General description
The COPri ontology aims at assisting software engineers while designing privacy-aware systems that belong to various domains by providing a generic and expressive set of key privacy concepts and relationships, which enable for capturing privacy requirements of the system-to-be in their social and organizational context.

Evaluation purpose
This evaluation aims at verifying whether the COPri ontology in terms of its concepts and relationships suffers from several common pitfalls in ontologies. The list of the pitfalls we consider in this evaluation can be found in the next page, followed by the COPri ontology.

Privacy and confidentiality statement
All responses, including any personal information will be kept strictly confidential. Your data will only be used in combination with the responses of others participating in the evaluation.

For any problem do not hesitate to contact: Mohamad Gharib mohamad.gharib@unifi.it
Expert participation: you are kindly asked to carefully evaluate the COPri ontology in terms of its concepts and relationships against several common pitfalls in ontologies. The list of the pitfalls we consider in this evaluation are listed below, and the COPri ontology can be found on the next page.

P1. Creating polysemous elements: an ontology element whose name has different meanings is included in the ontology to represent more than one conceptual idea. For example, the class “Theatre” is used to represent both the artistic discipline and the place in which a play is performed.

P2. Creating synonyms as classes: several classes whose identifiers are synonyms are created and defined as equivalent. As an example we could define “Car”, “Motorcar” and “Automobile” as equivalent classes. Another example is to define the classes “Waterfall” and “Cascade” as equivalents. This pitfall is related to the guidelines presented in [14], which explain that synonyms for the same concept do not represent different classes.

P7. Merging different concepts in the same class: a class is created whose identifier is referring to two or more different concepts. An example of this type of pitfall is to create the class “StyleAndPeriod”, or “ProductOrService”.

P17. Specializing too much a hierarchy: the hierarchy in the ontology is specialized in such a way that the final leaves cannot have instances, because they are actually instances and should have been created in this way instead of being created as classes. Authors in [14] provide guidelines for distinguishing between a class and an instance when modeling hierarchies. An example of this type of pitfall is to create the class “RatingOfRestaurants” and the classes “1fork”, “2forks”, and so on, as subclasses instead of as instances. Another example is to create the classes “Madrid”, “Barcelona”, “Sevilla”, and so on as subclasses of “Place”. This pitfall could be also named “Individuals” are not Classes.

P21. Using a miscellaneous class: to create in a hierarchy a class that contains the instances that do not belong to the sibling classes instead of classifying such instances as instances of the class in the upper level of the hierarchy. This class is normally named “Other” or “Miscellaneous”. An example of this type of pitfall is to create the class “HydrographicalResource”, and the subclasses “Stream”, “Waterfall”, etc., and also the subclass “OtherRiverElement”.

P24. Using recursive definition: an ontology element is used in its own definition. For example, it is used to create the relationship “hasFork” and to establish as its range the following the set of restaurants that have at least one value for the relationship “hasFork”.
The conceptual model of COPri

In this section, we present the conceptual model of COPri in terms of its concepts and relationships. Figure 1 shows the meta-model of COPri as a UML class diagram. The concepts of COPri are organized into four main dimensions:

**Organizational dimension**: proposes concepts to capture the social and technical components of the system in terms of their capabilities, objectives, and dependencies.

**Risk dimension**: proposes concepts to capture risks that might endanger privacy needs at the social and organizational levels.

**Treatment dimension**: proposes concepts to capture countermeasure techniques to mitigate risks to privacy needs.

**Privacy dimension**: proposes concepts to capture the stakeholders’ (actors) privacy requirements/needs concerning their personal information.

(1) **Organizational dimension**, includes concepts for capturing the organizational aspects of the system, which are further organized into several categories such as agentive, intentional and informational entities, social dependencies and social trust. In what follows, we define each of these categories in terms of their concepts and relationships.

**Agentive entities**: captures the active entities of the system, we have three concepts along with two relationships:

- **Actor** represents an autonomous entity that has intentionality and strategic goals within the system, and it covers two entities: a role and an agent:
  - **Role** represents an abstract characterization of an actor in terms of a set of behaviors and functionalities within some specialized context. A role can be a specialization (**is-a** relationship) of one another.
  - **Agent** represents an autonomous entity that has a specific manifestation in the system. An agent can **plays** a role or more within the system, where an agent inherits the properties of the roles it plays.

