

# A Model-based Approach to Support Privacy Compliance

Majed Alshammari and Andrew Simpson

Department of Computer Science, University of Oxford  
Wolfson Building, Parks Road, Oxford OX1 3QD, UK  
`firstname.secondname@cs.ox.ac.uk`

## Abstract

**Purpose** - Concerns over data-processing activities that may lead to privacy violations or harms have motivated the development of legal frameworks and standards. Further, software engineers are increasingly expected to develop and maintain privacy-aware systems that both comply with such frameworks and standards and meet reasonable expectations of privacy. This paper aims to facilitate reasoning about privacy compliance, from legal frameworks and standards, with a view to providing necessary technical assurances.

**Design/methodology/approach** - We show how the standard extension mechanisms of the UML meta-model might be used to specify and represent data-processing activities in a way that is amenable to privacy compliance checking and assurance.

**Findings** - We demonstrate the usefulness and applicability of the extension mechanisms in specifying key aspects of privacy principles as assumptions and requirements, as well as in providing criteria for the evaluation of these aspects to assess whether the model meets these requirements.

**Originality/value** - First, we show how key aspects of abstract privacy principles can be modelled using stereotypes and tagged values as privacy assumptions and requirements. Second, we show how compliance with these principles can be assured via constraints that establish rules for the evaluation of these requirements

## 1 Introduction

Privacy is a multifaceted concept that has legal, social and political aspects. It is subjective in nature and culturally variable, not least because it is derived from societal demands, expectations and culture, which are all in turn influenced by political, social and economic factors, as well as information technology advancements (Onn *et al.*, 2005). The complexity challenges software engineers to understand and translate abstract privacy principles and various perceptions into operational requirements (Gürses *et al.*, 2011; Spiekermann, 2012). The variability challenges software engineers to understand and consider multiple stakeholders' expectations and concerns,

which, in turn, requires a specific expertise, contextual analysis and resolution of stakeholders' security and privacy interests (Gürses *et al.*, 2011).

Typically, privacy is articulated at a high level of abstraction. Accordingly, its concrete manifestations are ambiguous to those concerned with data protection and those responsible for developing and maintaining systems (Kost *et al.*, 2011; Shapiro, 2010). Further, incorporating privacy requirements into the early stages of the development process requires an appropriate interpretation of legal, social and political concerns (Gürses *et al.*, 2011). These challenges lead to a disconnect between policy-makers and software engineers with regards to the actual meaning of privacy, its related concepts, and the ways in which systems can be developed to comply with legal frameworks and standards, as well as data subjects' expectations (Spiekermann, 2012). Accordingly, there is a need for modelling techniques that support the effective translation of abstract privacy principles, privacy risk models and privacy mechanisms into implementable requirements (Shapiro, 2010; Spiekermann, 2012).

Privacy by Design (PbD) (Cavoukian, 2011) has been advocated as a solution. It aims to achieve privacy assurance by meeting regulatory compliance requirements and mitigating potential privacy risks. However, the principles of PbD are given at a high level of abstraction. As a consequence, it is not clear how the foundational principles of PbD can be appropriately translated into engineering activities (Gürses *et al.*, 2011).

The privacy-aware conceptual model for handling personal data (Antignac *et al.*, 2016) was proposed to bridge the semantic gap between legal and technical concepts. It aims to support reasoning about the compliance of a technical design with respect to legal frameworks and/or standards. It adopts Data Flow Diagrams (DFDs) as a convenient representation to validate software design. However, it does not capture all aspects of abstract privacy principles that are necessary for compliance checking (Antignac *et al.*, 2016). To overcome such limitation, the Abstract Personal Data Lifecycle (APDL) model (Alshammari and Simpson, 2017a) was developed to serve as a stepping stone for modelling privacy-related concepts along with associated properties and relationships, and for representing data-processing activities in a way that is amenable to risk analysis and compliance checking. The Unified Modeling Language (UML) (Object Management Group, 2015) has been adopted to help support the APDL model's main concepts and integration into software engineering processes. In particular, a UML profile for the APDL model (Alshammari and Simpson, 2017b) was proposed to describe data-processing activities at a fined-grained level, with the possibility of expressing how these activities are performed, their effects in terms of changes of states, when they take place in terms of lifecycle stages, and where they take place in terms of lifecycle roles (Alshammari and Simpson, 2017b). Based on the APDL model and its UML profile, we present a model-based approach to facilitate reasoning about privacy compliance with abstract privacy principles.

This paper is organised as follows. Section 1 frames the overall problem and gives an overview of the UML profile of the APDL model; it also motivates and justifies the contribution. Section 2 illustrates how the extension mechanisms can be used to facilitate reasoning about privacy compliance. Section 3 introduces the ePetition system, the aim of which is to implement the European Citizens' Initiative (ECI), which

we shall use as an illustrative case study. In addition, it illustrates the applicability of the UML profile through the case study. Finally, Section 5 summarises the contribution, and outlines our plans for future work.

## 2 Background and Motivation

### 2.1 The Problem

The challenge of ensuring that the processing of personal data is done fairly and lawfully has motivated the development of legal frameworks and standards (such as the EU General Data Protection Regulation (GDPR) (The European Parliament, 2016)). Typically, legal frameworks and standards in general, and those related to privacy and data protection in particular, do not rely on rigorous models that specify key concepts along with their properties and relationships (Antignac *et al.*, 2016); rather, the principles of such frameworks and standards are often given at a high level of abstraction. This leads to practical challenges with respect to translating legal, social and political concerns into systems requirements (Gürses *et al.*, 2011) and reasoning compliance with these principles. Further, it is difficult to use such principles as a means of describing design options or justifying architectural choices (Oetzel and Spiekermann, 2014). Thus, there is a semantic gap (Antignac *et al.*, 2016). As such, privacy principles need to be translated into concrete, auditable and functionally enforceable goals to aid engineers in specifying design decisions to meet privacy compliance requirements (Oetzel and Spiekermann, 2014).

