Phase 4 [27, 30] (and previous phases)

Accordingly, activities executed during this phase are automatic and produce data synthesis and graphical representations, which respectively are stored in the DAM model and in the DEFeND Document Management System. 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").

Summarizing, within DAC and the ATr module, MM-Assess completes the GDPR self-assessment of the organization, by translating automatically the answers, collected until this phase, for generating data syntheses and graphical representations, save in the DAM Model and the Document Management System.

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 Figure 11. 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 (Figure 6) within the DPAC component and the DPIA Analysis, Threats Analysis, and Data Minimization Analysis modules (Figure 3). The user of the organization for this phase is typically a privacy/security analyst (Figure 6), which could collaborate with the business analysts that used the DEFeND platform in the previous steps. Most of the activities performed

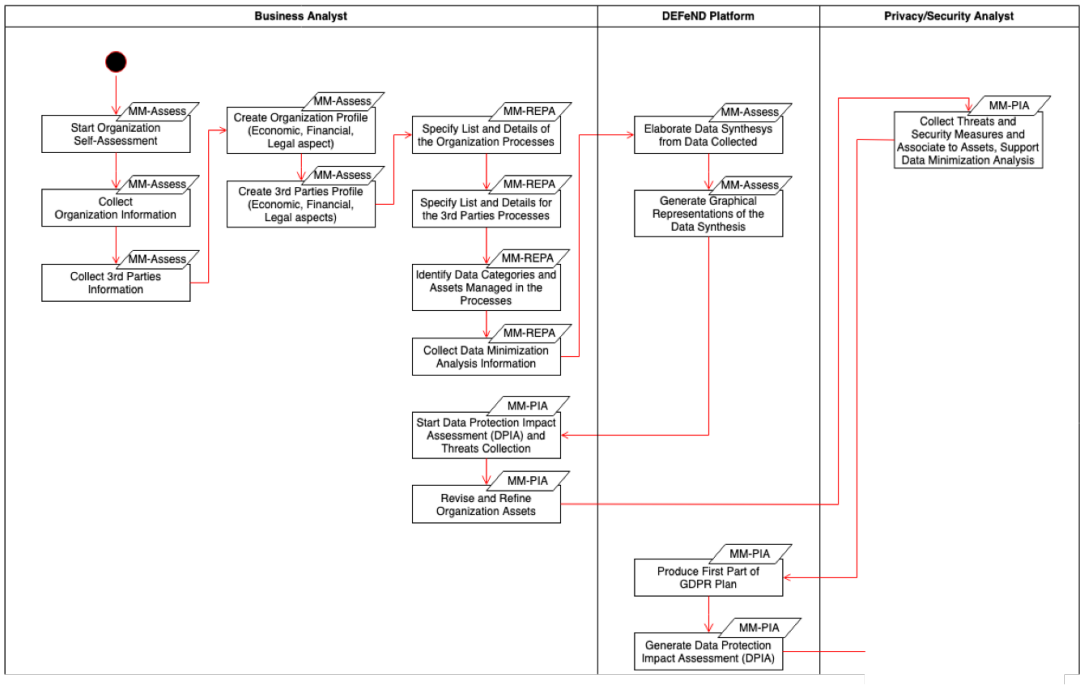

Fig. 11. DSM Flow Activities of Phase 5 [27, 30] (and previous phases)

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).

Furthermore, it is interesting to note that, in the DSM flow, MM-REPA and MM-PIA

are executed in this order, because MM-REPA collects assets and save them in DAM, then MM-PIA reads this information from DAM and, allows the analyst to revise and refine them, and to save them in other data categories of DPM, as outlined above.

Therefore, in summary, within DPAC and the DPIA Analysis, Threats Analysis, and Data Minimization modules, MM-PIA reads the DAM Model for supporting the organization in identifying, revising and refining organization assets and related GDPR risks, collecting related threats and impacts, for producing a DPIA and a GDPR compliance plan. Moreover, in the next phases, these elements collected, produced and saved in the DPM model, are relevant and important also for the next phases and are read and used by other DSM tools for other purposes. For instance, SecTro and RAE use them mainly for Threat Analysis. Furthermore, MM-PIA prepares the first elements needed for creating the GDPR plan; however, the full GDPR plan will be created by the Dashboard; in fact, it is the responsibility of the GDPR Planning Service to prepare the full GDPR plan, while DSM has the responsibility to collect information that could be needed also for the creation of the full GDPR Plan; MM-PIA read from DAM and write the information produced in the DPM Model.

