Modeling and Reasoning About Privacy-Consent Requirements

Marco Robol, Elda Paja, Mattia Salnitri, Paolo Giorgini

▶ **To cite this version:**

Marco Robol, Elda Paja, Mattia Salnitri, Paolo Giorgini. Modeling and Reasoning About Privacy-Consent Requirements. 11th IFIP Working Conference on The Practice of Enterprise Modeling (PoEM), Oct 2018, Vienna, Austria. pp.238-254, 10.1007/978-3-030-02302-7_15 . hal-02156452

HAL Id: hal-02156452

https://inria.hal.science/hal-02156452

Submitted on 14 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Modeling and reasoning about privacy-consent requirements

Marco Robol[1], Elda Paja[1], Mattia Salnitri[2], and Paolo Giorgini[1]

[1] University of Trento, Trento, Italy
[2] Politecnico di Milano, Milan, Italy

Abstract. Since the origin of the web, up to social networks, and now to the internet of things, the quantity of personal information produced and shared is uncontrollably increasing. Privacy regulations protect our right to have the control on our personal data. According to the recent General Data Protection Regulation (GDPR), entered into force in May 2018, infringements can be very costly to organizations, ranging from 10s to 100s of thousands of Euros. In order to ensure compliance with such regulations, privacy should be taken into consideration as early as at requirements time, so to avoid expensive after-the-fact fixes. Modeling frameworks have been proposed to support the analysis of requirements in complex socio-technical systems, however, even if a primary role is given to security, for privacy more work need to be done. In this paper, starting from the social concept of consent, we propose a modeling language and define the formal framework for the analysis of privacy-consent requirements. We report on our experience in the analysis of privacy in the medical domain, in the context of a research project with the Trentino health-care provider (APSS).

Keywords: privacy, regulations, consent, socio-technical systems, requirements, modeling, automated reasoning

1 Introduction

The European General Data Protection Regulation (GDPR) [12] has entered into force on May 2018. Compliance is of utmost importance for organizations, in order to avoid monetary penalties which can be up to 20 million Euros. Moreover, public debates on privacy, such as the recent one about personal data being sold away by a well known social network [8], have a strong impact on people, with consequences for organizations that can be even worst than actual fines.

The way how companies do their business is shifting from a traditional closed one, to an approach more open to collaborations with external parties. This way of doing business is supported by *consent*, a key element in privacy regulations that allows the processing and sharing of personal data among organizations, yet it gives users control over their own data. Personal data are stored and shared among organizations by humans through information system. In such complex socio-technical systems, compliance with privacy regulations should be considered starting from social components, down to technical ones.

Compliance with privacy regulations should be handled as soon as possible, considering it as an early requirement, so to avoid unexpected costs of re-engineering [22].

Most importantly, given the relevance of humans in organization activities, the analysis of privacy requirements must carefully represent the social context, the actors involved, their interactions, as well as their expectation and responsibilities in light of privacy regulations.

Methods for the analysis of privacy requirements has been often discussed [18, 3, 6], however, most of these works either does not put social aspects first, or does not take into consideration regulations. Nómos [29, 30] propose a solution for regulatory compliance of software specifications, while other works analyze requirements under a social perspective, such as, *i** [36] or Tropos [7], and focus on security as Secure Tropos [14] or STS [11]. In previous work [27], we have presented preliminary results of a framework for privacy requirements, which includes a modeling language based on STS and automated reasoning capabilities.

In this paper, we propose a method, based on STS, for the analysis of privacy and *consent* requirements, to support compliance with regulations. It is based on (i) the modeling of the domain, (ii) the specification of privacy and *consent* requirements, and (iii) automated reasoning to support compliance with regulations.

The contributions of the paper are as follows:

- a modeling language, goal-oriented, focused on privacy and consent;
- A reasoning framework to automate the detection of conflicts between system operations and consent provided by users;
- A validation of the modeling language and the reasoning framework in collaboration with privacy experts in the medical domain, including legal experts, experts in the organization processes, and technical people.

The paper is structured as follows. Section 2 presents the problem of privacy and consent requirements, followed by a motivating case study in Section 3 and the baseline in Section 4. Section 5 introduces the modeling language for privacy and consent requirements. Section 6 discusses the formal framework for automated reasoning. Section 7 presents the results of the validation, with experts from APSS. Section 8 discusses related work, and Section 9 concludes.

2 From regulations to privacy and consent requirements

Regulations are composed of a set of principles that impact on interactions between organizations and users. Compliance with regulations is not straightforward and should be reached throughout a careful analysis of the regulations with respect to the organization.

The European Union (EU) has developed a new privacy law, the General Data Protection Regulation (GDPR)[12], to improve privacy safeguard of all European citizens. The GDPR is based on general principles which include privacy-by-default, transparency, data minimization, storage limitations, accuracy, and integrity. **Privacy-by-default** prevent the collection, processing, or use of personal data if it is not the case that the user has previously agreed on such operation. **Transparency** imposes organizations to inform the user on the performed processing operations. **Data minimization** is on the minimal set of personal data that are necessary for the provision of a service.

Storage limitation impose constraints on the storage of the data, such as, time constraints or the right to be forgotten. **Accuracy** and **integrity** require for reliable and non corrupted data and adequate security measures.