Importantly, privacy goals can be used by multiple stakeholders to express their privacy concerns and expectations, as they describe how personal data is collected and processed. Based on these goals, engineers can derive privacy requirements, specify design options that fulfil these requirements and choose appropriate technologies that implement these options (Hansen *et al.*, 2015). This, in turn, necessitates a classification of universal and comprehensive privacy goals (Beckers, 2012) that can be applied in a variety of contexts in various jurisdictions. A number of privacy requirements engineering approaches have been proposed to help support the elicitation of privacy requirements, such as the PriS method (Kalloniatis *et al.*, 2008) and LINDDUN (Mina *et al.*, 2011). However, these approaches differ in their privacy goals, methods, notions and terminology (Beckers, 2012).

In recent years, Privacy by Design (PbD) has been proposed as a universal guideline to address these challenges (Gürses *et al.*, 2011). PbD, however, does not address the establishment of engineering methodologies. The principle of data minimisation has been proposed as a necessary and foundational first step for engineering systems according to the principles of PbD (Gürses *et al.*, 2011). However, ensuring data minimisation is in itself a challenge (Antignac *et al.*, 2016): software engineers need support in determining the appropriate type of data minimisation for a given system. Several factors need to be taken into account, including the nature and semantics of the data (Antignac *et al.*, 2016), as well as stakeholders' expectations, applicable laws and regulations, and appropriate threat models (Spiekermann and Cranor, 2009).

There is more to PbD than data minimisation: it aims to mitigate potential privacy risks, achieve accountability and enhance user trust (Cavoukian *et al.*, 2014). One attempt to address these challenges in a broader context is the *privacy design strategies* of (Hoepman, 2014), which serve as a fundamental approach to realise specific architectural goals - derived from privacy principles and data protection regulations - to achieve a particular level of privacy protection. Another example is the *privacy enhancing architecture methodology* of (Kung, 2014), which illustrates how to design, analyse and evaluate software architectures through quality attributes, architectural tactics and design patterns. However, both design strategies and architectural tactics jump directly from abstract privacy principles into software architecture (Martín *et al.*, 2014) without providing criteria for the evaluation of architectural choices and the selection of appropriate design patterns and the corresponding technologies.

The first step towards bridging the semantic gap between technical and normative concepts involves providing an appropriate conceptual model for privacy engineering. The APDL model (Alshammari and Simpson, 2017a) expresses the processing of personal data in terms of *states* (data items), *operations* (activities), and *roles* (actors). It facilitates the protection, traceability and management of personal data during its lifetime in an abstract manner. However, the APDL was initially presented informally without formal semantics, which would facilitate analysing potential privacy risks and reasoning about privacy compliance with legal frameworks and/or standards. In order for the APDL model to be integrated into an appropriate software engineering process, a widely-used modelling language, needs to be adopted to help support its main concepts. UML is ideal for this purpose.

## 2.2 UML and its Extensions

UML is a well-defined language that can be used for specifying, representing and documenting the artefacts of software-based systems at different levels of abstraction and from different viewpoints. Importantly, it can be extended to communicate new intentions in particular domains. *Stereotypes* can be used to extend the vocabulary of UML by creating new model elements derived from existing ones but that have specific properties suitable for the domain of interest. *Tagged values* can be used to extend the properties of a UML model element to add new information in the specification of that element. As its value applies to the element itself and not its instances, tagged values serve as metadata.

To specify unambiguous constraints, Object Constraint Language (OCL) (Object Management Group, 2012) can be used to describe expressions on UML models that are typically not sufficiently refined to provide all aspects of specification. These expressions specify unambiguous platform-, method- or domain-specific constraints in terms of invariants, pre- and post-conditions.

An example of an extension is UMLsec (Jürjens, 2002), which facilitates the consideration of security requirements from the early stages of the development process and the evaluation of security vulnerabilities at the design level. The extension is presented in the form of a UML profile using the standard UML extension mecha-

nisms. Stereotypes, along with tag definitions, are used to specify key aspects of security as assumptions and requirements, and constraints are used to specify criteria to evaluate whether the requirements are satisfied by the system design (Jürjens, 2002). The UML profile for privacy-aware data lifecycle models (Alshammari and Simpson, 2017b) was initially proposed to adapt UML to the domain of privacy-aware systems. It allows the aforementioned APDL model to be represented in UML as a meta-model that represents the personal data lifecycle in terms of stages that involve data-processing activities that consist of concrete events and corresponding actions, and roles that define a set of responsibilities performed by different actors according to their capabilities. Its stereotypes are defined to extend existing metaclasses with the aim of using privacy-related terminology whether in place of, or in addition to, the terminology used for the extended metaclasses (Alshammari and Simpson, 2017b). The constraints needed to express privacy-related concepts of the APDL model are limited only to association multiplicities, pre- and post-conditions of stage activities and actions. The APDL profile was defined to represent data-processing activities in a contextual and fined-grained manner to support risk analysis and compliance checking. Here, we focus on the role of the APDL profile in facilitating compliance checking.

### 2.3 The Contribution

In order to collect and process personal data fairly and lawfully, regulatory compliance requirements need to be concretely modelled to ensure that designs fulfil them. The UML profile for the APDL model represents privacy-related concepts using the standard extension mechanisms of the UML meta-model. Stereotypes and tagged values are used to specify data-processing activities along with the key aspects of privacy principles as a requirements model, and constraints provide criteria to determine whether the model fulfils these requirements (Alshammari and Simpson, 2017b). However, it was not explicitly illustrated in (Alshammari and Simpson, 2017b) how the standard extension mechanisms of UML can be used to facilitate reasoning about privacy compliance with abstract privacy principles. This leaves a number of open questions:

1. How might one model the purposes for which personal data is collected and processed in a way that is amenable for compliance checking?
2. How might one model the key aspects of abstract privacy principles as assumptions, requirements and constraints against which privacy compliance checking can be performed?
3. How can the standard extension mechanisms be used to facilitate reasoning about privacy compliance with legal frameworks and standards?