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 Figure 12. High-level goals of this phase concern to

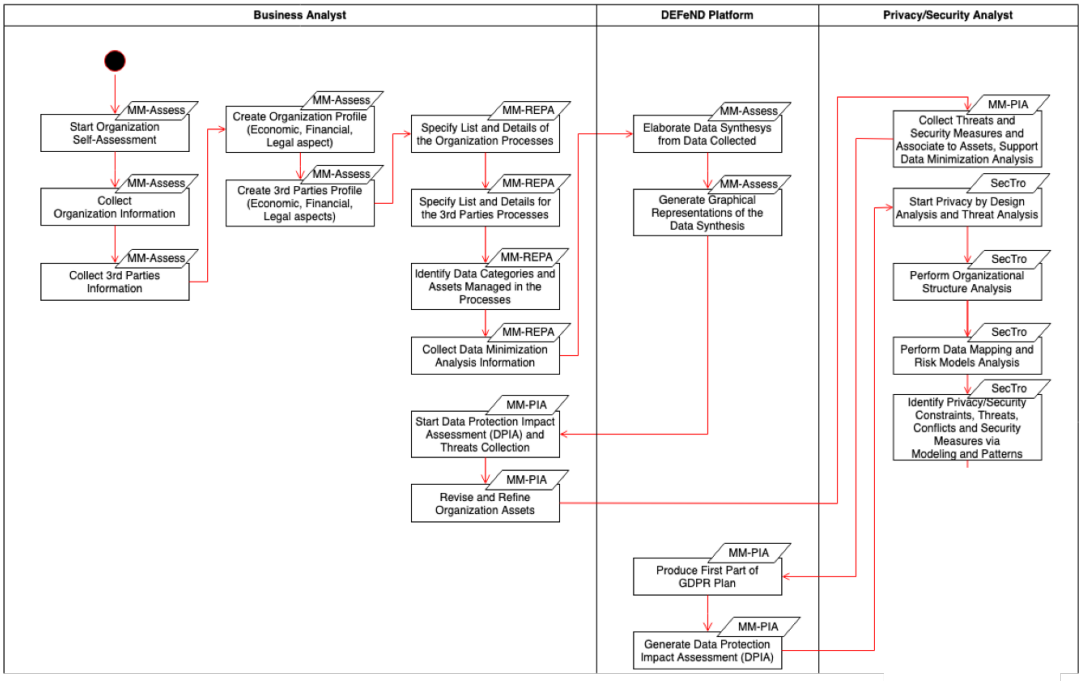

Fig. 12. DSM Flow Activities of Phase 6 [27, 30] (and previous phases)

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 (Figure 6), and its method, extended in DEFeND, within the DPAC component and the Privacy/Security and Threat Analysis modules (Figure 3).

Users of this phase are privacy/security analysts (Figure 6). 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. for the identification of: **(i)** Requirements conflicts and mechanisms for solving them [1]; **(ii)** Privacy/security measures for fulfilling privacy/security requirements and mitigating potential attacks to vulnerabilities. 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:

> "*The system helps also in modelling the data mapping with organizational assets, identifying the different data categories managed by the organization, and assigning data retention periods to them.*"

Therefore, in the Data Mapping View it is possible also to specify other important information

regarding data categories, where them are managed within data processing activities, and expressing relevant information for them (e.g., data retention periods). Also, the next extract of the storyline 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 [1], 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. Furthermore, SecTro supports the analyst in solving conflicts [1] via libraries of patterns including well-established solutions [1]. 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. Specifically, in the attack view the analyst can model in detail potential kinds of attacks that an attacker can execute, and individuate more in details new vulnerabilities to consider, more attacks and security mechanisms to employ.

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''.

