COPri - a Core Ontology for Privacy Requirements Engineering

Mohamad Gharib[1✉], John Mylopoulos[2] and Paolo Giorgini[2]

[1] University of Florence - DiMaI, Viale Morgagni 65, Florence, Italy
mohamad.gharib@unifi.it
[2] University of Trento - DISI, 38123, Povo, Trento, Italy
john.mylopoulos, paolo.giorgini@unitn.it

Abstract. In their daily practice, most enterprises collect, store, and manage personal information for customers in order to deliver their services. In such a setting, privacy has emerged as a key concern as companies often neglect or even misuse personal data. In response to this, governments around the world have enacted laws and regulations for privacy protection. These laws dictate privacy requirements for any system that acquires and manages personal data. Unfortunately, these requirements are often incomplete and/or inaccurate as many RE practitioners might be unsure of what exactly are privacy requirements and how are they different from other requirements, such as security. To tackle this problem, we developed a comprehensive ontology for privacy requirements. To make it comprehensive, we base our ontology on a systematic review of the literature on privacy requirements. The contributions of this work include the derivation of an ontology from a previously conducted systematic literature review, an implementation using an ontology definition tool (Protégé), a demonstration of its coverage through an extensive example on Ambient Assisted Living, and a validation through a competence questionnaire answered by lexical semantics experts as well as privacy and security researchers.

Keywords

Privacy Ontology, Privacy Requirements, PbD, Conceptual Modeling

1 Introduction

It is common practice for most companies today to collect, store, and manage personal information to deliver their services. Therefore, privacy has emerged as a key concern since such companies need to protect the privacy of personal information in order to comply with various privacy laws and regulations (e.g., GDPR in the EU [1]) that many governments have enacted for privacy protection. Accordingly, dealing with privacy concerns is a must these days [2]. However, most of such concerns can be tackled if the privacy requirements of the system-to-be were considered and addressed properly during requirements engineering [3,4].

Unfortunately, most requirements engineers are unfamiliar with privacy requirements and how they differ from other requirements, such as security or vanilla quality requirements [5]. Even when requirements engineers have familiarity with privacy concerns, they focus mainly on confidentiality, and overlooking important privacy aspects such as unlinkability, unobservability [3].

Although privacy concepts have been studied for more than a century, they are still elusive and vague concepts to grasp [6,3]. In recent years, there have been numerous attempts to define privacy based on various related concepts such as confidentiality, anonymity, risk, transparency, etc. [6,7,8,9]. However, there is no consensus on the definitions of many of these concepts nor which of them should be used to analyze privacy [6]. In addition, many of these concepts are overlapping, thereby contributing to the confusion while dealing with privacy. Ontologies have proven to be a key factor for reducing the conceptual vagueness and terminological confusion by providing a shared understanding of related concepts [10]. In this context, the main objective of this work is to propose, implement, evaluate and validate a well-defined ontology that captures key privacy-related concepts.

Privacy is a social concept in that it depends on how others treat an individual's personal information and it strongly depends on the social context where that information is captured and used [5]. Accordingly, the privacy ontology should conceptualize privacy in their social and organizational context. In previous research [5], we worked toward addressing this problem by proposing a preliminary ontology for privacy requirements that has been mined through a systematic literature review.

In this paper, we propose COPri (a Core Ontology for Privacy requirements engineering) that has been mined from the results of what is proposed in [5] with new and more refined concepts concerning both personal information and privacy. Moreover, we implement the ontology, apply it to an Ambient-Assisted Living (AAL) illustrative example, and then validate it by querying the ontology instance (the AAL example) depending on a set of competency questions. Finally, we evaluate the ontology against common pitfalls in ontologies with the help of some tools, lexical semantics experts, and privacy and security researchers.

The rest of the paper is organized as follows; Section (§2) presents an illustrative example, and we describe the process we followed for developing COPri in Section (§3). Section (§4) presents the conceptual model of COPri, and we implement and validate COPri in Section (§5) and (§6) respectively. We evaluate the ontology in Section (§7). Related work is presented in Section (§8), and we conclude and discuss future work in Section (§9).

2 Illustrating example: the Ambient-Assisted Living (AAL) System

Our motivating example concerns an old person called Jack that suffers from diabetes disease. Jack lives in a home that is equipped with an AAL system, which relies on various interconnected body sensors (e.g., Continuous Glucose

Monitoring (CGM), location, and motion sensors). These sensors collect various information about Jack's vital signs, location, and activities. This information is transmitted to Jack's Personal Digital Assistant (PDA) that assesses his health situation and provides required notifications accordingly. Jack's PDA may also forward such information to a nearby caring center, where a nurse called Sarah can monitor such information, and she can also monitor some of Jack's activities (e.g., watching TV, sleeping, etc.) by collecting location and motion related-information. Sarah can detect unusual situations and react accordingly, she also has access to all Jack's health records and she may contact the required medical professional that might be needed depending on Jack's situation. Jack, like many other users, wants to preserve his privacy by controlling what is collected and shared concerning his personal information, who is using such information, and for which reasons.