Thus we aim to enrich the APDL profile with useful extension mechanisms that aid software engineers in: modelling the purposes in a way that is amenable for compliance checking; modelling the key aspects of privacy principles as assumptions and requirements; and specifying suitable criteria as constraints for the evaluation of these aspects to ensure that the model fulfils these requirements.

### 3 An Approach to Support Privacy Compliance

To facilitate reasoning about privacy compliance, the abstract purposes for which personal data is collected and processed need to be operationalised. The abstract purpose is a nonoperational objective to be directly achieved by the collection and processing of personal data. By operationalising it, we wish to move towards a position in which the objective is expressed in terms of personal data items, data-processing activities that consist of concrete actions and events that cause the execution of these actions, and roles that define a set of responsibilities performed by different actors according to their capabilities. Further, abstract purposes are made operational through constraints that specify conditions to be satisfied before, or to be guaranteed after, the execution of corresponding operations. Accordingly, the main aspects of abstract privacy principles are specified as constraints on the abstract purpose. In particular, the purpose operationalisation consists of four separate, but related activities: refinement, conceptualisation, representation and evaluation.

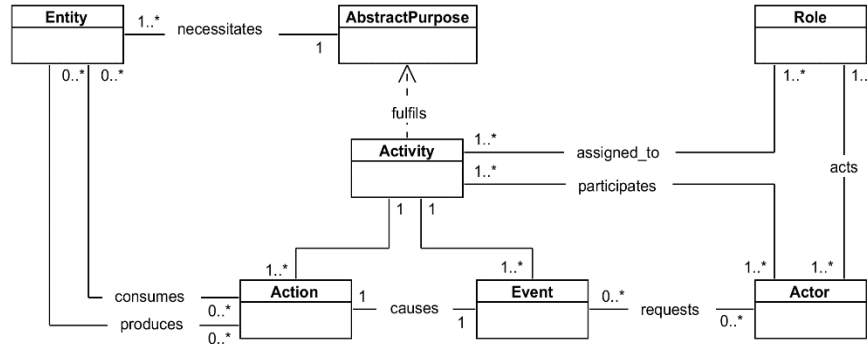
#### 3.1 Refinement

This activity aims to refine the abstract purpose into a set of concrete purposes that can be assigned to actors as responsibilities. This aim can be achieved by specifying the abstract purpose at a certain level of detail as concrete purposes - i.e. data-processing activities - to capture how each activity participates in the fulfilment of the abstract purpose.

Instead of specifying abstract purposes in terms of either entities or activities, they are better specified as objectives from which entities and activities can be derived. They can be represented in a hierarchical structure of which the lowest level represents concrete purposes as data-processing activities, which can be assigned as the responsibility of actors, as per **Error! Reference source not found..** The activities are operationalised by actions that are triggered by events, which are requested by the actors responsible for each activity according to the role to which they are assigned in such a way that activities fulfil the abstract purpose. Actors may also process certain objects of personal data that are derived when refining abstract purposes. These objects can be consumed or produced by certain actions as inputs or outputs. At a high level of abstraction, the abstract purpose necessitates objects of personal data; at a low level, objects are necessary for an activity if these objects are processed by an action that operationalises the activity.

In summary, the main steps of the refinement activity are as follows.

1. The abstract purpose needs to be refined into concrete purposes.
2. Each concrete purpose needs to be expressed in terms of actions and events that trigger the execution of these actions.
3. The minimum amount necessary of personal data needs to be derived from the actions of the concrete purposes.
4. The concrete purposes need to be assigned to the capable actors according to their roles and associated responsibilities.



**Figure 1** The refinement of purposes.

### 3.2 Conceptualisation

This activity aims to derive and model the key aspects of abstract privacy principles. These include, for example, the EU General Data Protection Regulation (GDPR) (The European Parliament, 2016), the Global Privacy Standard (GPS) (Cavoukian, 2006) and the Generally Accepted Privacy Principles (GAPP) (American Institute of Certified Public Accountants and Canadian Institute of Chartered Accountants, 2009). This aim can be achieved by classifying the primary terms into *concepts*, along with associated properties, meanings, possible values, and useful and potentially usable *actions*, together with associated constraints. The concepts are to represent the primary terms, whereas the actions are to ensure that the main operations make sense in the context of privacy and data protection. The constraints specify conditions to be satisfied before, or to be guaranteed after, the execution of corresponding actions. Each aspect of these principles can be modelled by one or more concepts, which can be characterised by a set of properties, or one or more actions, which can be restricted by a set of constraints.

Such a conceptual model can be used as a common language for privacy engineering to express stakeholders' expectations and concerns. It can be used by multiple stakeholders - both those concerned with privacy and data protection, and those responsible for developing and maintaining privacy-aware systems.

In summary, the main steps of the conceptualisation activity are as follows.

1. The sources from which knowledge can be acquired need to be identified.
2. The most appropriate technique for deriving useful and potentially usable concepts and actions needs to be used.
3. A list of concepts, their meanings and properties needs to be identified.
4. A list of actions and their conditions needs to be identified.
5. A conceptual model that describes the problem and its solution in terms of the domain vocabulary needs to be developed.

### 3.3 Representation

This activity aims to model the abstract and concrete purposes together with the key aspects of abstract privacy principles as a requirements model. This aim can be achieved by adopting the UML profile for the APDL model as a means for representation.

The UML profile provides all necessary concepts for operationalising the abstract purpose. It represents the concrete purposes in a hierarchical structure as data-processing activities according to the stages of the personal data lifecycle. Further, it represents involved actors, along with their assigned roles and responsibilities, and captures how those actors participate in data-processing activities according to their roles.