Summarizing, within the Privacy/Security and Threat Analysis modules, SecTro reads the DAM and DPM models, including data produced by MM-PIA (mainly related to threats and vulnerabilities), and, on the basis of this input, shows, to the privacy/security analyst, semi-populated models in different views. For example, within the organizational view it will be shown the organizational structure model. Then, the analyst is supported by SecTro in carrying out different kinds of important analysis activities, through modelling, by enriching models, for instance to refine and discover new threats and vulnerabilities, she is supported by ready-to-use libraries of patterns concerning privacy/security requirements conflicts resolution [1] and individuation of security mechanisms mitigating threats. The output of those analyses is written in DPM and in the last phase by another DSM tool called RAE and described in the next subsection.

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 Figure 13. Goals of this phase are to support the organi-

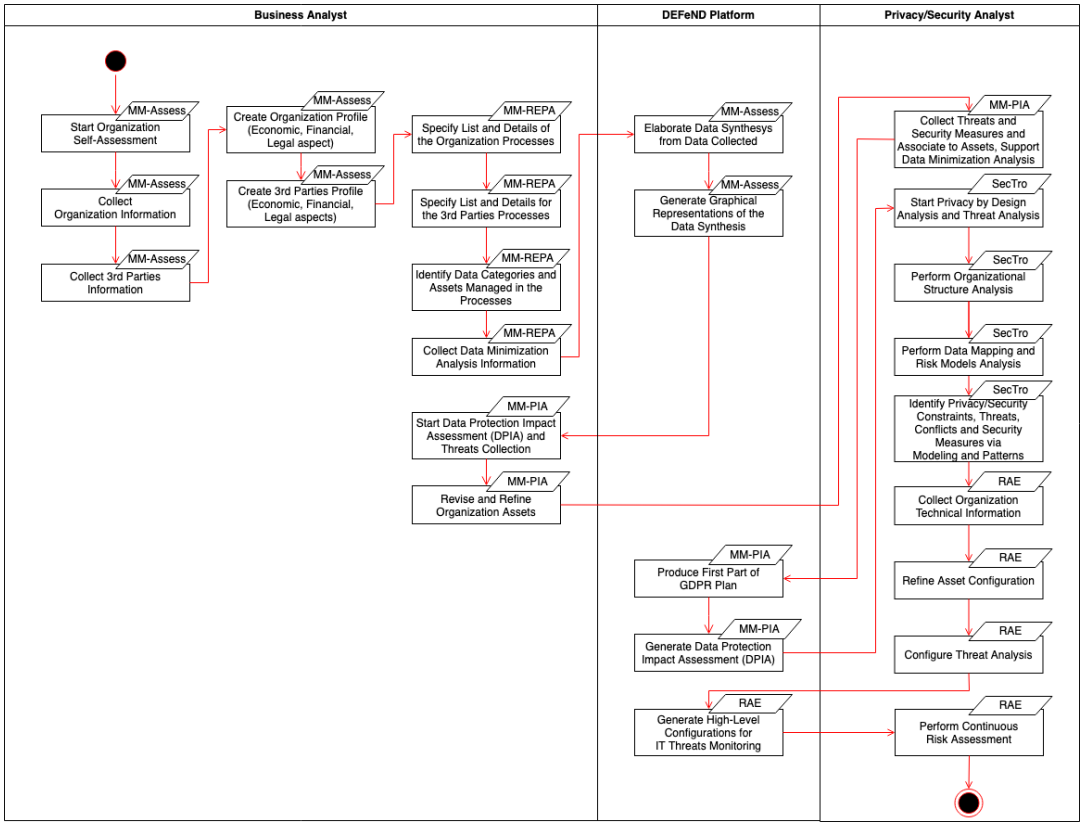

Fig. 13. DSM Flow Activities of Phase 7 [27, 30] (and previous phases)

zation 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 (Figure 6) within the DPAC component and the Threat Analysis module (Figure 3). Users of this phase are privacy/security analysts (Figure 6). 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, gen-*

> *erates 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, organization aspects such as organization debts of the hospital (data category ''Organization General Information'' of the DAM model in Figure 4), verification activities on data obtained by the lab (DAM data categories such as ''Processing List'' and ''Processing Description, Scope, Purpose and Legal Basis'', 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''). Assets that could be involved in threats such as the computer, the patient medical record and the exam results (DPM data category ''Privacy related Resources and Assets'', Figure 4). 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.