3 The process for developing the COPri ontology

The process for developing COPri (depicted in Figure 1) has been constructed based on [11,12], and it is composed of five main phases, two of them (in gray) were addressed in [5] while the remaining three are addressed in this paper:

- *Step 1. Scope & objective identification,* COPri aims at assisting software engineers while designing privacy-aware systems by providing a generic and expressive set of key privacy concepts and relationships, which enable for capturing privacy requirements in their social and organizational context.
- *Step 2. Knowledge acquisition* aims at identifying and collecting knowledge needed for the construction of the ontology. In [5], we have conducted a systematic literature review for identifying the concepts and relationships used in the literature for capturing privacy requirements as well as the semantic mappings between them[3]. The systematic literature review has identified 38 privacy-related concepts and relationships.
- *Step 3. Conceptualization* aims at deriving an ontology that consists of key concepts and relationships for privacy [12]. In [5], we have proposed a preliminary ontology consisting of 38 concepts and relationships. In this paper, we extend and refine our earlier proposal to a comprehensive ontology consisting of 52 concepts and relationships.
- *Step 4. Implementation* aims at codifying the ontology in a formal language. This requires an environment that guarantees the absence of lexical and syntactic errors from the ontology, and an automated reasoner to detect inconsistencies and redundant knowledge.
- *Step 5. Evaluation and Validation* aims at ensuring that the resulting ontology meets the needs of its usage [12]. Following [13], we validated COPri by applying it to the Ambient-Assisted Living (AAL) illustrating example and querying the ontology instances depending on Competency Questions (CQs). Then, evaluating whether the ontology captures enough detailed knowledge about the targeted domain to fulfill the needs of its intended use.

[3] A detailed version of the systematic literature review can be found at [14]

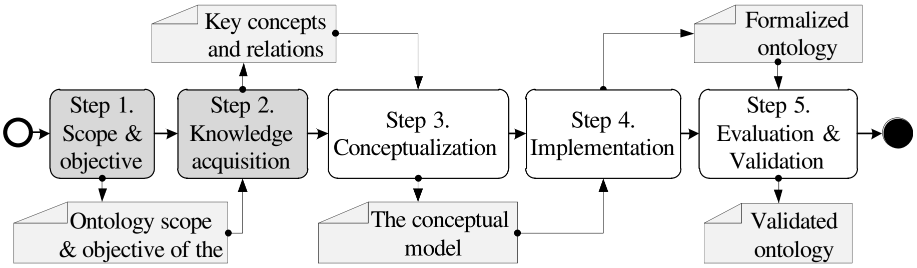
Key concepts and relations
Formalized ontology
Step 1. Scope & objective
Step 2. Knowledge acquisition
Step 3. Conceptualization
Step 4. Implementation
Step 5. Evaluation & Validation
Ontology scope & objective of the
The conceptual model
Validated ontology

Fig. 1: The process for developing the COPri ontology

4 The COPri ontology

The ontology is presented as a UML class diagram in Figure 2. For reasons of readability, multiplicity and other constraints have been left out. The concepts of the ontology are organized into four main dimensions:

(1) Organizational dimension includes concepts for capturing the social and organizational aspects of the system, which are organized into several categories:

Agentive entities captures the active entities of the system, and it includes the following concepts: *Actor* represents an autonomous entity that has intentionality and strategic goals, and it covers two entities: a *Role* represents an abstract characterization of an actor in terms of a set of behaviors and functionalities. A role can be a specialization (*is_a*) of one another; an *Agent* represents an autonomous entity that has a specific manifestation, and it can *play* a role or more, where an agent inherits the properties of the roles it plays.

Intentional entities includes the following concepts: *a goal* is a state of affairs that an actor **aims** to achieve. When a goal is too coarse to be achieved, it can be refined through *and/or-decompositions* of a root goal into finer subgoals, where the first implies that the achievement of the root-goal requires the achievement of all of its sub-goals, and the latter implies that the achievement of the root-goal requires the achievement of any of its sub-goals.

Informational entities includes the following concepts: *Information* represents a statement provided or learned about something or someone. Information can be atomic or composed of several parts, and we rely on *partOf* relationship to capture the relationship between an information entity and its sub-parts. We differentiate between two types of information: *Public information,* any information that cannot be *related* (directly or indirectly) to an identified or identifiable legal entity, and *Personal information,* any information that can be *related* to an identified or identifiable legal entity (e.g., medical records).

Sensitivity level & situation, personal information has a *sensitivity level* [15,4]. Based on [16], we adopt four different sensitivity levels ordered as *(R)estricted*, *(C)onfidential*, *(S)ensitive*, and *Secre(T)*, where *Secre(T)* is the most sensitive. Moreover, the sensitivity of personal information can be linked to when and where such information has been collected and for what purposes,