The abstract purpose can be represented using the «Purpose» stereotype. The concrete purposes need to be classified according to the typical stages of the personal data lifecycle as data-processing activities. The lifecycle stages can be represented using the «LifecycleStage» stereotype; data-processing activities can be represented using the «StageActivity» stereotype. For each stage of the lifecycle, a data-processing activity can be assigned to a role, represented using the «LifecycleRole», as a responsibility through which an actor, represented using the «LifecycleActor» stereotype, can participate in the performance of the activity. Furthermore, each data-processing activity can be represented in terms of actions and events via «StageAction» and «StageEvent» respectively. Each event is requested by an actor; each action consumes and/or produces specific personal data items. These items can be represented using the «PersonalData» stereotype.

The UML profile is based upon a conceptual model that captures the key aspects of abstract privacy principles. It adopts the principles of the aforementioned GPS and GAPP. The concepts identified in the conceptualisation activity can be modelled as stereotypes or tag definitions for specific stereotypes to add additional information in the specification of these stereotypes. In doing so, the profile can be used as an acquisition language understood by multiple stakeholders.

In summary, the main steps of the representation activity are as follows.

1. The main steps of the refinement activity need to be conducted.
2. The main steps of the conceptualisation activity need to be conducted.
3. The abstract and concrete purposes, along with the concepts and actions derived from the abstract privacy principles, can be modelled using the stereotypes and tag definitions of the UML profile for the APDL model.

### 3.4 Evaluation

This activity aims to specify the constraints through which the abstract purpose can be operationalised. These constraints can be used to specify conditions on both *concepts* and *actions* identified in the conceptualisation activity. As such, the constraints are used to establish a set of suitable rules against which privacy compliance checking can be performed.

The UML profile for the APDL model supports the specification of these constraints in two different ways. First, OCL expressions can be modelled as invariants on the stereotypes and their associated tag definitions when necessary. Second, OCL expressions can be modelled as preCondition or postCondition and localPreCondition or localPostCondition of the «StageActivity» and «StageAction» stereotypes respectively. In this case, the OCL expressions are modelled as tagged values for tag definitions of the stereotypes.

In summary, the main steps of the evaluation activity are as follows.

1. For each concept identified in the conceptualisation activity, all possible invariant conditions need to be established.
2. For each action identified in the conceptualisation activity, all possible pre- and post-conditions that must be satisfied need to be established.
3. The established rules need to be specified using the OCL as constraints on the stereotypes.

## 4 An Example

### 4.1 Overview

This section introduces the ePetition system, the aim of which is to implement the European Citizens' Initiative (ECI) (European Commission, 2012) (which was previously used as an illustrative case study in (Alshammari and Simpson, 2017a)). The ECI is an online collection system used to support a formal request provided by organisers to a particular authority for submitting a proposal for a legal act. It enables one million EU citizens from at least seven EU Member States to invite the European Commission to propose a legal act on issues where it has competence to legislate. The abstract purpose of collecting and processing personal data is to verify and certify the number of valid signatures that support an initiative. In this paper, we focus on a petition that can be submitted in electronic form.

A citizens' committee of at least seven EU citizens acts in its capacity as the official organiser of the initiative and is responsible for preparing and launching the initiative. The organisers also need to find a hosting service provider when signatures are to be collected electronically by an online collection system. A Signatory, who acts as a data subject, is able to support a specific initiative by submitting a statement of support that requires providing 'identifying' personal data, such as full names, permanent resident, data of birth and nationality. However, in some Member States, such as France and Spain, personal identification numbers are required. Once the required number of signatures is reached, organisers should send statements of support to relevant competent national authorities for verification and certification. Having received all certificates from competent national authorities, organisers should submit the initiative by sending these certificates to the European Commission.

Organisers and competent national authorities act as data controllers, who must ensure that collected personal data is not used for purposes other than those specified for supporting the initiative and verifying the statements of support respectively. Further, the data controllers must destroy all statements of support and any copies one month

after submitting the initiative to the Commission or issuing the certificate respectively.

## 4.2 An Illustration

**Refinement.** The *abstract purpose* is to verify and certify the valid number of statements of support for a proposed citizens' initiative.

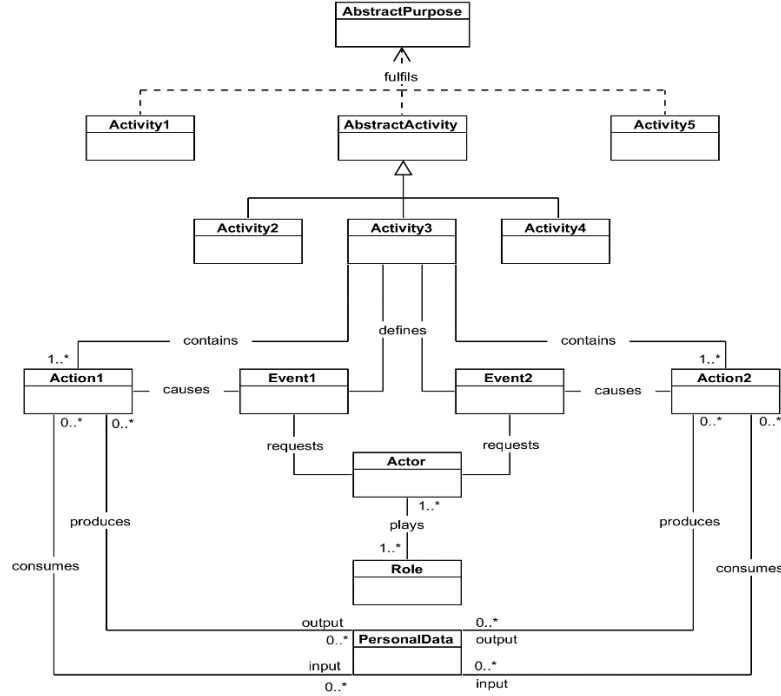
The first step is to refine the abstract purpose into concrete purposes. Due to space limitations, we list the most important ones and express only one activity in terms of actions, events, roles and actors (**Figure 2**).

