Personal and Ubiquitous Computing Journal manuscript No.
(will be inserted by the editor)

User Consent Modeling for Ensuring Transparency and Compliance in Smart Cities

Javier D. Fernández · Marta Sabou · Sabrina Kirrane · Elmar Kiesling · Fajar J. Ekaputra · Amr Azzam · Rigo Wenning

Received: date / Accepted: date

Abstract Smart city infrastructures such as transportation and energy networks are evolving into so-called *Cyber-Physical Social Systems* (CPSSs), which collect and leverage citizens' data in order to adapt services to citizens' needs. The privacy implications of such systems are, however, significant and need to be addressed. Current systems either try to escape the privacy challenge via anonymization or use very rigid, hard coded work flows that has been agreed with a data protection authority. In the case of the latter, there is a severe impact on data quality and richness, whereas in the former, only these hard coded flows are permitted resulting in diminished functionality and potential. We address these limitations via *user modeling* in terms of investigating how to model and semantically represent user consent, preferences and data usage policies that will guide the processing of said data in the data lake. Data protection is a horizontal field and consequently very wide. Therefore we focus on a concrete setting where we extend the domain-agnostic SPECIAL policy language for a smart mobility use case supplied by Vienna's largest utility provider. To that end (1) we create an extension of SPECIAL in terms of a core CPSS vocabulary that lowers the semantic gap between the domain agnostic terms of SPECIAL and the vocabulary of the use case; (2) we propose a workflow that supports defining domain specific vocabularies for complex

Funded by Austrian Research Promotion Agency (FFG): grant no. 861213 (CitySPIN), and the European Unions Horizon 2020 research and innovation programme grant 731601 (SPECIAL).

Javier D. Fernández, Sabrina Kirrane and Amr Azzam
Vienna University of Economics and Business, Vienna, Austria
E-mail: firstName.secondName@wu.ac.at

Marta Sabou, Elmar Kiesling and Fajar J. Ekaputra
Technical University of Vienna, Vienna, Austria
E-mail: firstName.secondName@ifs.tuwien.ac.at

Rigo Wenning
ERCIM/W3C, Sophia-Antipolis, France
E-mail: rigo@w3.org

CPSS systems; and (3) show that these two contributions allow successfully achieving the goals of our setting.

Keywords Cyber Physical (Social) Systems · Smart Mobility · User Consent Modeling · Privacy · GDPR · Linked Data

1 Introduction

Large-scale Smart City infrastructures such as smart transportation or smart energy networks typically span the boundaries of the physical, cyber and social spheres. Sensors in the physical world are used to collect (real-time) data, which is then processed pragmatically by a cyber-component to determine appropriate actuation/adaptation strategies. Increasingly, participants and users of these infrastructures provide data to the system (e.g., through social sensing) and can even act as actuators to optimize the system. Such complex systems, are referred to as *Cyber-Physical Social Systems* (CPSSs) [43].

Considering that CPSSs often make use of and integrate personal information from various sources, privacy protection needs to be at the core of such systems. The reality is often different with systems adopting a take-it-or-leave-it approach [16]. At a first glance, giving users the choice to accept or reject a request to participation in a CPSS seems to be an efficient and simple solution. But such an approach regularly attracts harsh criticism. Once trust is sufficiently eroded via this simple approach, people will reject all those systems by default. It is therefore important to earn and sustain the trust into CPSSs. This philosophy is also underlying the framework that the EU has laid down with the creation of the GDPR.

The GDPR defines a set of obligations for controllers and processors of personal data, including, but not limited to, having a lawful reason for processing personal data and providing full transparency to data subjects with respect to the processing of their personal data. Several tools [22,28,31] that assist companies in assessing their compliance with the GDPR have recently been developed. They are, however, targeted at self-assessment (i.e., companies complete standard questionnaires in the form of privacy impact assessments). The self assessment is used to check a pre-set and fixed work flow against the legal rules. The challenge for CPSSs is to allow for a maximum amount of data to be collected with a maximum of use and re-use permissions granted by data subjects. Confronted with the resulting large amount of legalese that comes with such an approach, privacy scholars start to talk about the end of data self determination[19]. Despite the nominal transparency, data subjects and regulators alike are just overwhelmed by the complexity of CPSSs.

The EU H2020 SPECIAL[1] project, strives to enable companies to work with the data subjects to sustain trust in complex systems, such as CPSSs. With the SPECIAL system, preferences, consent, and legal grounds for processing can be integrated at run time into a CPSS. Because of the semantics involved, the

[1] `https://www.specialprivacy.eu/`

CPSS becomes privacy-aware. Algorithms in the CPSS can react on privacy concerns within the system with a high degree of flexibility. This serves the data controllers and the data subjects alike. Data self-determination allows data subjects to participate in value creation via those CPSSs. The legal or technical term for such participation is consent. If other grounds for processing are used, the quest for trust imposes a high level of digestible transparency, here again semantics plays a critical role.