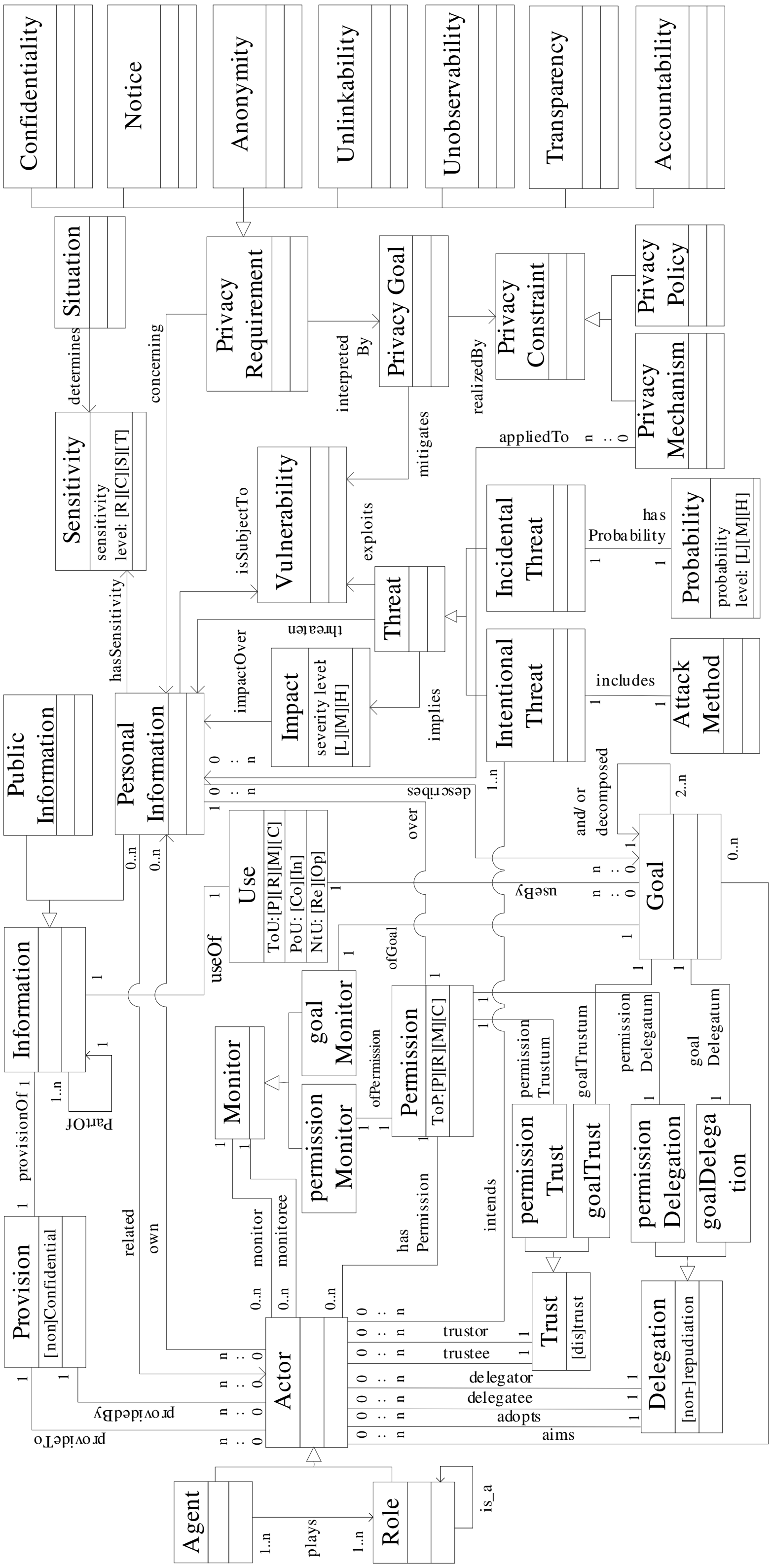

Fig. 2: The conceptual model of COPri

i.e., the context/state of affairs related to such information. Thus, we also adopt the concept of *situation* as a mean to determine the sensitivity level.

Information use is a relationship between a goal and information, and it has three attributes: *(1) Type of Use (ToU),* our ontology provide four types of use, we consider sufficient for covering main information processing tasks: *Produce, Read, Modify,* and *Collect,* indicates that information is created, consumed, altered and acquired respectively. *(2) Need to Use (NtU)* captures the necessary of use that has two types: *Require* and *Optional*, wherein the first the use of information is required for the goal achievement, and in the later is not [17]. *(3) Purpose of Use (PoU),* we differentiate between two types: *Compatible* and *Incompatible*, where the first indicates that the purpose for which information is used is compliant with the rules that guarantee the best interest of its owner; and in the later, it is not compliant.

Describes is a relationship where information characterizes a goal (activity) while it is being pursued by some actor[4].

Ownership & Permissions includes the following concepts: *Own* indicates that an actor is the legitimate owner of information. *Permission* is a consent that identifies a particular use of a particular information in a system. Information owner (data subject[5]) has full control over the use of information it owns, and it depends on *permissions* for such control. In COPri, a permission has a type that can take as values (P)roduce, (R)ead, (M)odify and (C)ollect, which cover the four relationships between goals and information that our ontology proposes.

Entity interactions: the ontology adopts three types of interactions: *(1) Information provision* captures the transmission of information among actors, and it has a type that can be either *confidential* or *nonConfidential*, where the former guarantee the confidentiality of the transmitted information, while the last does not. *(2) Delegation* indicates that actors can delegate obligations and entitlements to others, where the source of delegation called the delegator, the destination is called delegatee, and the subject of delegation is called delegatum. The concept of *delegation* is further specialized into two concepts: *Goal delegation,* where the delegatum is a goal; and *Permission delegation,* where the delegatum is a permission. *(3) Adoption* is considered a key component of social commitment, and it indicates that an actor accepts to take responsibility for the delegated objectives and/ or entitlements from another actor.

Entities social trust: the need for trust arises when actors depend on one another for goals or permissions since such dependencies might entail risk [18]. *Trust* has a type that can be either: *(1) Trust* means the trustor expects that the trustee will behave as expected considering the trustum (e.g., a trustee will not misuse the trustum), and *(2) Distrust* means the trustor expects that the trustee may not behave as expected considering the trustum. Moreover, the concept of

[4] The Ontology has been extended with *Collect* and *Describes* to capture situations when information *describing* some activities performed by a data subject (personal information) is being *collected* by others

[5] We treat “information owner” and “data subject” as synonyms

Trust is further specialized into two concepts *GoalTrust,* where the trustum is a goal; and *PermissionTrust,* where the trustum is a permission.

Monitoring: is the process of observing and analyzing the performance of an actor in order to detect any undesirable performance. We adopt the concept of *monitoring* to compensate for the lack of trust or distrust in the trustee concerning the trustum. The concept of *monitor* is further specialized into two concepts *GoalMonitor,* where the subject of the monitoring is a goal; and *PermissionMonitor,* where the subject of the monitoring is a permission.

(2) Risk dimension includes risk related concepts that might endanger privacy needs at the social and organizational levels:

A vulnerability is a weakness in the current state-of-affairs that may be *exploited* by a *threat.*

A threat is a potential incident that *threatens* personal information by *exploiting* a *vulnerability* concerning such information [19]. *Threat* has a *probability* that measures the likelihood of its occurrence, and it is characterized by three different values *high*, *medium* or *low.* In COPri, we differentiate between two types of threat: *(1) Incidental threat* that is a casual, natural or accidental threat that is not caused by a *threat actor* nor does it require an *attack method. (2) Intentional threat* is a threat that require a *threat actor* and *includes* a presumed *attack method* [14].