Consent is a key element adopted by the majority of privacy regulations, including EU GDPR [12] and US HIPAA [2], to put the user in control of his own personal data. In relation with consent, the GDPR provides a set of principles, such as, purpose limitation, free, informed and explicit consent. For the **purpose limitation** principle, consent must have a well-defined purpose, clearly stated in the privacy notice, no general consent is allowed. **Free consent** is the freedom of deciding whether to consent or not on personal data processing, without being forced by the organization. **Informed consent** imposes organizations to provide users with a privacy notice with clear and understandable details on the processing. **Explicit consent** requires companies to demonstrate a legally compliant acquisition of consent from each user. For example, this can be enforced by asking the user to sign a paper version of the consent.

3 Motivating case study: Trentino health-care provider

For what concerns privacy regulations, the medical domain is one of the most critical, because of the big quantity of highly confidential clinical data involved. Here, a trade-off between privacy and accessibility of data is fundamental. If on one hand there is the need for privacy of patients, on the other, availability of data can be of vital interest. For this reason, the traditional management of user consent, paper-based and not integrated in the organization processes, is not a viable solution. This makes the analysis of requirements related to consent not straightforward, also considering that the health-care system is an evolving complex socio-technical system, where requirements identified at social and organizational level impact on operational processes, such as, accountability of the transmission of medical reports between doctors, impacts on the processes and on technical components.

We report on the analysis of privacy that we have conducted on the Trentino health-care provider of the province of Trento (Italy), the APSS (Azienda Provinciale Servizi Sanitari). The APSS, not only directly provides health care services, but also collaborates with external organizations so to integrate them in the national health-care system. We focus on the newborn Italian national register of citizens' medical data (FSE), which is going to become operative in the near future, and for which, the APPS is working toward the implementation for what concern the province of Trento. With the FSE, public and private Italian medical service providers will be all interconnected, giving the possibility to doctors to access patient medical and administrative data from everywhere.

4 Baseline

The work presented in this paper is based on STS [11], a security requirement engineering method to support the design of secure systems. STS focuses on social aspects as the main causes of security problems in complex systems. It includes (i) modeling languages to represent the system both at a socio-technical level and at the business process (procedural) level, (ii) an automated reasoning framework, (iii) a supporting tool.

STS-ml is the formally defined goal-based modeling language provided by the STS method. It is used to model socio-technical systems as a composition of intentional actors, which represent either a single instance (agent) or a class (role) of either technical components or humans. Such overall representation of the system allows to focus on actor interactions, i.e. goal delegations and document transmissions. The language is multi-view, so to capture and focus on different aspects of the same system separately.

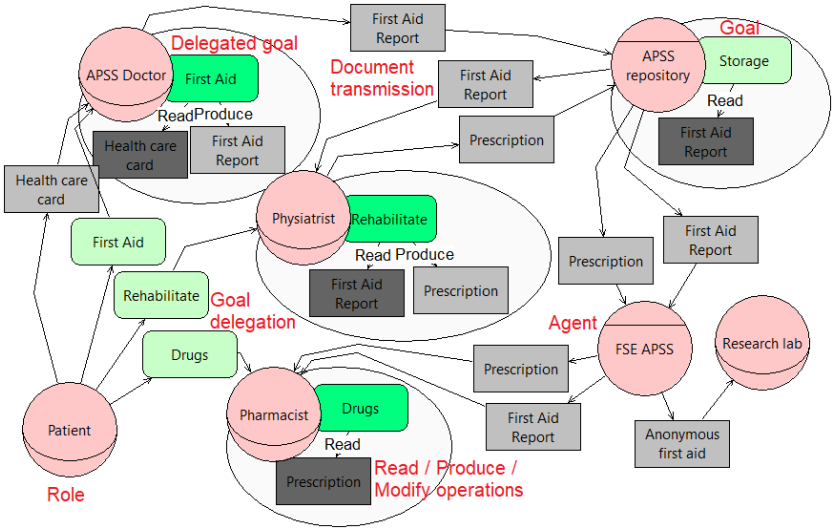

Fig. 1. APSS social view

Figure 1 shows an excerpt of the **social view**, representing the dependencies between actors, from a model of the case study [3], where a patient interacts with a doctor of the APSS, who provides the first aid, collects his personal data, produces a report, and uploads it to the FSE system, then the patient interacts with a physiatrist for the rehabilitation, who reads the first aid report and produces a prescription, and with a pharmacist, who reads the prescription to provide drugs. A research lab obtains an anonymous version of the first aid report.

Figure 2 shows an excerpt of the **information view**, representing the structure of documents and their informative content, from the model of the case study. Patient first aid data are made tangible by the first aid report document possessed by the APSS doctor. Prescription data are made tangible in the document possessed by the physiatrist. Last information is the health care identifier of the patient, which is made tangible in the health care card.

Figure 3 shows an excerpt of the **authorization view**, representing the operations permitted by owners of information. In this example, *Pharmacist* is authorized by the *Physiatrist* to read the *Prescription data* in the context of *providing drugs*, while he is not authorized to modify or produce these.