If some of this information has been collected in the previous phases, for instance through questionnaires by MM-PIA or modelling by using SecTro, related questions will be automatically filled in. 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.

Therefore, in summary, within DPAC and the Threat Analysis module, RAE reads from the DAM and DPM models information produced in DAC concerning the organization and list of data processing activities and details, and information written in DPM and produced by MM-PIA and using SecTro mainly in relation to risks, threats and vulnerabilities. Furthermore, in DPAC, if some of this information have not been collected in the previous phases, RAE collects this, via technical questionnaires, plus other technical information of the organisation including IT assets monitoring configuration (assets are known as targets in RAE's terminology). Then, RAE, on the basis of all this information collected, provides the analyst with a high-level configuration for making RAE doing IT threats monitoring and continuous risk assessment, within the scope of other services. The analyst can confirm, and optionally revise, such configuration. RAE reads information both from the DAM and DPM models, and writes information only in the DPM model.

3.5. *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 individuals. For this reason, we used individuals for our preliminary evaluation of our research claims, and in particular members of the pilot organisations that participate in the DEFeND project. Therefore, we included members from the pilots (i.e., users of the DSM) who 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 individuals, 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

In this Section we outline advances from the industry related to our work, and we also describe a number of relevant research challenges identified by the literature and how our work satisfies them.

4.1. *Industrial Advances*

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 [9]. 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 [31], 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'' [31]. The IAAP's report classifies the solutions into two key categories: Privacy Program Management and Enterprise Privacy Management. The first are grouped into six subcategories: assessment managers, consent managers, data mapping, incident response, privacy information managers and website scanning. The second are grouped in four 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 [8] released a report evaluating the twelve most significant providers in the market of EU GDPR compliance and privacy management. Platforms are evaluated with ten 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 for accountability. DSM supports this functionality via the Organization Data Collection module, where organizational data is collected and transformed automatically in a Data Assessment Model.

4.2. *Research Novelty*

This section briefly discusses relevant 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 (referred to as Data Protection by Design and by Default), but it is widely accepted that only few efforts exist to support practical implementation of PbD [6, 15, 17, 33]. The Data Scope Management service facilitates the structured implementation of PbD principles using methods and techniques from privacy requirements engineering, and privacy design. In addition, privacy requirements are frequently in conflict with security requirements, for example anonymity against accountability, which may induce system vulnerabilities and ultimately data breaches. Nevertheless, most of the approaches in the literature consider security and privacy as one, and they do not provide a way of eliciting and resolving conflicts. The Data Scope Management service facilitates the identification and resolution of conflicting privacy and security requirements by considering the context of the requirements.

Data Minimisation. There have been different approaches related to data minimisation at the data collection stage. In [3] the authors present a formal framework that enables designers to reason about data minimisation requirements. Data minimisation is presented as an optimisation problem in [2], and a set of algorithms is presented which solves the optimisation problem. The Data Scope Management Service facilitates the collection of personal data that is adequate in relation to the purpose that is pursued, relevant in light of the purpose, and limited to what is necessary, by continuously highlighting the correlation between personal data and data processing activities.

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 [32], while state of the art methodologies and tools to implement PIA are very few (for instance [14]). 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 for Privacy by Design (PbD), and a toolkit, DSM (the Data Scope Management service of the DEFeND EU project platform), supporting them, for carrying out major activities for Privacy by Design GDPR compliance. These activities include, for instance, identification of data processing activities and third parties, privacy/security, threats and risk assessment analysis, and all of them encapsulate a method for continuous model-based GDPR compliance.

In particular, we individuated, extended, and integrated candidate tools, the MM-Assess, MM-REPA, MM-PIA, SecTro, and RAE tools, for creating DSM, to support PbD activities. DSM tools are communicating through data models. In addition, we proposed a flow, supported by DSM, which organisations can follow in order to carry out the PbD activities in a systematic and structured way. The flow guides business analysts, and the privacy and security analysts, on how to conduct such activities. The proposed collection of methods, the toolkit, and the accompanied flow for the DSM service of the DEFeND platform, will be the result of the EU funded DEFeND project and will assist organisations to comply with GDPR using a PbD approach. To evaluate the proposed method, toolkit, and flow, we organised three 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 pilot organisations' responses were that the features of the method, toolkit, and flow satisfy their needs and have the potential to save them time and effort.