Threat actor is an actor that intends to achieve an *intentional threat* [19].

Attack method is a standard means by which a *threat actor* carries out an *intentional threat* [19,10].

Impact is the expected consequence of a *threat* over the personal information. An *impact* has a *severity* that captures the level of the impact [10], and takes values *high*, *medium* or *low.*

(3) Treatment dimension includes concepts to mitigate risks:

A privacy goal defines an intention to counter threats and prevent harm to personal information by satisfying privacy properties.

A privacy constraint is a design restriction that is used to realize/satisfy a privacy goal, constraints can be either a privacy policy or privacy mechanism.

A privacy policy defines permitted and forbidden actions to be carried out by actors toward information.

A privacy mechanism is a concrete technique that operationalizes a privacy goal. Some mechanisms can be directly *applied to personal information* (e.g., anonymity, unlinkability).

(4) Privacy dimension includes concepts to capture the actors' privacy requirements/needs concerning their personal information:

Privacy requirements capture information owners' privacy needs. *Privacy requirements* can be *interpretedBy privacy goals*, and it is further specialized into seven more refined concepts[6]:

[6] The right to erasure (right to be forgotten) is essential in several privacy laws, yet we did not consider it since the use of information is limited to a specific, explicit, legitimate purpose (a goal), i.e., information will not be kept after achieving the goal

Confidentiality means personal information should remain inaccessible to incidental or intentional threats [6,15,4]. We rely on three principles to analyze confidentiality: *(1) Non-disclosure,* personal information can only be disclosed if the owner's consent is provided [6,15,4]. Therefore, *non-disclosure* can be analyzed depending on the existence of read permission as well as the confidentiality of information provision. *(2) Need to Know (NtK),* can be analyzed depending on *Need to Use (NtU)* that captures the necessity of use, i.e., personal information can only be used if it is strictly necessary for completing a certain task [4]. *(3) Purpose of Use (PoU),* personal information can only be used for specific legitimate purposes and not in ways that are incompatible with those purposes [6,15], i.e., if the *PoU* is *compatible* with the rules that guarantee the best interest of its owner.

Anonymity means personal information can be used without disclosing the identity of its owner [15,6,7]. Personal information can be *anonymized* (e.g., removing identifiers) depending on some *privacy mechanism.*

Unlinkability means that it should not be possible to link personal information back to its owner [20,3,7]. A *privacy mechanism* can be used to remove any linkage between personal information and its owner.

Unobservability means the identity of information owner should not be observed by others, while performing an activity [3,7]. *Unobservability* can be analyzed relying on the *describes* relationship, which enables for detecting situations where personal information that describes an activity (goal) being pursued by a data subject is being collected by some other actor [21].

Notice means information owner should be notified when its information is being collected [6,15]. *Notice* can be analyzed depending on collect relationship and its corresponding permission. In the case where personal information is being collected and there is no permission to collect it, a notice violation will be raised. Providing a permission to collect implies that the actor has been already notified and agreed upon the collection of his information.

Transparency means information owner should be able to know who is using its information and for what purposes [15], we rely on two principles to analyze transparency: *(1) Authentication* a mechanism aims at verifying whether actors are who they claim they are, and it can be analyzed by verifying whether i) the actor is playing a role that enables the identification of its main responsibilities; and ii) the actor is not playing any threat actor role. *(2) Authorization* a mechanism aims at verifying whether actors can use information in accordance with their credentials [15].

Accountability means information owner should be able to hold information users accountable for their actions concerning its information [15]. We rely on the *non-repudiation* principle to analyze accountability, which can be analyzed relying on the adoption relationship, i.e., if a delegatee did not adopt the delegatum, a *non-repudiation* violation can be raised.

This ontology extends the one proposed in [5] by new concepts concerning personal information and privacy requirements. Accordingly, we have extended and refined the organizational, risk, treatment and privacy dimensions to cover the new extensions, and to allow for performing a more comprehensive analysis.

5 The implementation of COPri[7]

We have implemented the COPri ontology[8] using the on Protégé tool[9] that supports the creation, modification, visualization and consistency-checking for an ontology. Protégé also offers a plug-in for using SPARQL to query an ontology. In particular, we have implemented COPri relying on classes and object properties (relationships) in Protégé, had to amend and/or create new classes and relationships during this process. Moreover, for each class that has attributes with quantitative values, we have created a class (called a Value Partition pattern) to present such attributes, and several individuals (instances) to cover all quantitative values of their corresponding attributes.

In our implementation, all *primitive siblings* classes (e.g., Personal and Public Information) have been made *disjoint*, which helps the reasoner to detect inconsistencies. Moreover, we have used Probe Classes, which are classes that are subclasses of two or more disjoint classes to test and ensure that the ontology does not include inconsistencies. Additionally, we have used a covering axiom to solve the open-world assumption in OWL-based ontologies, where a covering axiom is a class that results from the union of the classes being covered. Properties are used to link individuals from domain to range classes. Thus, we have defined the domain and range for each of the object properties, which can be used by the reasoner to make inferences and detect inconsistencies. Moreover, we defined only one inverse property to minimize the number of object properties. Finally, we have used cardinality restrictions to specify the number of relationships between classes depending on at *least*, at *most* or *exactly* keywords.

6 The validation of COPri[10]

We validated the COPri ontology by applying it to the AAL illustrating example, and then query the ontology instance relying on Competency Questions (CQs) and check whether these queries can return comprehensive answers. In particular, CQs represent a set of queries that the ontology must be capable of answering to be considered competent for conceptualizing the domain it was intended for [11,13]. The CQs are meant to assist and guide requirements engineers while dealing with privacy requirements by capturing main wrong/bad design decisions (we call *violations*) related to the four dimensions of our ontology. 26 CQs[11] have been defined (shown in Table 1), which we consider sufficient for capturing all violations to the privacy requirements considered in our ontology.