The abstract purpose is partially refined into five concrete purposes. CollectingStatementsOfSupport is represented as *Activity1*, which collects the specified personal data from at least one million EU citizens who act as signatories for signing-up to a citizens' initiative. ExportingAllStatementsOfSupport is represented as *Activity2*, which exports all statements of support and displays the total distribution of signatures. ExportingSpecificStatementsOfSupport is represented as *Activity3*, which exports the selected statements of support according to the Member State and/or the date of submission for reporting purposes. DeletingSpecificStatementOfSupport is represented as *Activity4*, which deletes the selected statement of support according to the signature identifier and optionally the date of submission. *Activity2-Activity4* are generalised by an abstract activity - MonitorAndExportStatementsOfSupport - which is represented as *AbstractActivity*. ConfirmingValidStatementsOfSupport is represented as *Activity5*, which verifies and certifies the number of valid statements of support.

The second step is to express each activity in terms of actions and events that trigger the execution of these actions. We illustrate how to express the *Activity3*. The activity coordinates its execution via two actions and two corresponding events. ExportByMemberState is represented as *Action1*. ExportByCountry is represented as *Event1*. It specifies the occurrence of retrieving specific statements of support according to the Member State to export their corresponding data. ExportBySubmissionDate is represented as *Action2*. ExportByDate is represented as *Event2*. It specifies the occurrence of retrieving specific statements of support according to the submission date to export their corresponding data.

The third step is to derive the minimum amount of required personal data items. For *Action1* and *Action2* to accomplish their execution, they consume and/or produce personal data items, including name, address and date of birth. These data elements, the mandatory fields of the statement of support form, are represented as *PersonalData* in **Figure 2**.

The fourth step is to assign this activity to capable actors. ECIUser, represented as *Actor*, is an actor who is capable of, and responsible for, performing the activities of the ECI organiser. ECIOrganiser, represented as *Role*, is a data controller role that involves logically related activities for monitoring and exporting the collected statements of support.



**Figure 2** The refinement of the abstract purpose.

**Conceptualisation.** Instead of ‘starting from scratch’, we adopt the conceptual model developed in (Alshammari and Simpson, 2017b), which was developed based upon the principles of the aforementioned GPS and GAPP. We show only how its derived concepts and actions are related to the principles of the GPS.

**Purposes.** The purposes must be clear, limited and relevant to the circumstances. The principle is related to another principle in these sources: *Notice*. This aims to provide notice to data subjects regarding the purposes for, and the manner in which, personal data is collected and processed. Two key concepts are derived from this principle: the first concept is ‘purpose’ to specify the actual purpose, along with its fairness, lawfulness and proportionality. The second concept is ‘notice’ to specify the manner in which personal data is collected, used and disclosed.

**Consent.** The type of consent depends on the nature of personal data and the manner in which is to be processed, unless a law or regulation specifically requires otherwise. The principle is related to another principle in these sources: *Choice*. This aims to specify the choices available to data subjects with respect to the collection, use and disclosure of personal data. Two key concepts are derived from this principle: the first concept is ‘choice’ to specify a set of available choices. The second concept is ‘consent’ to specify the type of consent to be obtained. Two actions are derived from this principle: the notice - including available choices and the means by which data sub-

ject can exercise these choices - must be communicated to data subjects at or before the collection time; and the consent must be obtained at or before the collection time in a freely and knowledgeable basis.

**Collection Limitation.** This principle aims to ensure that the collection of personal data is lawful, fair and limited to that which is necessary for the specified purposes. It is closely related to the principle of ‘data minimization’, which aims to strictly minimise the amount of data to be collected. Collection actions need to be restricted by conditions: the notice - including data elements, data sources, available choices and collection methods - must be communicated to data subjects at or before the collection time; the consent must be obtained at or before the collection time; and the collection methods must be fair and lawful.

**Use, Retention, and Disclosure Limitation.** This principle aims to ensure that the use of personal data is limited to the purposes for which data subjects have provided implicit or explicit consent; personal data is retained for no longer than necessary to fulfil the specified purposes, unless a law or regulation specifically requires otherwise, and thereafter securely anonymised, disposed of, or destroyed; and the disclosure of personal data is limited to third parties only for the specified purposes for which data subjects have provided implicit or explicit consent. It is closely related to the principle of ‘data minimization’, which aims to strictly minimise the amount of data to be processed and the number of actors who have access to, use, or to whom personal data is disclosed according to their roles and assigned responsibilities. These actions need to be restricted by a set of conditions: the notice must be communicated to data subjects at or before the collection time; the consent must be obtained at or before the collection time; personal data must be used only by authorized actors according to their roles and responsibilities; and the specified retention time and the disclosure of personal data are in conformity with the purposes, in agreement with the consent and in compliance with applicable laws and regulations.

**Access.** This principle aims to give data subjects the ability to access, review and update their personal data to verify the accuracy of the data and the lawfulness of the processing, unless a law or regulation specifically requires otherwise. Access actions need to be restricted by conditions: the notice - including the means by which data subjects may gain access to their personal data - has been communicated to data subjects at or before the collection time; the identity of data subjects who request access to their personal data is authenticated; and the access request is in compliance with applicable laws and regulations.

Other concepts of the GPS principles - such as accuracy, security, openness, accountability and compliance - cannot easily be directly addressed by technical means. Accordingly, all relevant actions related to these aspects are treated as assumptions.