Acknowledgements

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

References

1. D. Alkubaisy, L. Piras, M. G. Al-Obeidallah, K. Cox, and H. Mouratidis. ConfIs: A Tool for Privacy and Security Analysis and Conflict Resolution for Supporting GDPR Compliance through Privacy-by-Design. In *16th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE)*, 2021.

2. N. Anciaux, B. Nguyen, and M. Vazirgiannis. Limiting data collection in application forms: A real-case application of a founding privacy principle. In *2012 Tenth Annual International Conference on Privacy, Security and Trust*, pages 59--66, 2012.
3. Thibaud Antignac and Daniel Le Métayer. Trust driven strategies for privacy by design. In Christian Damsgaard Jensen, Stephen Marsh, Theo Dimitrakos, and Yuko Murayama, editors, *Trust Management IX*, pages 60--75, Cham, 2015. Springer International Publishing.
4. Steven Blank. *The Four Steps to the Epiphany: Successful Strategies for Products that Win*. John Wiley & Sons, 2007.
5. Erik Paolo S. Capistrano and Jengchung Victor Chen. Information privacy policies: The effects of policy characteristics and online experience. *Computer Standards & Interfaces*, 2015.
6. Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfilment of Privacy Requirements. *Requirements Engineering Journal*, 2011.
7. Vasiliki Diamantopoulou, Konstantinos Angelopoulos, Julian Flake, et al. Privacy Data Management and Awareness for Public Administrations: A Case Study from the Healthcare Domain. In *Privacy Technologies and Policy*. Springer, 2017.
8. The Forrester New Wave^{TM}. `https://www.forrester.com/report/The%20Forrester%20New%20Wave%20GDPR%20And%20Privacy%20Management%20Software%20Q4%202018/-/E-RES142698`.
9. Gdpr temperature tool. `http://gdprtool.cyberwatching.eu/Pages/Home.aspx`.
10. M. Gharib, M. Salnitri, E. Paja, P. Giorgini, et al. Privacy Requirements: Findings and Lessons Learned in Developing a Privacy Platform. In *IEEE 24th International Requirements Engineering Conference (RE)*, pages 256--265, 2016.
11. Mohamad Gharib, John Mylopoulos, and Paolo Giorgini. COPri - a Core Ontology for Privacy Requirements Engineering. In *International Conference on Research Challenges in Information Science (RCIS)*, pages 472--489. Springer, 2020.
12. Seda Gürses, Carmela Troncoso, and Claudia Diaz. Engineering privacy by design. *Computers, Privacy & Data Protection*, 14(3):25, 2011.
13. Jaap-Henk Hoepman. Privacy design strategies. In *IFIP International Information Security Conference*, pages 446--459. Springer, 2014.
14. Martin Horák, Václav Stupka, and Martin Husák. GDPR Compliance in Cybersecurity Software: A Case Study of DPIA in Information Sharing Platform. In *14th International Conference on Availability, Reliability and Security*, 2019.
15. Christos Kalloniatis, Petros Belsis, and Stefanos Gritzalis. A Soft Computing Approach for Privacy Requirements Engineering: The PriS Framework. *Applied Soft Computing*, 2011.
16. Blagovesta Kostova, Seda Gürses, and Carmela Troncoso. Privacy engineering meets software engineering. on the challenges of engineering privacy by design. *arXiv preprint arXiv:2007.08613*, 2020.
17. Christian Kurtz, Martin Semmann, et al. Privacy by Design to Comply with GDPR: a Review on Third-Party Data Processors. In *Americas Conf. on Inf. Systems*, 2018.
18. Martin Maguire. Methods to Support Human-Centred Design. *Int. Journal of Human-Computer Studies*, 2001.
19. Aleecia M McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *ISJLP*, 2008.
20. Haralambos Mouratidis. Secure Software Systems Engineering: the Secure Tropos Approach. *JSW*, 2011.
21. Haralambos Mouratidis, Nikolaos Argyropoulos, and Shaun Shei. *Security Requirements Engineering for Cloud Computing: the Secure Tropos Approach*. Springer, 2016.
22. Haralambos Mouratidis and Paolo Giorgini. Secure Tropos: a Security-Oriented Extension of the Tropos Methodology. *International Journal of Software Engineering and Knowledge Engineering*, 2007.
23. Michalis Pavlidis and Shareeful Islam. SecTro: A CASE tool for Modelling Security in Requirements Engineering Ssing Secure Tropos. In *Ceur Workshop Proceedings*, volume 734, pages