In particular, **CQ1-3** are dedicated for organizational aspects, e.g., identifying violations related to permissions delegation without trust or monitoring

[7] Available in greater detail in [22]

[8] The COPri ontology is available in OWL formal at `https://goo.gl/AaqUxx`.

[9] `http://protege.stanford.edu/`

[10] Available in greater detail in [22], formalization of the CQs (SPARQL queries), and the validation we performed

[11] Note that the main focus of the CQs is privacy requirements, not goal analysis

Table 1: Competency Questions for validating the COPri ontology

Organizational dimension	
CQ1.	Who are the delegators that delegate produce, read, modify, or collect permission, which is not accompanied by trust nor monitoring?
CQ2.	Who are the delegators that delegate produce, read, modify, or collect permission accompanied by both trust and monitoring?
CQ3.	Which is the personal information of sensitivity Restricted [Confidential, Sensitive or Secret]?
Risk dimension	
CQ4.	Which are the existing vulnerabilities and which personal information are subject to them?
CQ5.	Which are the existing vulnerabilities and which are the threats that can exploit them?
CQ6.	Which are the existing vulnerabilities that are not mitigated by privacy goals?
CQ7.	Which are the existing threats and which is the personal information that are threatened by them?
CQ8.	Which are the existing threats that have an impact with severity level Low [Medium, High] over personal information?
CQ9.	Which are the existing intentional threats and which is the personal information that are threatened by them?
CQ10.	Who are the threat actors and which are the intentional threats that they intend to perform?
CQ11.	Which are the existing attack methods and to which intentional threats they can be used for?
CQ12.	Which are the existing incidental threats and which is the personal information that are threatened by them?
CQ13.	Which are the existing threats of probability Low [Medium | High]?
Treatment dimension	
CQ14.	Which are the privacy goals that are realized by privacy constraints?
CQ15.	Which are the existing privacy mechanisms and which is the personal information that such mechanisms are applied to?
Privacy dimension	
CQ16.	Which is the personal information that is read without read permission?
CQ17.	Which is the personal information that is transferred relying on non-confidential provision?
CQ18.	Which is the personal information that is used by a goal, where their usage (NtU) is not strictly required (i.e., optional)?
CQ19.	Which is the personal information that is used by goals, where their purpose of use (PoU) is incompatible with the best interest of its owner?
CQ20.	Which is the personal information that is not anonymized?
CQ21.	Which is the personal information that can be linked back to their owners?
CQ22.	Which is the personal information that describes a goal, and it is also being collected by some actor?
CQ23.	Who are the actors that are collecting personal information without collect permissions?
CQ24.	Who are the actors that do not play any role or they play a threat actor role?
CQ25.	Who are the actors that are using (producing, reading, modifying, or collecting) personal information without the required permission?
CQ26.	Who are the delegatees that have not adopted their delegatum?

(*CQ1*), the existing of trust and monitoring concerning the same trustum that is considered as a bad design decision (*CQ2*), and *CQ3* can be used for returning different sets of personal information based on their sensitivity levels (e.g., Secret, Sensitive, etc.).

CQ4-13 are dedicated for risk aspects, e.g., identify violations related to existing vulnerabilities and information subject to them (*CQ4*), threats that can exploit such vulnerabilities (*CQ5*), unmitigated vulnerabilities (*CQ6*), existing threats (*CQ7*), and *CQ8* can be used to identify threats based on the severity levels of their impact (e.g., Low, Medium, or High). *CQ9-11* can be used to identify existing intentional threats, threat actors and attack methods respectively. While *CQ12* can be used to identify existing incidental threats, and *CQ13* is used to identify different sets of threats based on their probability levels (Low, Medium, or High).

CQ14-15 are dedicated for treatment aspects, e.g., identify violations related to unrealized privacy goals (*CQ14*) and *CQ15* can be used to identify privacy mechanisms and personal information that such mechanisms are applied to.

CQ16-26 are dedicated for privacy requirements violations. In particular, *CQ16-19* are used for analyzing *Confidentiality*, where *CQ16-17* are used for analyzing non-disclosure by detecting and reporting when personal information is read without the owner's permission (*CQ16*), or it has been transferred relying on non-confidential transmission means (*CQ17*). *CQ18* is used for analyzing Need to Know (NtK) principle by verifying whether personal information is strictly required by goals using them, i.e., if the Need to Use (NtU) of the goal is optional, *CQ18* will report such violation. *CQ19* is used for analyzing the Purpose of Use (PoU) principle by verifying whether personal information is used for specific, explicit, legitimate purposes that have been permitted to be used for, i.e., if the PoU is incompatible, *CQ19* will report such violation. *CQ20* is used for analyzing *Anonymity* by verifying whether the identity of information owner can be sufficiently identified, i.e., if personal information has not been anonymized relying on a privacy mechanism, *CQ20* will report such violation. *CQ21* is used for analyzing *Unlinkability* by verifying whether it is possible to link personal information back to its owner, i.e., if an unlinkability mechanism has not been applied to personal information, *CQ21* will report such violation.

CQ22 is used for analyzing *Unobservability* by verifying whether the identity of information owner can be observed by others while performing some activity. Consider for example that Jack does not want his activities to be monitored while he is in the bathroom. Then, "Jack's location" should not be collected when he is in the bathroom since such information can be used to infer activities that Jack does not want it to be observed. If such information is collected, *CQ22* will report such violation. *CQ23* is used for analyzing *Notice* by verifying whether personal information is being collected without notifying its owner. In case, personal information is being collected and there is no permission to collect, *CQ23* will detect and report such violation.