In this context, the data controller is challenged to make sure that personal data processing actually conforms to the promises made to the data subject. This is of increased importance since the fines for misbehaviour have become significant with the advent of the GDPR. If the data subject sets a preference in the CPSS, e.g. via his mobile device, the CPSS needs to make sure said preference are followed in the subsequent complex work flows that sometimes even transcend organizational borders. The idea here is to enable the system to automatically ingest and interpret the preferences without having a programmer setting switches. The challenge is then to manage the data flows. The SPECIAL approach addresses this challenge by attaching semantics to personal data that specifies possible usages, in the form of a *usage policy*. The SPECIAL engine is capable of using those semantics in order to perform both ex-ante and ex-post compliance checking, and to provide digestible transparency to data subjects concerning what happened to their personal data, why and when.

In this paper we analyze the suitability of the SPECIAL policy language for CPSS user consent modeling and showcase its extension to cover the needs of the *Smart Mobility* use cases provided by Vienna's largest utility provider, Wiener Stadtwerke (WStW). To that end, we adopt a three stage approach. First, we create an extension of SPECIAL with a vocabulary that is generically applicable to CPSS use cases (the SPECIAL-CPSS core vocabulary) and introduces a set of CPSS-specific terms, thus lowering the effort of extending the domain-agnostic policy language to the need of concrete use cases. The core vocabulary is grounded on an overview of CPSS systems obtained with a literature review, and as such aims to be reusable across various CPSS domains, beyond smart mobility. Second, we propose a practical work flow to support CPSS owners in analyzing their complex systems and deriving user consent modeling vocabularies needed for their use cases. Third, we check the usefulness of these two contributions by using them in the context of the WStW's smart mobility use case and successfully deriving use case specific usage policies. In a nutshell, the novelty of this paper lies in a non-trivial extension of the semantics used within Cyber-Physical-Systems in order to prepare a much more sophisticated approach to data protection and GDPR compliance for CPSS. Concretely, we make the following contributions:

– the SPECIAL-CPSS core vocabulary, which serves as a means for describing usage constraints across a variety of CPSS domains and are usable within a SPECIAL – like system;

- the practical work flow, which enables CPSS preference, constraint and consent modeling in general;
- a practical use case to show how these techniques can be applied in a CPSS setting.

The remainder of this paper is organized as follows. In Section 2 we present the state of the art in policy languages and GDPR transparency and compliance. Then, in Section 3 we describe the main components of the SPECIAL consent, transparency and compliance framework, paying particular attention to the SPECIAL usage policy language and vocabularies, and the methodology used to extend SPECIAL in order to cater for CPSSs. The proposed extension is motivated and guided by our CitySPIN use cases, presented in Section 4. Our core CPSS vocabulary is introduced in Section 5. Section 6 presents a workflow to establish data subjects' consent and data usage policies for specific use cases. This workflow is validated using our CitySPIN use cases in Section 7. Finally, we conclude and present future work in Section 8.

2 Related Work

The European General Data Protection Regulation (GDPR) requires data controllers to obtain explicit consent for the processing of personal data from data subjects. Traditionally, this consent is obtained via a human-readable description (i.e., a *contract*, or *terms and conditions*), which does not allow for any automatic processing. Thus, formal policy languages are designed to unambiguously represent usage policies, which makes it possible to automatically verify whether data processing is covered by data subjects' given consent. In the following, we first review current alternative policy languages (Section 2.1) and GDPR transparency and compliance tools (Section 2.2).

2.1 Uage Policies

There are several potential candidates for the formal representation of usage policies, including semantic policy languages [42,23,7,25] and standard based policy languages [13,21]. KAoS [42] is a general policy language which adopts a pure ontological approach, whereas Rei [23] and Protune [7] use ontologies to represent concepts, the relationships between these concepts and the evidence needed to prove their truth, and rules to represent policies. Kolovski et al. [25] demonstrate how together description logic and defeasible logic rules can be used to understand the effect and the consequence of sets of access control policies. They share with our view the set of reasoning tasks over policies, and use description logics. On the other hand, they don't address complexity issues. The Platform for Privacy Preferences (P3P)[2] is a W3C recommendation that enables websites to express their privacy preferences in a

[2] P3P,http://www.w3.org/TR/P3P/

machine readable format. A more recent W3C recommendation known as the Open Digital Rights Language (ODRL)[3], released in early 2018, is a general rights language that can be used to define rights or to limit access to digital resources. In principle, any of these languages could be used to encode usage policies in a CPSS scenario. Still, there are other relevant considerations that suggest to define a usage policy language around the recent standard OWL2, and select language constructs carefully in order to adequately trade off expressiveness and computational complexity. This is the main objective of the SPECIAL policy language [5][8], developed within the EU H2020 SPECIAL project. In the next section, we provide an analysis of the policy language and its adaptation to CPSS needs.