[3] The complete model can be found at `disi.unitn.it/~marco.robol/`

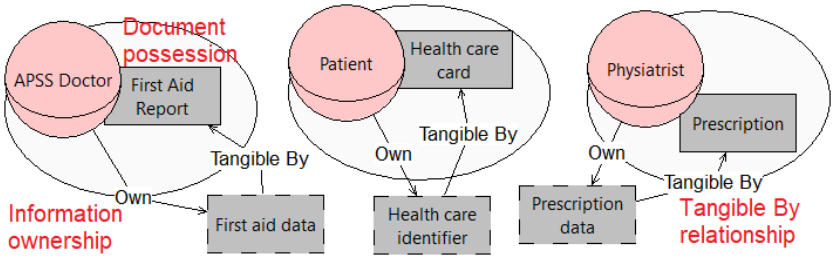

Fig. 2. APSS information view

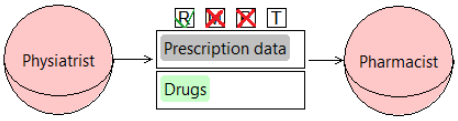

Fig. 3. APSS authorization view

Privacy in socio-technical systems has been discussed in [27], where a method based on STS has been proposed to reason on data protection requirements and normative aspects. The work describes first results of the development of a method to support the design of complex systems with the principle of privacy-by-design. Here we compliment the method with a support for consent, wich is crucial to comply with GDPR.

5 Modeling consent requirements

Graphical models can be used to ease the analysis of requirements both in the design of a new system or in the re-engineering of an existing one. In the analysis of privacy and consent requirements, it is important to model the system so to represent (i) personal data, (ii) data operations, and (iii) privacy consent. We propose a modeling language, based on STS-ml, for the analysis of privacy and consent requirements in complex and evolving socio-technical systems. The new modeling language includes the concepts of (i) *personal data*, based on linkability, (ii) *privacy operations*, based on a taxonomy that includes collection, processing, and disclosure, and (iii) *consent*, based on authorizations. In the rest of the section we go into more details on these aspects, for each of them we present how they are supported by the modeling language.

Our definition of **personal data** is aligned with the one provided in the GDPR[12]. Here an excerpt from Article 4(1): "personal data means any information relating to an ... identifiable natural person (‘data subject’); ... one who can be identified, directly or indirectly ..." [12]. The idea behind this definition is in the identifiability of the person in the information. Spiekermann and Cranor, in [33], relate the identifiability of users to **linkability** of data, they talk of personal data in case of linkability and anonymous data in the case of unlinkability. Information is not always linkable by its own, but could became such depending on how it is made tangible. Therefore, we define personal data as an information that is made tangible in a form so that it can be linked to an

identifiable user. We are aware that linkability is actually a very discussed and controversial topic. Deciding and demonstrating the linkability or unlinkability of data is not straightforward and several studies have been done on this, starting from k-anonimity [34] to l-diversity [23] and t-closeness [21]. The process of de-identification of personal information is critical and if not approached correctly, could lead to unwanted and malicious data breaches [13]. We suggest that the lack of linkability, is a property that must be carefully verified. The eventuality that an information could be linkable to the user should be always taken into consideration. In our modeling language, we introduce the *Linkable To* relationship to represent the potential linkability to a *Data Subject* actor, of a *Document*, our tangible form of *Information*.

We investigate on **data operations** that are relevant for privacy, and we propose a classification based on our interpretation of the privacy taxonomy presented in [31]. The taxonomy, in addition to the concept of personal data and data subject [27], is based on the concept of data holder, who is the performer of the following operations: (i) *information collection*, related with the means by which information is gathered from the user by the data holder, (ii) *information processing*, related with the consolidation and use of information and its transfer between information systems by the data holder, (iii) *information dissemination*, related with the disclosure to the public or to another person, (iv) *invasion*, related with intrusion in the private life of the user and interference with his decisions. In our modeling language we include three privacy-relevant operations, namely collection, processing, and dissemination, while we not included invasion since it does not necessary involve information and it is therefore not relevant to information analysis. We speak of *Collection* of personal data in the case of transmission of a document from the data subject, the actor to which the document can be linked to, to another actor. *Processing* of personal data is in the case of any of the operations of reading, modification or production, and also in the case of transmission of documents between actors that are part of the data holder. While *Dissemination* is any transmission made by the data holder, to any other actor that differs from the data holder, and the data subject.

To support **consent** requirements analysis we adopt the definition of consent given in Article 4(11) of the GDPR: "Consent of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". In our interpretation consent in an agreement between two actors, the data subject and the data holder, that consists in the permission for operating on personal data for a specific purpose. For the consent on the processing of personal data, defined in the GDPR, we propose a classification based on [32], that includes consent on the collection, processing (use), and dissemination. In our language we support the modeling of *Consent* as a social relationship between two actors, the *Data Subject* and the *Data Holder*, in terms of authorizations for operating over personal data. The set of actors authorized within a consent defines the *Scope* of the consent, with respect to which we speak of *consent to the collection* in case of authorizations for the transmission of personal data from the data subject to any actor part of the consent scope, *consent to the processing* in case of authorizations to read, modify, produce, or transmit, personal data between actors in the scope, and *consent*

to the dissemination in case of authorizations for the transmission of personal data to actors outside consent scope.