**Intentional entities**: captures the objectives that the actors aim to achieve. Therefore, we adopted the goal concept as well as and/or decomposition (refinement) relationships to represent such objectives.

- **A goal** is a state of affairs that an actor intends (**aims relationship**) to achieve. When a goal is too coarse to be achieved, it can be refined through **and-decompositions** of a root goal into finer sub-goals.
  - **and-decomposition** implies that the achievement of the root-goal requires the achievement of all its sub-goals.
  - **or-decomposition** is used to provide different alternatives to achieve the root goal, and it implies that the achievement of the root-goal requires the achievement of any of its sub-goals.

**Informational entities**: capture the Information related concepts and relationships:
Information represents any informational entity without intentionality. Information can be atomic or composite (composed of several parts), and we rely on partOf relationship to capture the relationship between an information entity and its sub-parts. Moreover, 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.

**Personal information** any information that can be related (directly or indirectly) to an identified or identifiable legal entity (e.g., names, addresses, medical records) \[2,35\].

Several researchers have advocated that not all personal information has the same sensitivity levels (e.g., \[35,6,21\]). Moreover, various sensitivity levels and categories for personal information have been proposed (e.g., \[5,9,34,20\]). To this end, we include sensitivity level concept that personal information has in our ontology. Based on \[34\], we adopt four different sensitivity levels that range from 1 to 4, where 4 is the most sensitive. Accordingly, we adopt four corresponding categories (we represent as classes) of personal information, namely Restricted, Confidential, Sensitive, and Secret. These categories have different privacy requirements, and therefore, can be used to facilitate the identification of privacy requirements based on the sensitivity of the personal information. The sensitivity levels and their corresponding categories of personal information along with a short description are shown in Table 1.

On the other hand, numerous works (e.g., \[26,1,27\]) have linked the sensitivity of personal information to when and where such information has been collected and for what purposes, i.e., the context/state of affairs related to such information. Thus, we adopt the concept of situation as a mean to determine the sensitivity level of personal information, where a situation can be defined as a partial state of affairs in terms of things that exist in that state, their properties, and interrelations \[16\].

**Use** is a relationship between goals and information, and it has three attributes:

<table>
<thead>
<tr>
<th>Sen. level</th>
<th>Potential effects on owner</th>
<th>Category</th>
<th>Description/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Loss of reputation, recognition, social acceptance, self-respect, loyalty, competence</td>
<td>Restricted</td>
<td>Information on political views, anti-social behavior</td>
</tr>
<tr>
<td>2</td>
<td>Loss of economic security and opportunities, employment. Disruption of family life</td>
<td>Confidential</td>
<td>Information on medical and psychiatric treatments, sexual deviations, extramarital affair, criminal history</td>
</tr>
<tr>
<td>3</td>
<td>Loss of civil rights, imprisonment, serious effects on mental and physical health</td>
<td>Sensitive</td>
<td>Self-reported information on illegal or anti-social behavior, information on medical and mental condition, psychiatric evaluations</td>
</tr>
<tr>
<td>4</td>
<td>Loss of life or physical safety</td>
<td>Secret</td>
<td>Information on the real identity of a police officer that is under-cover</td>
</tr>
</tbody>
</table>
Type of Use (ToU) our ontology provide four different types of use:

- **Produce** indicates that information is created by a goal that is responsible for its production;
- **Read** indicates that information is consumed by a goal;
- **Modify** indicates that information is modified/ altered by a goal;
- **Collect** indicates that information is acquired by a goal.

Need to Use (NtU) capture the necessary to use, and we differentiate between two main types:

- **Require** indicates that the use of information is required for the goal achievement \([12]\);
- **Optional** indicates that information is not required for the goal achievement \([12]\).

Purpose of Use (PoU) we differentiate between two main purposes of use:

- **Compliant** indicates that the purpose for which information is used is compliant with the rules that guarantee the best interest of its owner;
- **Incompliant** indicates that the purpose for which information is used does not comply with the rules that guarantee the best interest of its owner.