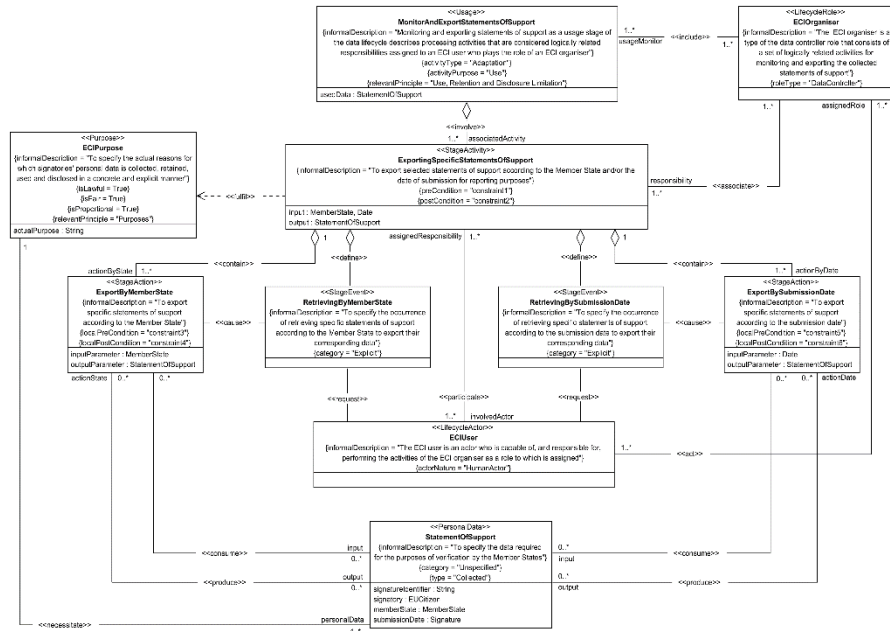
**Representation.** Due to space limitations, we consider only one concrete purpose to illustrate how to model the abstract and concrete purposes, along with the key concepts and actions derived from the abstract privacy principles, (see *Figure 3*).

ECIPurpose is a class stereotyped by «Purpose» to represent the abstract purpose for which signatories' personal data is collected and processed. At the domain level, its tag definitions - isFair, isLawful and isProportional - capture the main attributes of the concept of the purpose derived from the abstract privacy principles. At the instance level, the value of its actualPurpose attribute is to verify and certify the valid number of the statements of support for a proposed citizens' initiative. StatementOfSupport is a class stereotyped by «PersonalData» to represent the data required for the purpose of verification and certification by the Member States. At the domain level, its tag definitions - category and type - capture the main attributes of the concept of the personal data derived from the abstract privacy principles.

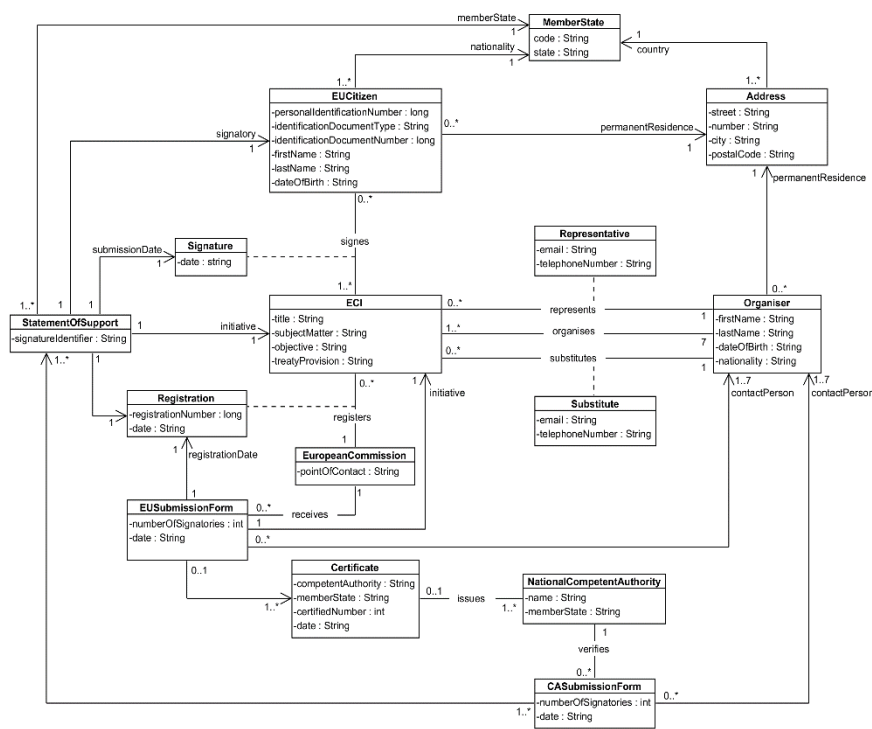
The mandatory fields that are required to sign up to an initiative vary between Member States. For illustrative purposes, we consider the mandatory fields of a statement of support form by France and Italy. These elements are modelled as StatementOfSupport class, which has four attributes: signatureIdentifier, signatory, memberState and submissionDate. The memberState is to capture the EU country that verifies a statement of support, whereas signatory is typed as EUCitizen. The EUCitizen is a class to represent the mandatory fields of a statement of support, as per *Figure 4*.

MonitorAndExportStatementsOfSupport is a class stereotyped by «Usage» to represent ExportingSpecificStatementsOfSupport as a data-processing activity among others that are considered logically related responsibilities. ExportingSpecificStatementsOfSupport is a class stereotyped by «StageActivity» as a concrete purpose to export the selected statements of support according to the Member State or the date of submission for reporting purposes. At the domain level, its tag definitions - preCondition and postCondition - capture the main constraints of its concrete actions derived from the abstract privacy principles. This activity coordinates its execution by containing two actions and defining two events that cause the execution of these actions. ExportByMemberState and ExportBySubmissionDate are classes stereotyped by «StageAction» to export specific statements of support according to the Member State and the submission date respectively. At the domain level, their tag definitions - preCondition and postCondition - capture the main constraints to be satisfied before and after the execution of the actions. RetrievingByMemberState and RetrievingBySubmissionDate are classes stereotyped by «StageEvent» to specify the occurrence of retrieving specific statements of support according to the Member State and the submission data respectively to export their corresponding data.