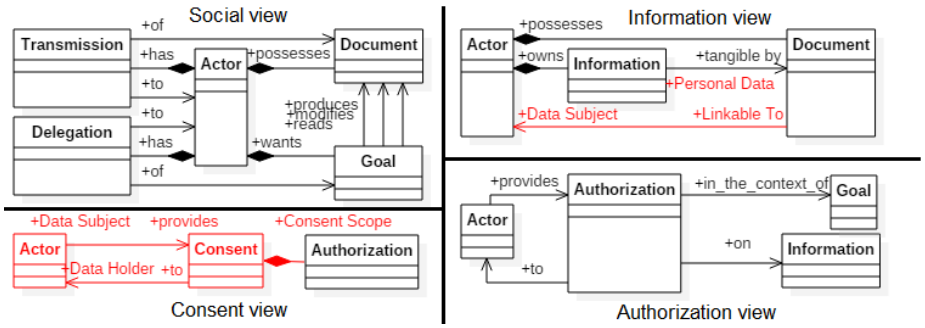

Fig. 4. Meta-model of the proposed modeling language

Figure 4 shows the **meta-model** of the modeling language, in red the elements related to privacy and consent. The meta-model is organized in four diagrams, each representing one of the views proposed in the modeling language. Concepts of the language shared between views are here represented separately in each of the respective meta-model diagrams. The replication of such concepts between the views can be automated by the supporting tool, so to help the modelers in creating consistent diagrams. The modeling language splits across a total of four different views. With respect to STS-ml we added a fourth one, the consent view, to model consent and analyze its requirements. Consent is represented as a relationship between a data subject and a data holder, and consists in a set of authorizations. Permissions specified in this view differs from the ones in the authorization view because: (i) they includes the operations of collection and dissemination and (ii) such permissions are related to a specific consent.

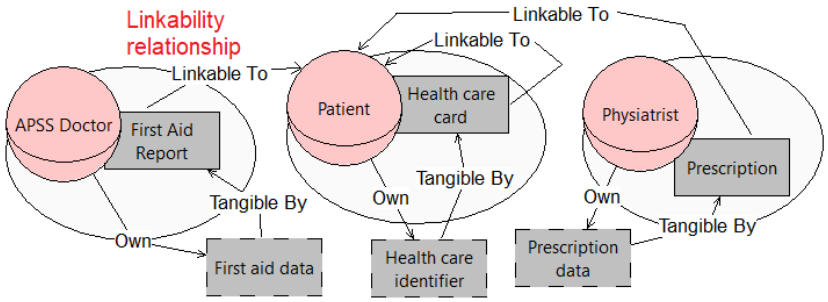

Fig. 5. APSS information view modified

Figure 5 shows an excerpt of the **information view** from the model of the case study, modified with respect to its STS version presented in Figure 2. Patient first aid report is represented as a document, possessed by the actor APSS doctor, such document is *linkable to* the patient himself, meaning that the patient is identifiable. Similarly, the

prescription document, that is possessed by the physiatrist, is also linkable to the patient, and the same is for the health care card, also linkable to the patient.

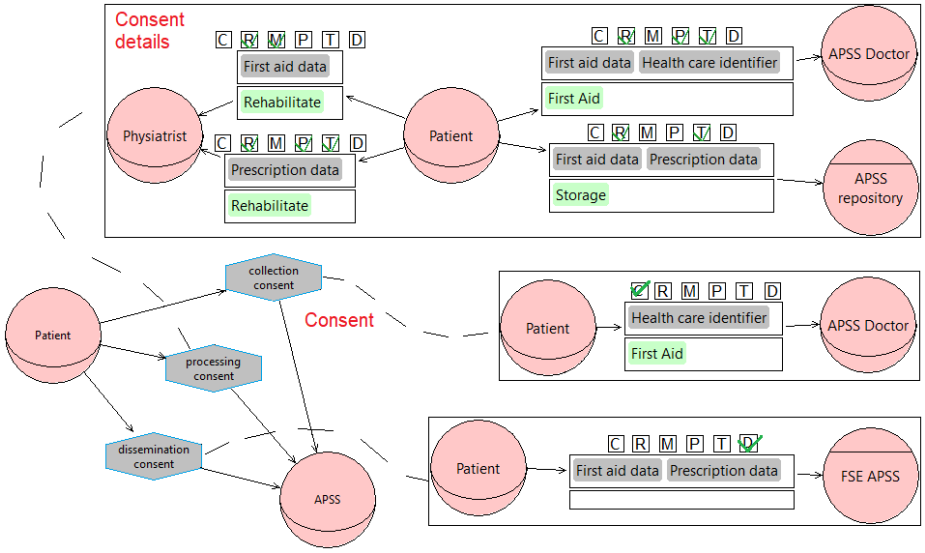

Fig. 6. APSS consent view

Figure 6 shows an excerpt of the consent view from the model of the case study, that includes the details of the three consents provided by the patient to APSS. On the top, details of the processing consent provided to the APSS, which includes in its scope physiatrist, APSS doctor and APSS repository. Letters C, R, M, P, T, and D stands respectively for Collect, Read, Modify, Produce, Transmit, and Disseminate. The APSS doctor is authorized to read, produce and transmit information of the patient within the scope of the consent. Physiatrist is authorized to read and modify first aid data, and read produce and transmit prescription data. On the right, details of the collection consent provided to the APSS, where the APSS doctor is authorized to collect information from the patient. On the bottom, details of the dissemination consent provided to the APSS, which consists in authorizing the FSE in disseminating first aid data and prescription data of the patient to actors outside the scope of this consent.

6 Formal framework and automated reasoning

Modeling languages should be simple enough to easily identify inconsistencies but, as the models start growing, to adequately represent real world cases, this could became harder [35]. Automated reasoning, based on formal languages, can support users in the identification of potential inconsistencies in the models. The formalization we propose relies on [11, 27], for further supporting consent requirements for privacy.