**Describes** is a relationship between information and goal, where information describes the goal while it is pursued by some actor. In particular, actors are able to collect information that describes activities (goals) performed by some other actors \([13]\).

**Information ownership & Permissions:** capture the relationships among personal information, the legal entities who own them, and how such categories control the use of such information by others.

- **Own** indicates that an actor is the legitimate owner of information, where information owner has full control over the use of information it owns.
- **Permission** is consent that identifies a particular use of a particular object in a system \([30]\). Information owner (data subject\(^1\)) controls the use of its own information depending on permissions over such information. In COPri, a permission has a type that is (P)roduce, (R)ead, (M)odify and (C)ollect, which cover the four relationships between goals and information that our ontology proposes.

**Entities interactions:** capture the interactions/dependencies among actors of the system concerning their objectives and entitlements. The ontology adopts three types of interactions:

- **Information provision** captures the provision of (provisionOf) information by an actor (provisionBy) to another one (provisionTo), where the source of the provision relationship is the provider and the destination is the requester. Moreover, information provision 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.

\(^1\) Information owner and data subject are synonyms in this paper
Delegation indicates that actors can delegate obligations and entitlements to one another, 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.

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 [3].

Entities social trust: the need for trust arises when actors depend on one another for goals or permissions since such dependencies might entail risk [4]. Therefore, our ontology adopts the concept of trust to capture the actors’ expectations of one another concerning their delegations. The source of trust called the trustor, the destination is called trustee, and the subject of trust is called trustum. The concept of Trust is specialized into two concepts GoalTrust, where the trustum is a goal; and PermissionTrust, where the trustum is a permission.

Trust has a type that can be either trust or distrust, where the former indicates the expectation of trustor that the trustee will behave as expected considering the trustum (e.g., a trustee will achieve the delegated goal, or it will not misuse the delegated permission); while the last indicates the expectation of trustor that the trustee will not behave as expected considering the trustum (e.g., a trustee will not achieve the delegated goal, or it will misuse the delegated permission).

Monitoring: can be defined as the process of observing and analyzing the performance of an actor in order to detect any undesirable performance [15]. We adopt the concept of monitoring to compensate the lack of trust or distrust in the trustee concerning the trustum [10,38], where the source of monitoring is called the monitor, the destination is called monitoree. 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 along with their interrelationships (e.g., threat, vulnerabilities, attack, etc.) concerning personal information. In what follows, we define each of these concepts and their interrelationships:

A vulnerability is a weakness in the system that can be exploited by a threat [29,23,31].

A threat is a potential incident that threaten personal information by exploiting a vulnerability concerning such information [24,31,19]. A threat can be either natural (e.g. disaster), accidental (e.g. hardware or software failure), or intentional (e.g. theft of personal information) [36,33]. COPri differentiates between two types of threat:

Incidental threat: a casual, natural or accidental threat that is not caused by a threat actor nor require an attack method. Incidental threat has a probability
that measures the likelihood of its occurs, and it is is characterized by three different values high, medium or low.

**Intentional threat** a threat that require a threat actor and includes a presumed attack method [22][23].

**Threat actor** is an actor that intends to achieve an intentional threat [29][24][27].

**Attack method** is a standard means by which a threat actor carries out an intentional threat [24][8][33].

**Impact** is the consequence of the threat over (impactOver) the personal information. An impact has a Severity (hasSeverity) that captures the level of the impact [37][33], which is characterized by high, medium or low.

(3) **Treatment dimension**, includes countermeasure concepts to mitigate risks. COPri proposes a high abstraction level concepts to capture the required protection/treatment level (e.g., privacy goal), which can be refined into concrete protection/treatment constraints (e.g., mechanisms or policies) that can be implemented. The concepts of the treatment dimension are:

A **privacy goal** defines an aim to counter threats and prevents harm to personal information by satisfying privacy criteria concerning such information.

A **privacy constraint** is defined as 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** is a privacy statement that defines the permitted and/or forbidden actions to be carried out by actors of the system toward information.