CQ24-C25 are used for analyzing *Transparency*, where *CQ24* analyzes the authentication principle by verifying whether an actor can be authenticated based on the role(s) she/he is playing[12]. Accordingly, *CQ24* will report whether an actor can be authenticated. While *CQ25* analyzes the authorization principle by verifying that actors are not using personal information without the required

[12] If an actor is not playing any role, it will be impossible to authenticate it

permissions. Finally, *CQ26* is used for analyzing *Accountability* relying on the non-repudiation principle by verifying that actors cannot repudiate that they accepted delegations, which can be done depending on the adoption concept, if there exists a delegatee without an adopt relationship to the delegatum, *CQ26* will detect and report such violation.

The formulation of the CQs was an iterative process i.e., several CQs have been refined before having the final set of CQs. Note that the concepts of the ontology have been refined and extended as well while formulating the CQs because some limitations in the ontology have been revealed.

7 Evaluation

We evaluate the COPri ontology against the common pitfalls for ontologies identified in [23], where the authors classify 20 of these pitfalls by criteria under 1- *Consistency pitfalls* verify whether the ontology includes or allows for any inconsistencies; 2- *Completeness pitfalls* verify whether the domain of interest is appropriately covered; and 3- *Conciseness pitfalls* verify whether the ontology includes irrelevant elements or redundant representations of some elements with respect to the domain to be covered. The pitfalls classification by criteria is shown in Table 2, where we can also identify the four different methods we followed to evaluate the COPri ontology:

1- Protégé & HermiT Reasoner[13]: Both Protégé & HermiT have been used. In particular, HermiT is able to detect cycles in the hierarchy (*P6.*). *P4.* has been verified depending on OntoGraf plug-in that enables for visualizing the ontology. Concerning *P10.*, we have already made all *primitive siblings* classes *disjoint*. We have manually checked whether the domain and range of all object properties have been defined (*P11.*). Moreover, we verified *P14.* depending on Probe Classes. COPri ontology cannot suffer from *P15.* since we did not use complement operators to describe/define any of the classes, i.e., all defined classes have been defined depending on both necessary and sufficient conditions. The concepts of the ontology are general enough to avoid both *P17.* and *P18.*. No miscellaneous class have been identified (*P21.*), since the names of all classes and their sub-classes have been carefully chosen.

2- Evaluation with OntOlogy Pitfall Scanner (OOPS!): OOPS! is a web-based ontology evaluation tool[14] for detecting common pitfalls in ontologies. The COPri ontology was uploaded to the OOPS! pitfall scanner, which returned an evaluation report[15]. In particular, two suggestions have been returned, proposing to characterize both is_a and partOf relationships as symmetric or transitive. We took these suggestions into account, characterizing both of these relationships as transitive. 53 minor pitfalls (*P13.*) have been identified. However, as mentioned earlier we defined only one inverse property to minimize the number

[13] http://www.hermit-reasoner.com/

[14] http://oops.linkeddata.es/index.jsp

[15] Evaluation with OOPS! has been performed after evaluating the ontology with Protégé & HermiT, i.e., several pitfalls have been already detected and corrected

Table 2: Pitfalls classification by criteria and how they were evaluated

			Protégé	OOPS!	Experts	Researchers
Consistency	**P1.**	Creating polysemous elements	-	-	✓	-
	P5.	Defining wrong inverse relationships	-	✓	-	-
	P6.	Including cycles in the hierarchy	✓	✓	-	-
	P7.	Merging different concepts in the same class	-	✓	✓	-
	P14.	Misusing "allValuesFrom"	✓	-	-	-
	P15.	Misusing "not some" and "some not"	✓	-	-	-
	P18.	Specifying too much the domain or the range	✓	-	-	-
	P19.	Swapping intersection and union	-	✓	-	-
	P24.	Using recursive definition	-	✓	✓	-
Completeness	**P4.**	Creating unconnected ontology elements	✓	✓	-	-
	P9.	Missing basic information	-	-	-	✓
	P10.	Missing disjointness	✓	✓	-	-
	P11.	Missing domain or range in properties	✓	✓	-	-
	P12.	Missing equivalent properties	-	✓	-	-
	P13.	Missing inverse relationships	-	✓	-	-
	P16.	Misusing primitive and defined classes	✓	-	-	-
Conciseness	**P2.**	Creating synonyms as classes	-	✓	✓	-
	P3.	Creating the relationship "is" instead of using "subclassOf", "instanceOf" or "sameIndividual"	-	✓	-	-
	P17.	Specializing too much a hierarchy	✓	-	✓	-
	P21.	Using a miscellaneous class	✓	✓	✓	-

of properties/relationships in the ontology. Finally, only one critical pitfall has been identified stating that we are using is_a relationship instead of using OWL primitives for representing the subclass relationship (rdfs:subClassOf). However, is_a relationship is used in most Goal-based modeling languages, where we have adopted many of the concepts and relationships of the COPri ontology. Therefore, we chose not to replace it with the subClassOf relationship.

3- Lexical semantics experts: Two lexical semantics experts with main focus on Natural Language Processing (NLP) have been provided with the COPri ontology, and they were asked to check whether the ontology suffers from *P1, P2, P7, P17, P21,* and *P24* pitfalls[16]. Several issues have been raised by the experts concerning *P2*, *P21* and *P24*. Each of these issues has been properly addressed. The experts' feedback and how it was addressed can be found in [22].