6.1 Formalizing the modeling language

This section provides a formalization of the language based on the formalization of STS-ml provided in [11]. We use set theory and we define atomic variables with strings in typewriter with a leading capital letter (e.g., G, I); sets are defined with strings in the calligraphic font for mathematical expressions (e.g., $\mathcal{G}$, $\mathcal{I}$); relationship are defined in italics style with a leading non-capital letter (e.g.,*wants*, *possesses*). Table 1 lists the predicates used to represent concepts and relationships. We focused on the formalization of concepts related to privacy, such as, consent and personal data. See the information relationship of *linkableTo*(D, A) and the social relationship of *consents*$(\mathsf{A}, \mathsf{A}', \mathsf{S})$ and authorization.

Table 1. Predicates

Concepts: *actor*(A), *agent*(Ag), *role*(R), *goal*(G), *document*(D), *information*(I).
Intentional relationships ($\mathcal{IRL}$): *wants*(A, G), *possesses*(A, D), *decomposes*$(\mathsf{A}, \mathsf{G}, \mathcal{S}, \mathsf{DecT})$, where $\mathsf{DecT} \in \{and, or\}$, *reads/modifies/produces*$(\mathsf{A}, \mathsf{G}, \mathsf{D}, \mathsf{OpT})$, where $\mathsf{OpT} \in \{R, M, P\}$.
Social relationships ($\mathcal{SRL}$): *plays*$(\mathsf{Ag}, \mathsf{R})$, *delegates*$(\mathsf{A}, \mathsf{A}', \mathsf{G})$, *transmits*$(\mathsf{A}, \mathsf{A}', \mathsf{D})$, *authorizes*$(\mathsf{A}, \mathsf{A}', \mathcal{I}, \mathcal{G}, \mathcal{OP}, \mathsf{Tr})$, *consents*$(\mathsf{A}, \mathsf{A}', \mathcal{A}'', \mathcal{AUTH})$, where $\mathcal{OP} = (\{C, R, M, P, T, D\} \cup \{\overline{C}, \overline{R}, \overline{M}, \overline{P}, \overline{T}, \overline{D}\})$ and $\mathsf{Tr} \in \{true, false\}$.
Information relationships ($\mathcal{I}_{\mathcal{RL}}$): *owns*$(\mathsf{A}, \mathsf{I})$, *partOfI*$(\mathsf{I}_1, \mathsf{I}_2)$, *partOfD*$(\mathsf{D}_1, \mathsf{D}_2)$, *tangibleBy*$(\mathsf{I}, \mathsf{D})$, *linkableTo*$(\mathsf{D}, \mathsf{A})$.

Def. 1 (Intentional actor). An intentional actor is any agent or role that commits himself in the achievement of a set of goals. An intentional actor model AM is a tuple $\langle \mathsf{A}, \mathcal{G}, \mathcal{D}, \mathcal{IRL} \rangle$, where A is an actor, $\mathcal{G}$ is a set of goals, $\mathcal{D}$ is a set of documents, and $\mathcal{IRL}$ is a set of intentional relationships. An actor model $\mathsf{AM} = \langle \mathsf{A}, \mathcal{G}, \mathcal{D}, \mathcal{IRL} \rangle$ is well-formed if all intentional relationships are defined over actor A, goals in $\mathcal{G}$, and documents in $\mathcal{D}$.

For example, considering the actors in Figure 1, the *APSS Doctor* commit himself in the achievement of the goal *First aid provided to the patient*, goal delegated to him by the *Patient*. The model of actor *APSS Doctor* is composed by the goal *First Aid* and the documents *Health Care Card* and *First Aid Report* and the intentional relationships *Read* and *Produce* on the documents.

Def. 2 (Consent). *consents*$(\mathsf{A}, \mathsf{A}', \mathcal{A}'', \mathcal{AUTH})$ is a social relationship defined between a data subject A, a data holder A', and a set of actors representing the consent scope A''. Consent consists in a set of authorizations $\mathcal{AUTH}$ provided to actors in the consent scope. A *consents*$(\mathsf{A}, \mathsf{A}', \mathcal{A}'', \mathcal{AUTH})$ is well-formed only authorizations in the consent are provided only to actors in the consent scope, represented by the set of actors $\mathcal{A}''$.

For example, considering the Figure 6, the *patient*, as data subject, provides the *consent for the collection* of his personal data to the *APSS actor*, the data holder. Such

consent consists in permitting the *APSS doctor* to collect from the *patient* the information *health care identifier* in the context of achieving the goal *first aid*.

Def. 3 (Social model). We bind together actors models and social relationships to compose a social model of the system. A social model SM is a tuple $\langle \mathcal{AM}, \mathcal{SRL}, \mathcal{I}_{\mathcal{RL}} \rangle$ where $\mathcal{AM}$ is a set of intentional actor models, $\mathcal{SRL}$ is a set of social relationships, and $\mathcal{I}_{\mathcal{RL}}$ is a set of information relationships.

Def. 4 (Authorization closure). We define a closure over authorizations so that if no explicit authorization is provided, any operation on any information is implicitly forbidden. Let SM be a well-formed social model, the authorization closure over $\mathcal{SRL}$ in SM, denoted as $\triangle_{\mathcal{SRL}}$, is a super-set of $\mathcal{SRL}$ that makes prohibitions explicit, when no authorization is granted by any actor.

Def. 5 (Consent closure). We define a closure over consents so that if no consent provide an explicit permission to operate on linkable documents, then operating on linkable documents is implicitly forbidden.

6.2 Reasoning about privacy and consent

In this section, we present our contribution in the automated reasoning, proposed to support analysis related to privacy and consent.