A **privacy mechanism** is a concrete technique to be implemented for helping towards the satisfaction of privacy goal. Some mechanisms can be applied to personal information directly (e.g., anonymity, unlinkability).

(4) **Privacy dimension**, introduce concepts to capture the actors’ privacy requirements/needs concerning their personal information. The concepts of the privacy dimension are:

**Privacy requirement** that is used to capture information owners’ privacy needs at a high abstraction level concerning their personal information. Privacy requirements are interpretedBy privacy goals. Moreover, privacy requirement is further specialized into seven more refined concepts:

**Confidentiality** means personal information should be kept secure from any potential leaks and improper access [32][6][21]. We rely on the following principles to analyze confidentiality:

**Non-disclosure** personal information can only be disclosed if the owner’s consent is provided, i.e., the disclosure of the personal information should be under the control of its legitimate owner [32][6][21]. Note that non-disclosure also covers information provision (e.g., confidential information provision). Non-disclosure can be analyzed depending on both read permission as well as confidential information provision.
Need to Know (NtK) an actor should only use information if it is strictly necessary for completing a certain task [21]. NtK can be analyzed depending on Need to Use (NtU), i.e., if the type of NtU is not required (i.e., optional) a violation can be raised.

Purpose of Use (PoU) personal information should only be used for specific, explicit, legitimate purposes and not further used in a way that is incompatible with those purposes [35,32,6]. PoU can be analyzed depending on the type of PoU, if it is incompliant a violation can be raised.

Anonymity the identity of information owner should not be disclosed unless it is required [6,32,17,28], i.e., the primary/secondary identifiers of the data subject (e.g., name, social security number, address, etc.) should be removed if they are not required and information still can be used for the same purpose after their removal. Personal information can be anonymized depending on some privacy mechanism.

Unlinkability means that it should not be possible to link personal information should back to its owner. In other words, any identifiers that allow such linkage should be removed, and information can be used without linking it back to its owner [17,25,18,28]. A privacy mechanism can be used to remove any linkage between personal information and its owner.

Unobservability the identity of information owner should not be observed by others, especially third parties, while performing an activity (e.g., use a resource or service) [17,18,28]. Unlike Anonymity and Unlinkability that try to hide the identity of information owner, Unobservability aims to hide some activities that are performed by the information owner [28]. We analyze unobservability relying on the describes relationship, which enables us to detect situations where an actor tries to collect personal information that describes an activity (goal) while it is performed by some actor.

Notice information owner should be notified when its information is being collected [35,32,6]. Notice is considered mainly to address situations where personal information related to a legitimate entity is being collected without her knowledge. Notice can be analyzed depending on the collect relationship and its corresponding permission. In case, personal information is being collected and there is no permission to collect, a notice violation will be raised. Providing a permission to collect means that the actor has been notified and agrees his personal information to be collected.

Transparency information owner should be able to know who is using his/her information and for what purposes [35,32,6].

We rely on two principles to analyze transparency:

Authentication a mechanism aims at verifying whether actors are who they claim they are. We analyze authentication by verifying whether 1- the actor is playing a role that enables to identify its main responsibilities; and 2- the actor is not playing any threat actor role. If both of these rules did not hold, a violation can be raised.

Authorization a mechanism aims at verifying whether actors can use information in accordance with their credentials [6]. We analyze authorization
by verifying whether the actor has the required permissions to perform a task at hand.

**Accountability** information owner should have a mechanism available to them to hold information users accountable for their actions concerning information \cite{6,19}. We rely on the following principle to analyze accountability:

**Non-repudiation** the delegatee cannot repudiate he/she accepted the delegation \cite{19}. We analyze *non-repudiation* by relying on the adoption concept, if there exists a delegatee without an adopt relationship to the delegatum, a *non-repudiation* violation can be raised.
**Expert’s comments**

Please list the pitfalls you have identified, described by their identifier (e.g., P1., P2., P7., P17., P21. and P24.), element(s) affected (e.g., a class or a relationship), followed by a description of the identified pitfall:

<table>
<thead>
<tr>
<th>ID</th>
<th>Element(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


34. Rein Turn. Classification of personal information for privacy protection purposes. page 301, 1976.