2.2 Transparency and Compliance

Since the GDPR has come into effect, data controllers must provide transparency to data subjects with respect to the processing of personal data and *compliance*, i.e. the CPSS data controller must demonstrate that the usage of personal data complies with data subjects' consent (it respects/does not violate any requests).

Transparency. As for transparency about data processing, relevant work primarily focuses on the re-purposing of existing logging mechanisms as the basis for personal data processing transparency and compliance [6]. Many of the existing approaches use a secret key signing scheme based on Message Authentication Codes (MACs) together with a hashing algorithm to generate chains of log records that are in turn used to ensure log confidentiality and integrity [2] (cf. [6] for a summary of existing approaches). MACs use symmetric keys that are generated and verified using collision-resistant secure cryptographic hash functions. Only a few contributions [34,37], however, focused on personal data processing. An alternative distributed architecture to manage access to personal data based on blockchain technology has been proposed by Zyskind et al. [46]. The authors discuss how the blockchain data model and Application Programming Interfaces (APIs) can be extended to keep track of both data and access transactions. More recently, Sutton and Samavi [41] propose an extension of blockchain technology with *Linked Data* to create tamper-proof audit logs and non-repudiation. Very little research has been conducted, however, into transparency requirements and performance/scalability issues of such blockchain-based solutions.

Compliance. As for GDPR compliance, recently the Information Commissioner's Office (ICO) in the UK [22], Microsoft [28], and Nymity [31] have developed compliance tools that enable companies to assess the compliance

[3] ODRL,`https://www.w3.org/TR/odrl-model/`

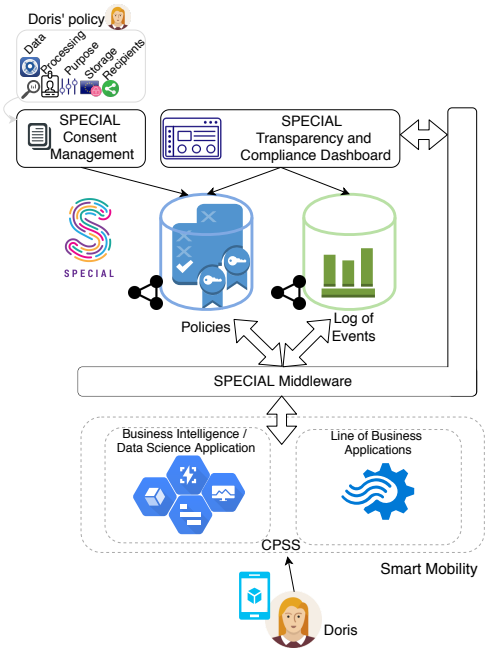

Fig. 1 The SPECIAL Consent, Transparency and Compliance framework

of their applications and business processes by completing a predefined questionnaire. Recent works also look at the challenges of representing GDPR concepts and obligations [32] as well as informed consent [17]. The management of events for business process compliance monitoring and process mining [27] can be seen as orthogonal work.

In contrast to existing approaches, in this paper we focus on vocabularies that can be used to record both usage policies and data processing and sharing events in a manner that supports automatic compliance checking.

3 Background and Methodology

In this section, we first present a high level overview of the SPECIAL consent, transparency and compliance framework (Section 3.1). Following on from this, we describe the SPECIAL policy language in detail (Section 3.2). Finally, we propose a methodology that can be used to adapt and extend the SPECIAL policy language and vocabularies to cope with CPSSs requirements in general, and, our smart mobility use case in particular (Section 3.3).

3.1 SPECIAL Consent, Transparency and Compliance Framework

In order to enable companies to comply with consent and transparency requirements stipulated by the GDPR, SPECIAL provides a policy language,

vocabularies and a consent, transparency and compliance framework, which can be adapted and extended specifically for CPSS needs. The SPECIAL framework (shown in Figure 1) consists of the following components:

(i) the *SPECIAL Consent Management* component, which is responsible for obtaining consent from the data subject and representing it in the form of a machine readable usage policy;
(ii) the *SPECIAL Transparency and Compliance Component*, which is responsible for presenting data processing and sharing events in an easily digestible manner and demonstrating that existing data processing and sharing complies with the respective usage control policies; and
(iii) the *SPECIAL Middleware*, which includes sub-components that connect the SPECIAL primary components with the access control mechanisms and business logic of existing Line of Business applications, and middleware that enables companies to perform policy aware business intelligence and data science.