The MonitorAndExportStatementsOfSupport includes ECIOrganiser as a lifecycle role that is played by ECIUser as a lifecycle actor according to their capabilities and responsibilities. ECIOrganiser is a class stereotyped by «LifecycleRole» to represent a type of the data controller role that consists of a set of logically related activities for monitoring and exporting the collected statements of support. ECIUser is a class stereotyped by «LifecycleActor» to represent an actor who is capable of, and responsible for, performing the activities of the ECI organiser.



**Figure 3** The operationalisation of the abstract purpose.



**Figure 4** The data model of the ECI.

**Evaluation.** We now illustrate how to model constraints on the ECIPurpose and its concrete purpose in the usage stage of the lifecycle. The pre- and post-conditions of the activity and its associated actions are referred to by names in *Figure 3*. We consider each element in turn.

*ECIPurpose.* To ensure that the abstract purpose is a singleton, the following OCL expression needs to be established as an invariant.

```
context ECIPurpose inv:
    ECIPurpose.allInstances()->size()<=1
```

To ensure that the abstract purpose is fair, lawful and proportional, the following OCL expression needs to be established as an invariant. (We assume these have been assessed by competent, authorised actors.)

```
context ECIPurpose inv:
    self.isFair=true and
    self.isLawful=true and
    self.isProportional=true
```

*PersonalData.* To ensure that the use of personal data is limited to the abstract and concrete purposes for which data subjects have given their implicit or explicit consent, the following OCL expression needs to be established as an invariant.

```
context StatementOfSupport inv:
    self.actionState->forAll(ms:ExportByMemberState|
ms.outputParameter->includes(self)) and
    self.actionDate->forAll(sd:ExportBySubmissionDate|
sd.outputParameter->includes(self))
```

*ExportingSpecificStatementsOfSupport.* To ensure that the use of personal data is limited to the abstract and concrete purposes for which data subjects have given their implicit or explicit consent, the following OCL expression needs to be established.

```
context MonitorAndExportStatementsOfSupport::ExportingSpecificStatementsOfSupport(country : MemberState, date : Date ): StatementOfSupport
pre constraint1:
    self.associatedActivity.actionByState.input->includes(country) or
    self.associatedActivity.actionByDate.input->includes(date)
post constraint2: result
    =self.associatedActivity.actionByState.input->select(s:StatementOfSupport|s.memberState = country) or
    result =self.associatedActivity.actionByDate.input->select(s:StatementOfSupport|s.submissionDate=date)
```

*ExportByMemberState.* To ensure that the items of personal data consumed or produced by the action as inputs or outputs to accomplish, or as a result of, its execution, are attributes of the modelled personal data, we use the following OCL expression.

```

context ExportingSpecificStatementsOfSupport::ExportByMemberState(country : MemberState ) : StatementOfSupport
pre constraint3: self.actionByState.input.memberState->includes(country)
post constraint4: result=self.actionByState.input->select(s:StatementOfSupport|s.memberState=country)

```

*ExportBySubmissionDate.* To ensure that the items of personal data consumed or produced by the action as inputs or outputs to accomplish, or as a result of, its execution, are attributes of the modelled personal data, we use the following OCL expression.

```

context ExportingSpecificStatementsOfSupport::ExportBySubmissionDate(date : Date ) : StatementOfSupport
pre constraint5: self.actionByDate.input.submissionDate->includes(date)
post constraint6: result=self.actionByDate.input->select(s:StatementOfSupport|s.submissionDate=date)

```

*ECIOrganiser.* To ensure that the role has the concrete purpose as an assigned responsibility, we use the following OCL expression.

```

context ECIOrganiser inv:
self.responsibility->notEmpty() implies
self.usageMonitor.associatedActivity->includes(self.reponsibility)

```

*ECIUser.* To ensure that the use of personal data is limited to the actors who are assigned to the role to which the concrete purpose is assigned as a responsibility, we use the following OCL expression.

```

context ECIUser inv:
self.assignedRole.responsibility->notEmpty() implies
self.assignedRole.responsibility->includes(self.assignedResponsibility)

```

### 4.3 Summary

To facilitate reasoning about privacy compliance with legal frameworks and/or standards, four main activities need to be conducted.

The first activity concerns modelling the purposes for which personal data is collected and processed at a certain level of abstraction to facilitate compliance checking. The abstract purpose needs to be refined into a set of concrete purposes that can be assigned to actors as responsibilities according to their roles and capabilities. In addition, concrete purposes need to be represented in terms of actions and events from which the minimum amount necessary of personal data can be derived and modelled.

The second activity concerns deriving and modelling the key aspects of abstract privacy principles stated in legal frameworks and/or standards as concepts and actions. The concepts, along with their properties and possible values, represent the primary terms, whereas the actions, together with their pre- and post-conditions, assure that the processing of personal data is performed fairly and lawfully.

The third activity concerns modelling the abstract and concrete purposes together with the useful and potentially usable concepts and actions derived from abstract privacy principles. The stereotypes and tag definitions identified in the UML profile for the APDL model can be used to model the abstract and concrete purposes in terms of personal data, activities, actions, roles and responsibilities. The concepts and actions derived from the abstract privacy principles can be modelled using the stereotypes and tag definitions of the profile.

The fourth activity concerns establishing and modelling suitable rules that provide a set of criteria against which the requirements model is evaluated to determine whether it fulfils the privacy requirements. The established rules can be modelled as constraints using OCL to specify semantics and conditions that must be maintained for the identified stereotypes and tag definitions.

## 5 Conclusion

Through a realistic case study, we have demonstrated the usefulness and applicability of the extension mechanisms that serve as a stepping stone to formally reason about the compliance of a requirements model with abstract privacy principles. The standard extension mechanisms have enriched the APDL profile with specific properties and constraints suitable for the domain of privacy and data protection.

We argue that privacy compliance can be checked and assured through operationalizing the abstract purpose for which personal data is collected and processed in terms of personal data, data-processing activities, actions, events, roles and actors. It can be achieved by adopting the APDL model as a requirements model that can be expressed as a common language understood by those concerned with privacy and data protection, and those responsible for developing and maintaining privacy-aware systems.