First example of automated reasoning is related to violated authorizations.

Def. 6 (Violated authorization). An authorization is violated when, even if it makes prohibition to an actor A to operate on an information I, the actor A actually operates on a document D that makes tangible the information I, also considering the operating context G. Formally, for each provided authorization *authorizes*(A, A', $\mathcal{I}$, $\mathcal{G}$, $\mathcal{OP}$, TrAuth), a violation is detected if exists an operation *reads/modifies/produces*(A', G, D, OpT) s.t. $\mathsf{G} \in \mathcal{G}$, OpT is negated in $\mathcal{OP}$, and exists a *tangibleBy*(I, D) s.t. $\mathsf{I} \in \mathcal{I}$.

In the example of Figure 1, considering only the authorizations specified in Figure 3, different authorizations are violated. The *APSS Doctor* can not read the *Health care card* of the *Patient*, the Physiatrist can not read the *First air report* of the *APSS Doctor*, and also the the *FSE repository* agent and the *FSE APSS* can not read the *First air report*, and they can not transmit any documents.

Consent requirement specifies the need of assessing user decision on the collection, processing, and disclosure of his personal data. We can automatically detect violation of consent reasoning on operations performed on document, and linkability of documents.

Def. 7 (Violated collection consent). Violation of collection consent is automatically detectable in the case of transmission of a document D, from actor A owner of some information I tangible in D, if it is the case that the document is linkable to A.

Def. 8 (Violated processing consent). Violation of processing consent is automatically detectable in the case of any processing operation on a document D, executed by an actor A', including the transmission toward an actor A'', whether it is the case that the document is linkable to another actor A, owner of some information I tangible in D, and both actors A' and A'' are part of the consent scope.

Def. 9 (Violated dissemination consent). Violation of dissemination consent is automatically detectable in the case of transmission of a document D, from an actor A' to an actor A'', if it is the case that the document is linkable to another actor A, owner of some information I tangible in D, and that the transmission is not authorized within the scope of any consent.

In the example of Figure 1, considering the consent view in Figure 6, some consent are violated. For example, the transmissions of *First Aid data* and *Prescription data*, from the *APSS repository* to the *FSE APSS*, raise a violation of the dissemination consent, because dissemination consent is provided only to the agent *FSE APSS*.

We provide automated reasoning for the minimization of personal data, based on the analysis of information used in the achievement of goals and consent. Excessive permissions are provided by consent in the case no such operations are performed.

Def. 10 (Excessive consent). A consent is excessive when the provided actor A, in the context of achieving any of the goals G in the authorization, does not perform any of the allowed operations on any document D that makes tangible any of the information in the authorization. Formally, an authorization *authorizes*(A, A', $\mathcal{I}$, $\mathcal{G}$, $\mathcal{OP}$, TrAuth) is excessive if given $\mathsf{I} \in \mathcal{I}$, $\mathsf{G} \in \mathcal{G}$, $\mathsf{OpT} \in \mathcal{OP}$, do not exists any *reads/modifies/produces*(A', G, D, OpT), s.t *tangibleBy*(I, D).

For example, in the *processing consent*, represented in Figure 6, *APSS doctor* is authorized to transmit the *Health care identifier*, while this is not necessary in the achievement of any goal, as in Figure 1.

Core logic implementation In STS [11], automated reasoning has been implemented in DLV [1], and integrated in the graphical modeling editor tool. Different types of automated analysis are provided with the tool, such as, well-formedness and security analysis. We present here an excerpt of the core logic implementation, while the integration in the modeling tool is still under development.

Table 2. DLV rules for consent

(i) *violatedCollectionConsent*(A, I, D) :-
not *canCollect*(A, I, G), *transmits*(A', A, D), *tangibleBy*(I, D), *linkableTo*(D, A').
(ii) *violatedProcessingConsent*(A, I, G, D) :-
not *canProcess*(A, I, G), *reads/mod./prod./tx.*(A, G, D), *tangibleBy*(I, D), *linkableTo*(D, A').
(iii) *violatedDisseminationConsent*(A, I, D) :-
not *canDisseminate*(A, I, G), *transmits*(A, A'', D), *tangibleBy*(I, D), *linkableTo*(D, A').
(iv) *excessiveConsent*(A, I, G, OP) :-
(*canCollect*(A, I, G), not *transmits*(A', A, D))
$\vee$ (*canProcess*(A, I, G), not *reads/mod./prod./tx.*(A, D))
$\vee$ (*canDisseminate*(A, I, G), not *transmits*(A, A'', D)),
tangibleBy(I, D), *linkableTo*(D, A').

In Table 2 the rules implementing in DLV the following reasoning: (i), (ii), and (iii) identification of violation of consent for the collection, processing, and disclosure, (iv) excessive consent.

7 Evaluation

This section discusses the results of evaluating the method with domain practitioners. Research questions that we wanted to address are about the usability and completeness of the modeling language and the utility of the reasoning framework.

The evaluation has been done in a real case study provided by the Trentino healthcare provider (APSS), in the context of a research project consisting in the experimentation of the STS method for the certification of processes with the GDPR. The experiment design consists in an iterative process on the activities of refinement and validation of the modeling language and the reasoning framework. This required several interactions with domain experts. People involved had different backgrounds, different perspectives on privacy, and different levels of expertise in modeling languages. They included legal, business, and technical people from the APSS, such as, privacy and legal experts, organization experts, experts in the processes, and APSS system experts.