Underpinning the framework are a variety of existing data sources that support business operations (i.e., *Line of Business Applications*), and strategic decision making (i.e. *Business Intelligence / Data Science Applications*), and two additional SPECIAL data sources that are needed to support SPECIAL's consent, transparency and compliance framework: a *Policies* store, which is used to record the consent, regulatory and business policies; and an *Events* store, which is used to record (i.e. log) data processing or sharing events.

3.2 SPECIAL's Usage Policy Language

In this section, we provide a detailed overview of the SPECIAL usage policy language vocabularies, which we will analyze and extend in a practical CPSS scenario in subsequent sections.

SPECIAL usage policies are encoded in OWL 2 [30]. In the examples[4] that follow, the `spl` prefix represents `http://www.specialprivacy.eu/langs/usage-policy#`. Additional details, including the full policy expression grammar in Backus normal form (BNF), can be found in the SPECIAL documentation [5].

3.2.1 Data Usage Policy Model

Conceptually, a *usage policy* is meant to specify a *set of authorized operations*. According to the GDPR, these policies shall specify clearly which data are collected, what is the purpose of the collection, what processing will be performed, where the data is stored, and whether or not the data will be shared with others. The SPECIAL policy language, which was developed in close collaboration with legal experts, consists of five core elements, collectively known as the *minimum core model* (MCM), which is depicted in Figure 2:

[4] For the policy language examples we use the OWL functional syntax which is less verbose.

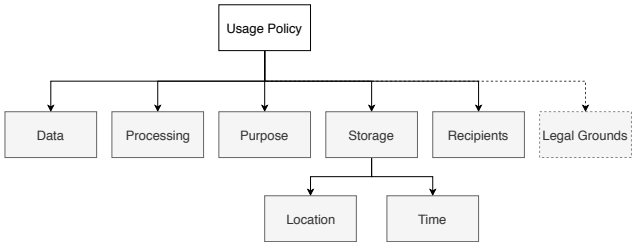

Fig. 2 The SPECIAL policy minimum core model (MCM), extended with optional legal grounds.

- *Data* describes the personal data collected from the data subject (e.g. contact information, financial data, etc).
- *Processing* describes the operations that are performed on the personal data (e.g. collection, analysis, etc).
- *Purpose* specifies the objective that is associated with data processing (e.g. health, marketing, etc).
- *Storage* specifies where data are stored and for how long.
- *Recipients* specifies who is going to receive the results of data processing and, as a special case, whom data are shared with.

Optionally, policies can be extended with zero or more legal ground(s) for processing. In this paper, we focus on consent, but other alternatives (such as *legitimate interest*) can be represented [4].

3.2.2 Encoding SPECIAL Usage Policies

A *basic usage policy* is composed of one or more policies, each of which is an OWL 2 expression of the form presented in Listing 1.

Listing 1 A basic usage policy

```
ObjectIntersectionOf(
    ObjectSomeValueFrom( spl:hasData SomeDataCategory )
    ObjectSomeValueFrom( spl:hasProcessing SomeProcessing )
    ObjectSomeValueFrom( spl:hasPurpose SomePurpose )
    ObjectSomeValueFrom( spl:hasStorage SomeStorage )
    ObjectSomeValueFrom( spl:hasRecipient SomeRecipient )
)
```

The policy presented in Listing 1, which follows the minimum core model (MCM), authorizes all operations on data that: (i) belong to *SomeDataCategory*, (ii) fall within the specified *SomeProcessing* category, (iii) have any purpose covered by the *SomePurpose* category, (iv) store the results of the processing in any place belonging to the *SomeStorage* category, and (v) disclose the results to any member(s) of the *SomeRecipient* category.

Additionally, SPECIAL provides several auxiliary vocabularies that provide a set of classes for *SomeDataCategory*, *SomeProcessing*, *SomePurpose*, *SomeRe-*

Table 1 SPECIAL auxiliary vocabularies for usage policies.

Category	Namespace	#Classes	Examples	Superclass
Data	svd:=(S)/data	27	svd:Activity, svd:Anonymized, svd:Financial, svd:Health, svd:Location, svd:Navigation, svd:Preference, svd:Profile, etc.	spl:AnyData
Processing	svpr:=(S)/processing	9	svpr:Aggregate, svpr:Analyze, svpr:Anonymize, svpr:Collect, svpr:Copy, svpr:Derive, svpr:Move, svpr:Query, svpr:Transfer	spl:AnyProcessing
Purpose	svpu:=(S)/purposes	31	svpu:Account, svpu:Arts, svpu:Delivery, svpu:Education, svpu:Feedback,svpu:Gaming, svpu:Health, svpu:Marketing, svpu:Payment, svpu:Search, etc.	spl:AnyPurpose
Recipient	svr:=(S)/recipients	6	svr:Delivery, svr:OtherRecipient, svr:Ours, svr:Public, svr:Same, svr:Unrelated	spl:AnyRecipient
Storage location	svl:=(S)/locations	7	svl:ControllerServer, svl:EU, svl:EULike, svl:ThirdCountries, svl:OurServers, svl:ProcessorServers, svl:ThirdParty	spl:AnyLocation
Storage duration	svdu:=(S)/duration	4	svdu:BusinessPractices, svdu:Indefinitely, svdu:LegalRequirement, svdu:StatedPurpose	spl:AnyDuration