89--96, 2011.
24. Michalis Pavlidis, Haralambos Mouratidis, Cesar Gonzalez-Perez, and Christos Kalloniatis. Addressing Privacy and Trust Issues in Cultural Heritage Modelling. In *CRiSIS 2015*. Springer, 2015.
25. Michalis Pavlidis, Haralambos Mouratidis, and Shareeful Islam. Modelling Security Using Trust Based Concepts. *International Journal of Secure Software Engineering (IJSSE)*, 2012.
26. Michalis Pavlidis, Haralambos Mouratidis, Emmanouil Panaousis, and Nikolaos Argyropoulos. Selecting Security Mechanisms in Secure Tropos. In *International Conference on Trust and Privacy in Digital Business (TrustBus)*. Springer, 2017.
27. L. Piras, M.G. Al-Obeidallah, M. Pavlidis, H. Mouratidis, A. Tsohou, E. Magkos, A. Praitano, A. Iodice, and B.G.N. Crespo. DEFeND DSM: A Data Scope Management Service for Model-Based Privacy by Design GDPR Compliance. In *International Conference on Trust, Privacy and Security in Digital Business (TrustBus)*. Springer, 2020.
28. L. Piras, F. Calabrese, and P. Giorgini. Applying Acceptance Requirements to Requirements Modeling Tools via Gamification: A Case Study on Privacy and Security. In *13th International Conference on the Practice of Enterprise Modeling (PoEM)*. Springer, 2020.
29. L. Piras, D. Dellagiacoma, A. Perini, A. Susi, P. Giorgini, and J. Mylopoulos. Design Thinking and Acceptance Requirements for Designing Gamified Software. In *13th Intern. Confer. on Research Challenges in Information Science (RCIS)*. IEEE, 2019.
30. Luca Piras, Mohammed Ghazi Al-Obeidallah, Andrea Praitano, Aggeliki Tsohou, Haralambos Mouratidis, Beatriz Gallego-Nicasio Crespo, Jean Baptiste Bernard, Marco Fiorani, Emmanouil Magkos, Andres Castillo Sanz, 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.
31. Privacy Tech Vendor Report.
32. K. Rantos, G. Drosatos, K. Demertzis, C. Ilioudis, A. Papanikolaou, and A. Kritsas. ADvoCATE: a Consent Management Platform for Personal Data Processing in the IoT Using Blockchain Technology. In *Int. Conf. on Security for Information Technology and Communications*. Springer, 2018.
33. Anna Romanou. The Necessity of the Implementation of Privacy by Design in Sectors where Data Protection Concerns Arise. *Computer law & security review*, 34(1):99--110, 2018.
34. Aggeliki Tsohou and Eleni Kosta. Enabling valid informed consent for location tracking through privacy awareness of users: A process theory. *Computer Law & Sec. Review*, 2017.
35. Aggeliki Tsohou, Emmanouil Magkos, Haralambos Mouratidis, George Chrysoloras, Luca Piras, Michalis Pavlidis, Julien Debussche, Marco Rotoloni, and Beatriz Gallego-Nicasio Crespo. Privacy, Security, Legal and Technology Acceptance Elicited and Consolidated Requirements for a GDPR Compliance Platform. *Information and Computer Security Journal*, 2020.
36. Aggeliki Tsohou, Emmanouil Magkos, Haris Mouratidis, George Chrysoloras, Luca Piras, Michalis Pavlidis, Julien Debussche, Marco Rotoloni, and Beatriz Gallego-Nicasio Crespo. Privacy, Security, Legal and Technology Acceptance Requirements for a GDPR Compliance Platform. In *Int. Workshop on Security and Privacy Requirements Eng. (SECPRE)*. Springer, 2019.
37. Wiser real-time assessment infrastructure (d5.2). `https://www.cyberwiser.eu/system/files/20160727_WISER_D5_2_v10%281%29_0.pdf`.