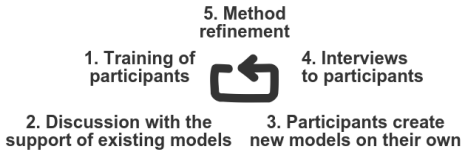

Fig. 7. User-Centred Evaluation process

Figure 7 shows the steps of each iteration. We first provided a quick introduction on the modeling language, then we discussed with participants with the support of models provided by us, then we let them use the language on their own to produce new models, finally we collected their opinions with respect to the research questions.

The first research question is related to the usability of the language, and how much it can be understand by non experts. Second research question is related to the completeness of the modeling language, and the missing concepts in representing the system and constraints imposed by regulations. Third research question is related to the utility of the automated reasoning framework in identifying privacy criticalities and problems. In the following, we discuss the results of the evaluation.

Usability Opinion of the participants was positive with respect to the usability of the modeling language. They were all able to understand the models, that have been successfully used to support the discussions. Some of the participants were also able to modify existing models and produce new ones. The continuous interactions provided us with new ideas on how to refine and improve the language, on the basis of feedback, comments, and suggestions. For example, the consent relationship between actors, and its detailed view in terms of authorizations, was initially spread among the social and authorization view. After some interaction with experiment participants, we have been able to improved the graphical aspects of the language, by introducing the consent view.

Completeness In the evaluation of completeness of the language, we focused on its ability to represent privacy and consent. In the fist iterations, opinion of the participants

was a lack of the language in representing consent. It was not clear if consent was a goal, a document, or an authorization, even if last one seems to be the more similar concept. There was a lack in the language in covering the concepts of collection, processing, and dissemination. We modified the language so to support consent, defined as an agreement between two actors consisting in a set of authorizations, and we also introduces the operations of collection, processing, and dissemination.

Utility of the reasoning framework We evaluate the utility of the reasoning framework in supporting humans in the analysis of privacy and consent. Models produced in the experimentation are composed by many organizations, departments, and technical systems, with a lot of dependencies, a complex structure of information and documents, and a not straightforward specification of security requirements, such as authorizations. It turned out that models were so complex to be as nearly as impossible for non expert users, and still difficult for experts like us, to analyze and reason on them by hands. For this reason, we improved some automated reasoning, in particular for the identification of violated consent on collection, processing, and dissemination. Opinion of the participants was positive with respect to the utility of the newly integrated automated reasoning on consent.

8 Related Work

Several description languages for authorizations and access control rules are available in the literature [5, 4, 24, 26], most of which are based on allow/deny rules, These can be used to protect personal data, as a good privacy practice, however, alone, they does not provide compliance with privacy regulations. A different use of authorization language, specific for controlling privacy in the web, is proposed in the P3P platform [10, 9], where users, surfing the web, are allowed to express their privacy preferences.

Frameworks for the analysis of privacy and security requirements are available in the literature. Qingfeng et al. in [18] presents a privacy goal-driven requirements modeling framework to support the design of a Role-Based Access Control (RBAC) system. Kalloniatis et al. in [3] present PriS, a requirement engineering method for security and privacy, for the analysis of the impact of privacy requirements on organizational processes, with the use of privacy-process patterns.

The use of automated reasoning to support analysis in requirements frameworks, is first proposed by Van Lamsweerde et al. in [35]. Giorgini et al. in [15], introduce automated reasoning in the Tropos method [7] to support the identification of conflicts. Giorgini et al. in [14, 25] present Secure Tropos, a security requirement engineering framework that extends Tropos. Breaux et al. in [6] propose a privacy requirements specification language, called Eddy, with automated reasoning feature. The language provides a set of privacy requirements, on which automated reasoning allows for the detection of conflicts between requirements. Nómos, a software requirement framework, is proposed by Siena et al. in [29, 30] and in its revisited versions [28] and [20], to tackle the problem of regulatory compliance of software. It includes a tool-supported modeling language that can detect conflicts between requirements. Privacy is not directly supported, but the framework can also be applied to privacy regulations.

Different designing frameworks, specific for privacy, have been proposed in the literature. Guarda et al. in [16] present an overview of legal aspects of privacy, which are considered of primary importance, in a technological interpretation. Spiekermann et al. in [33] propose an introduction to the privacy domain for engineers, by proposing two privacy design approaches: (i) privacy-by-policy, based on fair information practices; and (ii) privacy-by-architecture, based on data minimization. Gurses et al. in [17], provide an overview of privacy-by-design practices. The work focuses on data minimization and its importance in privacy-by-design. Hoepman et al. in [19] review mains PETs and patterns and propose privacy design strategies to integrate privacy-by-design in the software development life cycle.

9 Conclusion

We have proposed a modeling language and a reasoning framework for the analysis of consent and privacy requirements. Challenges were in the interpretation of the regulation, for example in the formalization of the concepts of personal data and consent, and in the definition of a language to support the analysis of compliance. We proposed a goal-oriented modeling language, a reasoning framework and a first evaluation based on a real case study in the medical domain. Future work includes (i) a detailed analysis and formalization of the requirements of privacy and consent and their integration in the modeling language and in the reasoning framework, (ii) the development of a supporting tool for the modeling and the automation of the analysis and (iii) the inclusion in the modeling language of other concepts related to privacy.