We will build upon this work in a number of ways. First, we will use additional case studies to further validate the approach and highlight its usefulness and practical impact in various domains. Second, we will define a systematic methodology for privacy risk-assessment, which necessitates defining an appropriate privacy risk model, along with assessment and analysis approaches. This will also require the definition of a risk assessment process that illustrates the activities necessary to prepare for, conduct and communicate the results of, risk assessments.

## Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments.

## References

- American Institute of Certified Public Accountants and Canadian Institute of Chartered Accountants (AICPA/CICA), (2009), “*Generally Accepted Privacy Principles*”, available at: <https://www.cippguide.org/2010/07/01/generally-accepted-privacy-principles-gapp/> (accessed 20 February 2018).
- Alshammari, M. and Simpson, A.C. (2017a), “Personal Data Management: An Abstract Personal Data Lifecycle Model”, in *Proceedings of the Workshop on Security and Privacy-enhanced Business Process Management (SPBP'17)*, Springer, pp. 685-697.
- Alshammari, M. and Simpson, A.C. (2017b), “A UML Profile for Privacy-Aware Data Lifecycle Models”, in *Proceedings of the 1st International Workshop on Security and Privacy Requirements Engineering (SECPRE 2017)*, Springer, pp. 189-209.
- Antignac, T., Scandariato, R. and Schneider, G. (2016), “A Privacy-Aware Conceptual Model for Handling Personal Data”, in *Proceedings of the International Symposium on Leveraging Applications of Formal Methods*, Springer, pp. 942-957.
- Beckers, K. (2012), “Comparing Privacy Requirements Engineering Approaches”, in *Proceedings of the 2012 Seventh International Conference on Availability, Reliability and Security (ARES)*, IEEE, pp. 574-581.
- Cavoukian, A. (2006), “Creation of a Global Privacy Standard”, IPC.
- Cavoukian, A. (2011), “Privacy by Design: Origins, Meaning, and Prospects for Assuring Privacy and Trust in the Information Era”, in *Privacy Protection Measures and Technologies in Business Organizations: Aspects and Standards*, IGI Global.
- Cavoukian, A., Shapiro, S. and Cronk, R.J. (2014), “Privacy Engineering: Proactively Embedding Privacy, by Design”, IPC.
- European Commission (2012), “*The European Citizens' Initiative*”, available at: <http://ec.europa.eu/citizens-initiative/public/welcome> (accessed 20 February 2018).
- Gürses, S., Troncoso, C. and Diaz, C. (2011), “Engineering Privacy by Design”, *Computers, Privacy & Data Protection*, Vol. 14 No. 3, p. 25.
- Hansen, M., Jensen, M. and Rost, M. (2015), “Protection Goals for Privacy Engineering”, in *Security and Privacy Workshops (SPW)*, IEEE, pp. 159-166.
- Hoepman, J.H. (2014), “Privacy Design Strategies”, in *ICT Systems Security and Privacy Protection*, Springer, pp. 446-459.
- Jürjens, J. (2002), “UMLsec: Extending UML for Secure Systems Development”, in *International Conference on The Unified Modeling Language*, Springer, pp. 412-425.
- Kalloniatis, C., Kavakli, E. and Gritzalis, S. (2008), “Addressing Privacy Requirements in System Design: the PriS Method”, *Requirements Engineering*, Vol. 13 No. 3, pp. 241-255.
- Kost, M., Freytag, J.C., Kargl, F. and Kung, A. (2011), “Privacy Verification Using Ontologies”, in *Proceedings of the Sixth International Conference on Availability, Reliability and Security (AREs 2011)*, IEEE, pp. 627-632.

Kung, A. (2014), “PEARs: Privacy Enhancing Architectures”, in *Privacy Technologies and Policy: Second Annual Privacy Forum (APF 2014)*, Springer, pp. 18-29.

Martín, Y.S., Del Alamo, J.M. and Yelmo, J.C. (2014), “Engineering Privacy Requirements: Valuable Lessons from Another Realm”, in *Proceedings of the 1st Workshop on Evolving Security and Privacy Requirements Engineering (ESPRe)*, IEEE, pp. 19-24.

Mina, D., Kim, W., Riccardo, S., Bart, P. and Wouter, J. (2011), “A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements”, *Requirements Engineering*, Vol. 16 No. 1, pp. 3-32.

Object Management Group (2012), “*OMG Object Constraint Language (OCL)*”, available at: <http://www.omg.org/spec/OCL/2.3.1/> (accessed 20 February 2018).

Object Management Group (2015), “*OMG Unified Modeling Language (OMG UML)*”, available at: <http://www.omg.org/spec/UML/> (accessed 20 February 2018).

Oetzel, M.C. and Spiekermann, S. (2014), “A Systematic Methodology for Privacy Impact Assessments: A Design Science Approach”, *European Journal of Information Systems*, Vol. 23 No. 2, pp. 126-150.

Onn, Y., Geva, M., Druckman, Y., Zyssman, A., Timor, R., Lev, I., Maroun, A., Maron, T., Nachmani, Y., Simsolo, Y., Sicklai, S., Fuches, A., Fishman, M., Packer, S. and Pery, L. (2005), “Privacy in the Digital Environment”, Haifa Center of Law & Technology, pp. 1-12.

Shapiro, S.S. (2010), “Privacy by Design: Moving from Art to Practice”, *Communications of the ACM*, Vol. 53 No. 6, pp. 27-29.

Spiekermann, S. (2012), “The Challenges of Privacy by Design”, *Communications of the ACM*, Vol. 55 No. 7, pp. 38-40.

Spiekermann, S. and Cranor, L.F. (2009), “Engineering Privacy”, *IEEE Transactions on Software Engineering*, Vol. 35 No. 1, pp. 67-82.

The European Parliament (2016), “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)”, *Official Journal of the European Union*, Vol. 59 No. L 119, pp. 1-88.