4- A survey with researchers: The main purpose of this survey was evaluating the adequacy and completeness of the COPri ontology in terms of its concepts and relationships for dealing with privacy requirements in their social and organizational context (*P9.*). The survey was closed, i.e., it was accessible through a special link that is provided to the invited participants only to

[16] The experts evaluation template can be found at `https://goo.gl/ZEhLnN`

avoid unintended participants. In total 25 potential participants were contacted to complete the survey, and they were asked to forward the email to anyone who fits in the participating criteria (e.g., has good experience in privacy and/or security). We have received 16 responses (64% response rate). The survey template[17] is composed of four main sections: *S1. General information about the survey, S2. Participant demographics, S3. Evaluation questions,* and *S4. Final remarks.*

S2. Result of demographic questions: 15 (93.8%) of the participants are researchers and 1 (6.2%) is a student. Concerning experience with privacy and/or security: 2 (12.5%) of the participants have both academic and industrial experience, and 14 (87.5%) have pure academic experience. Moreover, 3 (18.8%) have less than one year, 7 (43.8%) have between one and four years, and 6 (37.5%) have more than four years of experience.

S3. Result of evaluation questions: this section is composed of 10 subsections, each of them is dedicated to collect feedback concerning the adequacy and completeness of a specific dimension/category of concepts and relationships. In each of these subsections, we provide the definitions of the concepts and relationships of the targeted dimension/category as well as a diagram representing them. Followed by a mandatory question, asking the participant to grade the completeness of the presented concepts and relationships with respect to system aspects they aim to capture on a scale from 1 (incomplete) to 5 (incomplete). The result of the evaluation for each of these sections is summarized in Table 3. The result tends to demonstrate that most of the targeted dimension/category of concepts and relationships are properly covering the aspects they aim to represent.

Additionally, we have added an optional question in each of the 10 sections to evaluate the adequacy of the concepts and relationships by collecting suggestions to improve the category/dimension under evaluation. Some feedback suggested to refine, include or exclude some of the concepts/relationships, we took some of these suggestions into account while developing the final ontology.

S4. Result of remarks question: most of the feedback was valuable, has raised important issues and ranged from complementing to criticizing. For example, among the encouraging feedback, we received *"COPri covers a wide range of privacy-related concepts, with actor and goal-oriented perspectives, which looks promising. We look forward to seeing it used to capture real-world privacy problem context"*. Another feedback and suggestion was *"I think it is very precise and very good work. Maybe some other concepts could be expressed somewhere"*. One of the comments we received was *"How satisfaction of privacy requirements can be verified using it?"*. We also received criticisms such as the following one *"I have no idea how good it is unless it is applied to many real cases. I'm concerned that it is not grounded in reality. It's also very complicated, which makes it hard to apply in the industry"*. However, such criticism opens the way for future research directions.

Threats to the validity of our study, we have identified the following threats: *1. Authors' background,* the authors have good experience in goal modeling (es-

[17] The survey template can be found at https://goo.gl/bro8nG

Table 3: The result of the evaluation

	Strongly disagree	**Disagree**	**N. agree/ n. disagree**	**Agree**	**Strongly agree**
Q1. Agentive cat.	0 (0%)	1 (6.3%)	3 (18.8%)	6 (37.5%)	6 (37.5%)
Q2. Intentional cat.	0 (0%)	1 (6.3%)	4 (25.0%)	7 (43.8%)	4 (25.0%)
Q3. Informational cat.	0 (0%)	2 (12.5%)	4 (25.0%)	4 (25.0%)	6 (37.5%)
Q4. Goals & info cat.	0 (0%)	2 (12.5%)	2 (12.5%)	6 (37.5%)	6 (37.5%)
Q5. Ownership cat.	0 (0%)	1 (6.3%)	1 (6.3%)	5 (31.3%)	9 (56.3%)
Q6. Interactions cat.	0 (0%)	1 (6.3%)	1 (6.3%)	6 (37.5%)	8 (50.0%)
Q7. Social Trust cat.	0 (0%)	0 (0.0%)	4 (25.0%)	7 (43.8%)	5 (31.3%)
Q8. Risk dim.	0 (0%)	3 (18.8%)	0 (0.0%)	8 (50.0%)	5 (31.3%)
Q9. Treatment dim.	0 (0%)	0 (0.0%)	3 (18.8%)	7 (43.8%)	6 (37.5%)
Q10. Privacy dim.	0 (0%)	2 (12.5%)	2 (12.5%)	5 (31.3%)	7 (43.8%)

pecially in *i** languages). This may have influenced the selection and definitions of the concepts of the ontology. However, *i** languages have been developed to capture requirements in their social and organizational context, which is also a main objective of our ontology. *2. Survey result validity,* the number of participants can raise concerns about the validity of the result. However, most of them are experts with good experience in privacy. *3. Extensive evaluation,* the ontology has been evaluated against the common pitfalls in ontologies with the help of some tools, lexical semantics experts, and privacy researchers, yet it has not been applied in industry. However, applying our ontology to real case studies from different domains is on our list for future work.

8 Related work

Several ontologies have been proposed for dealing with privacy and security. For example, Palmirani et al. [24] proposed PrOnto, a first draft privacy ontology for supporting researchers and regulators while analyzing privacy policies through SPARQL queries. Oltramari et al. [25] developed PrivOnto, a semantic framework for analyzing privacy policies, which rely on an ontology developed to represent privacy-related issues to users and/or legal experts. On the other hand, Kalloniatis et al. [3] introduce PriS, a security requirements engineering method that considers users' privacy requirements as business goals and provides a methodological approach for analyzing their effect on the organizational processes. Dritsas et al. [15] developed an ontology for developing a set of security patterns that can be used to deal with security requirements for e-health applications. In addition, Labda et al. [4] propose a privacy-aware Business Processes framework for modeling, reasoning and enforcing privacy constraints. In summary, most existing works do not appropriately cover all four concept categories (e.g., organizational, risk, treatment, and privacy) we consider in this work, which was clear based on the results of the systematic literature review we conduct.