References

1. DLVSYSTEM S.r.l. | DLV, http://www.dlvsystem.com/dlv/
2. The Health Insurance Portability and Accountability Act (HIPAA). Washington, D.C.: U.S. Dept. of Labor, Employee Benefits Security Administration (2004)
3. Addressing privacy requirements in system design: the PriS method. Requirements Engineering 13.3, 241–255 (2008)
4. Ashley, P., Hada, S., Karjoth, G., Powers, C., Schunter, M.: Enterprise privacy authorization language (epal). IBM Research (2003)
5. Ashley, P., Hada, S., Karjoth, G., Schunter, M.: E-p3p privacy policies and privacy authorization. In: Workshop on Privacy in the Electronic Society. pp. 103–109. ACM (2002)
6. Breaux, T.D., Hibshi, H., Rao, A.: Eddy, a formal language for specifying and analyzing data flow specifications for conflicting privacy requirements
7. Bresciani, P., Giorgini, P., Giunchiglia, F., Mylopoulos, J., Perini, A.: Tropos: an Agent-Oriented Software Development Methodology. JAAMAS 8(3), 203–236 (2004)
8. Cadwalladr, C., Graham-Harrison, E.: Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian 17 (2018)
9. Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., Reagle, J.: The platform for privacy preferences 1.0 (p3p1. 0) specification. W3C recommendation 16 (2002)
10. Cranor, L.F.: Platform for Privacy Preferences (P3P). In: Encyclopedia of Cryptography and Security, pp. 940–941. Springer US, Boston, MA (2011)
11. Dalpiaz, F., Paja, E., Giorgini, P.: Security requirements engineering: designing secure socio-technical systems. MIT Press (2016)

12. 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 L119/59 (May 2016)
13. Garfinkel, S.L.: De-Identification of Personal Information. Tech. rep. (2015)
14. Giorgini, P., Massacci, F., Mylopoulos, J., Zannone, N.: Modeling security requirements through ownership, permission and delegation. In: Proceedings of 13th IEEE International Conference on Requirements Engineering. pp. 167–176. IEEE (2005)
15. Giorgini, P., Mylopoulos, J., Sebastiani, R.: Goal-oriented requirements analysis and reasoning in the tropos methodology. Engineering Applications of AI 18(2), 159–171 (2005)
16. Guarda, P., Zannone, N.: Towards the development of privacy-aware systems. Information and Software Technology 51.2, 337–350 (2009)
17. Gürses, S., Troncoso, C., Diaz, C.: Engineering privacy by design (2011)
18. He, Q., Antón, A.I., et al.: A framework for modeling privacy requirements in role engineering. In: Proc. of REFSQ. vol. 3, pp. 137–146 (2003)
19. Hoepman, J.H.: Privacy design strategies. In: IFIP International Information Security Conference. pp. 446–459. Springer (2014)
20. Ingolfo, S., Siena, A., Mylopoulos, J.: Goals and Compliance in Nomos 3. International Conference on Conceptual Modeling pp. 275–288 (2014)
21. Li, N., Li, T., Venkatasubramanian, S.: t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In: IEEE 23rd International Conference on Data Engineering. pp. 106–115 (2007)
22. Liu, L., Yu, E., Mylopoulos, J.: Security and privacy requirements analysis within a social setting. In: 11th International Requirements Engineering Conf. pp. 151–161. IEEE (2003)
23. Machanavajjhala, A., Kifer, D., Gehrke, J.: l-Diversity: Privacy Beyond k-Anonymity. ACM Trans. Knowl. Discov. Data 1(52) (2007)
24. Moses, T., et al.: Extensible access control markup language (xacml) version 2.0. Oasis Standard 200502 (2005)
25. Mouratidis, H., Giorgini, P.: Secure Tropos: A Security-Oriented Extension of the Tropos methodology. Int. J. Soft. Eng. Knowl. Eng. 17(2), 285–309 (2007)
26. Park, J., Sandhu, R.: The ucon abc usage control model. ACM Transactions on Information and System Security (TISSEC) 7(1), 128–174 (2004)
27. Robol, M., Salnitri, M., Giorgini, P.: Toward GDPR-Compliant Socio-Technical Systems: modeling language and reasoning framework. Leuven (2017)
28. Siena, A., Jureta, I., Ingolfo, S., Susi, A., Perini, A., Mylopoulos, J.: Capturing variability of law with nómos 2. ER 7532, 383–396 (2012)
29. Siena, A., Mylopoulos, J., Perini, A., Susi, A.: Designing Law-Compliant Software Requirements. pp. 472–486. Springer, Berlin, Heidelberg (2009)
30. Siena, A., Susi, A.: Engineering Law-Compliant Requirements - The Nomos Framework. Ph.D. thesis, University Of Trento (2010)
31. Solove, D.J.: A Taxonomy of Privacy (2005)
32. Solove, D.J.: Introduction: Privacy self-management and the consent dilemma. Harv. L. Rev. 126(7), 1880 (2012)
33. Spiekermann, S., Cranor, L.: Engineering Privacy. IEEE Transactions on Software Engineering 35(1), 67–82 (jan 2009)
34. Sweeney, L.: k-Anonymity: a Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5), 557–570 (2002)
35. Van Lamsweerde, A., Darimont, R., Letier, E.: Managing conflicts in goal-driven requirements engineering. IEEE transactions on Software Engineering 24(11), 908–926 (1998)
36. Yu, E.: Modelling strategic relationships for process reengineering. Social Modeling for Requirements Engineering 11, 2011 (2011)