cipient. Table 1 provides a high-level overview of the proposed vocabularies[5] that were defined in the context of the SPECIAL use cases. For instance, the policy in Listing 2 presents an example of a union of *basic usage policies*, in the context of an online fundraising website. The policy states that financial data can only be used for payment purposes and shall neither be stored nor disclosed to third parties, while the nickname can be used freely.

Finally, note that the `hasStorage` policy attribute is a structured object itself, with two attributes, and is specified in Listing 3, where *SomeLocation* and *SomeDuration* are expressed in terms of the corresponding storage location and duration auxiliary vocabularies.

Considering that it is clearly not possible to enumerate over all possible classes the policy language and by extension the vocabularies were designed to be extensible. This paper builds upon this extensibility to provide support for CPSS scenarios.

[5] All namespaces share the `S` which represents `http://www.specialprivacy.eu/vocabs`.

Listing 2 A policy composed of a union of basic usage policies

```
ObjectUnionOf(
   ObjectIntersectionOf(
      ObjectSomeValueFrom( spl:hasData svd:Financial )
      ObjectSomeValueFrom( spl:hasProcessing spl:AnyProcessing )
      ObjectSomeValueFrom( spl:hasPurpose svpu:Payment )
      ObjectSomeValueFrom( spl:hasStorage spl:Null)
      ObjectSomeValueFrom( spl:hasRecipient spl:Null ) )
   ObjectIntersectionOf(
      ObjectSomeValueFrom( spl:hasData svd:nickname )
      ObjectSomeValueFrom( spl:hasProcessing spl:AnyProcessing )
      ObjectSomeValueFrom( spl:hasPurpose spl:AnyPurpose)
      ObjectSomeValueFrom( spl:hasStorage spl:AnyStorage)
      ObjectSomeValueFrom( spl:hasRecipient spl:AnyRecipient ) )
)
```