9 Conclusions and Future Work

We proposed the COPri ontology for privacy requirements, and since it is based on a systematic literature review; it is more comprehensive in coverage than all ontologies included in our systematic review. Moreover, the ontology has been implemented and applied to an AAL illustrative example. In addition, we have validated it depending on CQs. Finally, we have evaluated the ontology against common pitfalls for ontologies with the help of some software tools, lexical semantics experts, and privacy and security researchers. The main purpose of developing COPri is assisting requirements engineers while eliciting privacy requirements for systems that handle personal data by providing a comprehensive set of necessary and sufficient concepts that allow for analyzing privacy requirements in their social and organizational context.

In this paper, we provide a preliminary validity check for the comprehensiveness of our proposal, which needs to be complemented in the future with empirical validation through controlled studies. The next step in this work is to develop a tool and a systematic methodology for privacy requirements that are founded on the COPri ontology. We also aim at better analyzing how the sensitivity level can be determined based on the situation, and how it can be used to facilitate the identification of privacy requirements. We will refine the analysis of the PoU property as *compatible/compatible* are too abstract to characterize such important property, and we will investigate how PoU can be determined based on the characteristics of the goal. Additionally, we are planning to develop a goal-oriented framework based on our ontology to be used for eliciting and analyzing privacy requirements.

References

1. General Data Protection Regulation: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, and repealing Directive 95/46. Official Journal of the European Union (OJ) **59** (2016) 1–88
2. Gharib, M., Salnitri, M., Paja, E., Giorgini, P., Mouratidis, H., Pavlidis, M., Ruiz, J.F., Fernandez, S., Siria, A.D.: Privacy Requirements: Findings and Lessons Learned in Developing a Privacy Platform. In: Proceedings - 24th International Requirements Engineering Conference, RE, IEEE (2016) 256–265
3. Kalloniatis, C., Kavakli, E., Gritzalis, S.: Addressing privacy requirements in system design: The PriS method. Requirements Engineering **13**(3) (2008) 241–255
4. Labda, W., Mehandjiev, N., Sampaio, P.: Modeling of privacy-aware business processes in BPMN to protect personal data. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, ACM (2014) 1399–1405
5. Gharib, M., Giorgini, P., Mylopoulos, J.: Towards an Ontology for Privacy Requirements via a Systematic Literature Review. In: International Conference on Conceptual Modeling. Volume 10650 LNCS. Springer (nov 2017) 193–208
6. Solove, D.J.: A Taxonomy of Privacy. University of Pennsylvania Law Review **154**(3) (2006) 477
7. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. Dresden University (2010) 1–98

8. Krasnova, H., Spiekermann, S., Koroleva, K., Hildebrand, T.: Online social networks: why we disclose. Journal of Information Technology **25**(2) (2010) 109–125
9. Awad, Krishnan: The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization. MIS Quarterly **30**(1) (2006) 13
10. Souag, A., Salinesi, C., Mazo, R., Comyn-Wattiau, I.: A security ontology for security requirements elicitation. In: Engineering Secure Software and Systems. Springer (2015) 157–177
11. Uschold, M.: Building Ontologies : Towards a Unified Methodology. Proceedings Expert Systems 1996, the 16th Annual Conference of the British Computer Society Specialist Group on Expert Systems (1996) 1–18
12. Fernández-López, M., Gómez-Pérez, A., Juristo, N.: METHONTOLOGY: From Ontological Art Towards Ontological Engineering. AAAI-97 Spring Symposium Series *SS-97-06* (1997) 33–40
13. Dong, H., Hussain, F.K., Chang, E.: Application of Protégé and SPARQL in the field of project knowledge management. Second International Conference on Systems and Networks Communications, ICSNC 2007 (2007)
14. Gharib, M., Giorgini, P., Mylopoulos, J.: Ontologies for Privacy Requirements Engineering: A Systematic Literature Review. preprint arXiv:1611.10097 (2016)
15. Dritsas, S., Gymnopoulos, M., Balopoulos, T., Kokolakis, S., Lambrinoudakis, C., Katsikas, S.: A knowledge-based approach to security requirements for e-health applications. Journal for E-Commerce Tools and Applications (2006) 1–24
16. Turn, R.: Classification of personal information for privacy protection purposes. (1976) 301
17. Gharib, M., Giorgini, P.: Modeling and Reasoning About Information Quality Requirements. In: Requirements Engineering: Foundation for Software Quality. Volume 9013., Springer (2015) 49–64
18. Gharib, M., Giorgini, P.: Analyzing trust requirements in socio-technical systems: A belief-based approach. In: IFIP Working Conference on The Practice of Enterprise Modeling. Volume 235., Springer (2015) 254–270
19. Mayer, N.: Model-based management of information system security risk. PhD thesis, University of Namur (2009)
20. Mouratidis, H., Giorgini, P.: Secure Tropos: A security-oriented extension of the Tropos methodology. Journal of Software Engineering and Knowledge Engineering **17**(2) (2007) 285–309
21. Gharib, M., Lollini, P., Bondavalli, A.: A conceptual model for analyzing information quality in System-of-Systems. In: 12th System of Systems Engineering Conference, SoSE 2017, IEEE 1–6
22. Gharib, M., Mylopoulos, J.: A Core Ontology for Privacy Requirements Engineering. preprint arXiv:2486619 (2018) `https://goo.gl/23ihY3`
23. Poveda-villalón, M., Suárez-figueroa, M.C., Gómez-pérez, A.: A Double Classification of Common Pitfalls in Ontologies. Development (2010) 1–12
24. Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., Robaldo, L.: PrOnto: Privacy Ontology for Legal Reasoning. Electronic Government and the Information Systems Perspective (2018) 139–152
25. Oltramari, A., Piraviperumal, D., Schaub, F., Wilson, S., Cherivirala, S., Norton, T.B., Russell, N.C., Story, P., Reidenberg, J., Sadeh, N.: PrivOnto: A semantic framework for the analysis of privacy policies. Semantic Web **9**(2) (2018) 185–203