Listing 3 Typical structure of the hasStorage policy attribute

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( spl:hasLocation SomeLocation )
   ObjectSomeValueFrom( spl:hasDuration SomeDuration )
   DataSomeValuesFrom( spl:durationInDays Interval )
)
```

3.3 Methodology

Figure 3 depicts the inputs, main steps and outputs of the methodology we adopted when extending the SPECIAL usage policy language vocabularies, in order to cater for a smart mobility use case, and more generally, making the first steps towards its use for the broader family of Cyber-Physical Social Systems.

The **inputs** to our work are the SPECIAL usage policy language presented in the previous section and the smart mobility use case, which will be discussed in detail in Section 4. The primary **output** is a vocabulary (i.e., a set of terms) that can be used in the current use case to specify usage policies.

SPECIAL's minimum core model (MCM, see Fig. 2) is highly generic (i.e., domain agnostic) and therefore offers little support in deriving vocabularies that are necessary to support specific use cases. To fill this gap, we propose a domain-agnostic approach that can be used to facilitate the creation of use case specific vocabularies, by first deriving a specialization of the SPECIAL MCM that captures terminology classes generically valid across a given domain. The objective being to derive a *core* ontology [38] for CPSS usage policies that are applicable and reusable across multiple CPSS subdomains. The proposed approach conforms with ontology engineering best practices, which suggest the development of layered ontology extensions from highly domain independent ontologies (e.g., generic ontologies), to core ontologies (e.g., domain ontologies) and then increasingly specific subdomain and task ontologies.

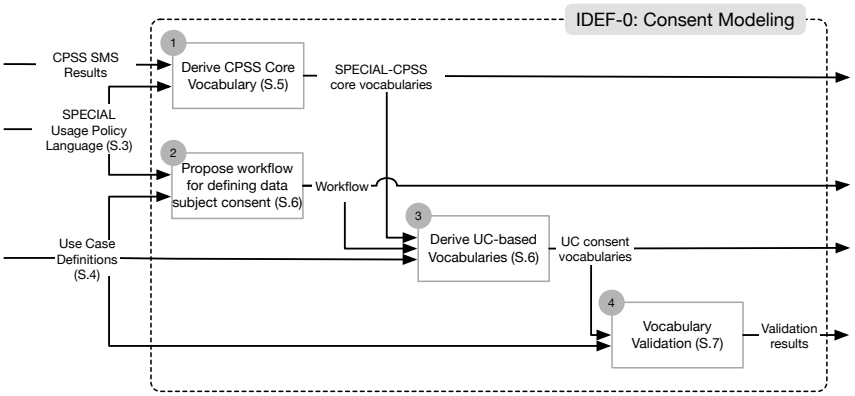

Fig. 3 Methodology for extending the SPECIAL usage policy language vocabularies for a smart mobility use case.

The proposed approach is composed of four steps, which can be summarized as follows:

1. **Step 1: Derive the CPSS core ontology.** In order to bridge the semantic gap between the SPECIAL vocabularies and domain-specific terms required to support our CPSS use cases, we first need to identify domain-specific terms. This core ontology serves as a starting point for extending SPECIAL not just to support smart mobility systems, but also to support other CPSS systems, such as smart manufacturing, smart grids or smart homes. In order to derive this generic CPSS ontology, we rely on a principled approach grounded in a *Systematic Mapping Study*. Specifically, we followed the methodology of Kitchenham et al [24] as we detail in Section 5. The output of this step is a `SPECIAL-CPSS core ontology`.
2. **Step 2: Propose workflow to define CPSS data subjects consent and data usage policies.** Before deriving specific usage policy vocabularies, it is first necessary to identify components, relationships and data sources based on use case descriptions. The concrete steps taken to that end are captured in a practical workflow to define CPSS data subjects consent and data usage policies, as described in Section 6.
3. **Step 3: Derive use case specific vocabularies** by applying the workflow for deriving domain specific vocabularies (at Step 2) based on a detailed use case description. This step results in a set of domain-specific vocabularies for usage policy specification (simply referred to as `UC consent vocabularies`). In this step, we rely on the output of Step 1, i.e., the `SPECIAL-CPSS core ontology`, and exemplify its usage to derive a domain ontology.
4. **Step 4: Vocabulary validation.** We validate both the `SPECIAL-CPSS core ontology` and the `UC consent vocabularies` by exemplifying their use to create usage policies required to support our smart mobility use cases. This

step results in a set of `validation results` w.r.t. concrete usage policies. We show a practical example in Section 7.

To sum up, the major, reusable outputs of our methodology are:

- The *SPECIAL-CPSS core ontology*, which can also serve as a starting point for describing usage policies vocabularies in other CPSS domains (output of Step1).
- The *workflow to define CPSS data subjects consent and data usage policies* (output of Step2);
- The *overall methodology* itself, can be followed whenever adapting SPECIAL to new domains. It provides guidance with respect to creating both core extensions of SPECIAL, as well as domain specific vocabularies.

4 CitySPIN Smart Mobility Use Case

In order to exemplify CPSS transparency and compliance requirements in the context of the GDPR, we present a general overview of a *Smart Mobility* use case that emerged in the CitySPIN project (Section 4.1) and subsequently describe four specific use case scenarios (Section 4.2).

4.1 General overview

As Viennas largest utility provider, Wiener Stadtwerke (WStW) manages a broad and diverse public transportation network. In their long-term planning activities, WStW aims to extend and optimize this network. In the shorter/medium term, the network needs to be adjusted temporarily, e.g., to cater for the transportation needs of large-scale events or to accommodate special situations such as refurbishing and temporary closure of transportation network stations. In this context, information about passenger flows, i.e., movement patterns generally or during (recurring) large-scale events, are a key factor in decision-making processes. Passenger access to the transportation network is currently not monitored digitally (e.g., through access gates). Such information could in principle be obtained from individual citizens, but this requires solid transparency mechanisms and means to ensure compliance, as described next. Note that the following descriptions are based to the CitySPIN project context and that they are not currently put in practice in the company's production environment.

4.2 Scenarios

In the following, we present four scenarios of the CitySPIN use case, exemplified with a generic WienMobil APP user, Doris, and a WStW transportation network planner, Eva.

Doris installs the WienMobil APP, provided by a subsidiary of WStW, which allows her to obtain real-time public transport routing information in Vienna. For a desired destination, the APP provides the best route from the current (or a specified) location by combining several modes of transportation (metro, bus, train, rent-bicycles etc) within Vienna.

During installation, Doris is guided through a number of privacy choices that determine the later behavior of the App. Those choices can later be changed in the settings. Depending on those choices, different policies will apply.

Scenario 1: A personalized mobility planning. The APP states that, in order to provide a more personalized routing service, the APP can record the history of her routing requests, including the GPS location at the moment of the search. This can be integrated with external non-personal data sources (e.g. traffic congestion and city events) and will be used to analyse her mobility patterns, including the attendance to events in the city (e.g. concerts, sport events), in order to recommend her best routes and notify about delays in the future. Additionally, the APP informs her that the data will be stored on the company servers in Austria for a period of 2 years after each collection point, in order to detect yearly recurrent events. The collected data can always be retrieved, amended or deleted via the privacy dashboard.

Doris accepts this option and starts using the APP. As she is a fan of one local soccer club, she makes intensive use of the APP to go to the soccer stadium. As soon as the end of the match is approaching, the WienMobil APP notifies Doris and shows the fastest route (avoiding any congestion) to her house or a restaurant she frequents regularly after matches.

As many people are using the APP, the service can alert Doris to wait and have a coffee in the surrounding until the congestion after the match is over. The APP issues an alarm sound once the conditions are good again.

Scenario 2: Event Partnership. At a certain point in time, WStW establishes a partnership with governmental organizations to promote non-profit cultural and heritage events. Thus, the APP asks for a potential policy update. The APP states that she can also receive partnership recommendations related to her mobility patterns for non-profit cultural and heritage events. In this case, she needs to consent to use the same collected information (history of routes, mobility patterns and GPS location) and demographic data (from her annual pass) for the new purpose. Doris consents to this update, and continues using the application.

Some time later, Doris is recommended to plan her visit to the "Long Night of Museums" (spanning activities around the full city). The day of the event, the APP suggests to keep her current GPS active in order to receive live updates of museum attendance, routes and sub-event recommendations. Doris enables this for 1 day, and the APP provides regular updates on her potential destinations, taking into account her profile (including already visited museums) and crowded locations.

Scenario 3: A Fully-fledged privacy dashboard. As there is a huge festival to promote local/regional products, the APP asks whether she would attend it, inviting her to an appetizer. She declines, and takes the opportunity to use the APP's privacy dashboard to check and modify some of the permissions given. She can also find the data gathered from her, how they were processed, where they were stored and for what purposes they were used.

Scenario 4: Decision support for WStW planners. Eva is a transportation network planner at WStW. She and her team are responsible for planning extensions of the public transport network infrastructure in order to respond to evolving mobility patterns in the city (e.g., creating new lines, increasing/decreasing the capacity and frequency of vehicles) as well as to offer advice on adjusting the transportation schedule during large-scale events (increasing/decreasing the capacity and frequency of vehicles on the transportation lines affected by and/or relevant for the event).

Thus, Eva integrates relevant information from multiple sources in a semantic data lake, together with the associated usage policies. At some point, she wants to check the validity of the existing personalized recommendations. Thus, she uses the WienMobile APP to collect feedback on alternative routes. Doris can now select between 2 suggested routes and add comments on why she has selected a given option. This is integrated into Doris' privacy dashboard, for a period of one year after the data has been collected.

When executing pattern detection algorithms, the system automatically checks to ensure that no usage policy is violated and keeps records about the processing of the data. Thanks to this logging facility, WStW can easily and transparently demonstrate (e.g., through user dashboards) that all data storage and processing complies with previously collected consents.

5 Deriving the SPECIAL-CPSS Core Vocabulary

A key goal of this paper is to illustrate how the SPECIAL vocabularies (presented in Section 3.2) can be extended to cope with practical CPSS scenarios, such as the CitySPIN smart mobility use case (Section 4). To this end, we first derive a CPSS-specific *core ontology* that reduces the semantic gap between the SPECIAL MCM and domain-specific vocabularies by providing a set of concepts that are semantically closer to the needs of the application domain than the SPECIAL MCM. Other benefits of this core vocabulary include that it can provide better guidance on deriving domain-specific vocabularies than just the very abstract SPECIAL MCM concepts. Indeed, the intention is to create a core vocabulary that can be reused for deriving usage policy vocabularies for CPSS in other domains as well (such as smart grids, smart home, smart manufacturing).

Methodologically, we ground the CPSS core ontology in information collected from literature describing a broad range of CPSSs. We collected this information by means of a Systematic Mapping Study (SMS) as proposed

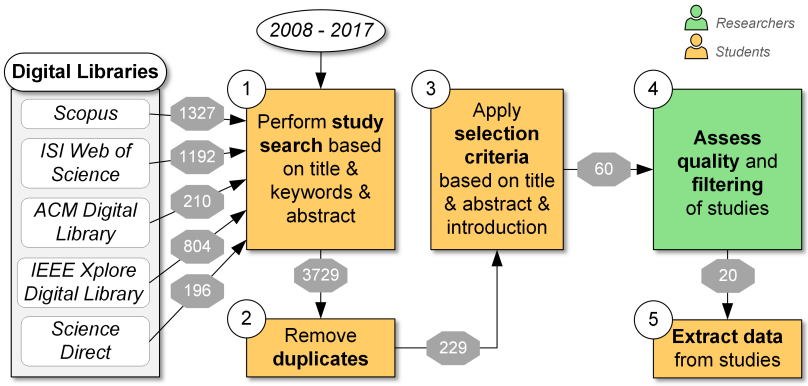

Fig. 4 Key stages of the Systematic Mapping Study that allowed extracting the information for creating the SPECIAL-CPSS core ontology.

by Kitchenham et al [24]. The goal of a SMS is to review a specific software engineering topic area and to classify primary research papers (i.e., papers describing concrete systems, but not papers that survey those systems or an aspect thereof) in that domain in order to provide an overview of a certain topic [24].

Before collecting and analyzing the literature, we detailed all envisioned study stages and their parameters in a study protocol [36] which can be consulted for further details. The study aimed to answer the following research questions in order to provide an in-depth understanding of CPSS as described in the literature: *RQ1: What is an overarching definition of CPSS? RQ2: What are application domains, goals and stages specific toCPSS? RQ3:What are main characteristics of CPSS that could be used for their classification? RQ4: What is the role of human and social elements in CPSS? RQ5: What data sources are typically used in CPSS? RQ6: How is data processed and distributed in CPSS? RQ7: What architectural approaches are applied to design and describe CPSS? RQ8: What are currently main research areas and topics and what are key challenges and emerging future work trends in CPSS?* The details of the results obtained with that protocol are available in [35] which we briefly sum up here.

Papers to be included in the study were found through a manually performed *(1) keyword-based search* in five of the largest scientific digital libraries (see Fig. 4). The search spanned the period 2008-2017 and focused on the paper title, keywords and abstract. For the selection of the query terms, the research team collected candidate terms that: were aligned with the focus of the CitySPIN research project on cyber-physical social systems as well as related areas of research such as ”internet of things”, ”sensor networks”, ”participatory sensing”. A number of search queries were formed from these terms and run on digital libraries, in order to determine the number of resulting papers that they would return, as retrieving an overly large number of papers would have made the study unfeasible. For each of the candidate queries, we also took a look at a sample of the returned papers to estimate

the quality of these papers, i.e., the level to which they fulfilled our selection criteria, especially IC1 (see details next). Finally, we settled for the following search query which lead to 3729 papers:

(cyber AND physical AND soci) OR (cyber AND physical AND human) OR (cyber AND physical AND soci* AND distributed) OR (cyber AND physical AND participatory)*

The 3729 papers were assessed for relevance based on their titles and collected into a spreadsheet which allowed *(2) duplicate detection and removal* and lead to a total of 229 papers. From these papers, 60 papers were identified as relevant for the study by *(3) applying a set of selection/exclusion criteria* on the information provided in their titles, abstracts and introductions. We tested three inclusion criteria:

- **IC1**: Studies focusing, proposing, leveraging, or analyzing a CPSS in detail. We were looking for papers that provide at least a minimal description of a concrete system in an application scenario or use case. At least one section of the paper should describe a system.
- **IC2**: Studies subject to peer review (e.g., journal papers, papers published as part of conference proceedings).
- **IC3** :Studies published since 2007.

We also checked the following exclusion criteria:

- **EC1**: Studies that are written in a language other than English, or that are not available in full-text.
- **EC2**: Secondary studies (e.g., systematic literature reviews, systematic map-ping studies, and surveys), which do not provide novel research results by their own and instead summarize work done by other researchers.
- **EC3**: Studies where a CPSS is only mentioned as a side-topic, e.g., this term appears only in the title or a reference or as an example.
- **EC4**: Studies focusing only on CPS in general, not on CPSS specifically.

Researchers involved in the study *(4) assessed the quality* of the candidate papers and selected 22 of them to include into the study. *(5) Data extraction* was guided by pre-defined extraction forms (see the study protocol [36]) which allowed to survey each paper in the same way (objectively) and reduced the room for bias. Besides bibliographic information, we collected data-items relevant to our research questions, e.g., *CPSS definition, application domain, CPSS purposes, CPSS process steps/activities, involvement of human actors, data sources, collected data*. The process of *analyzing and synthesizing* the collected data included the application of descriptive statistics and interpretation of the results with respect to the research questions.

The SPECIAL-CPSS core vocabulary is an extension of SPECIAL MCM and provides a point for further extension with use case specific vocabularies (see Figure 5). Specifically, extensions were made to the *Data* and *Purpose* concepts of SPECIAL, as described next and summed up in Tables 2 and 3. Note that we do not consider the extension of other MCM categories (*Processing*, *Storage*

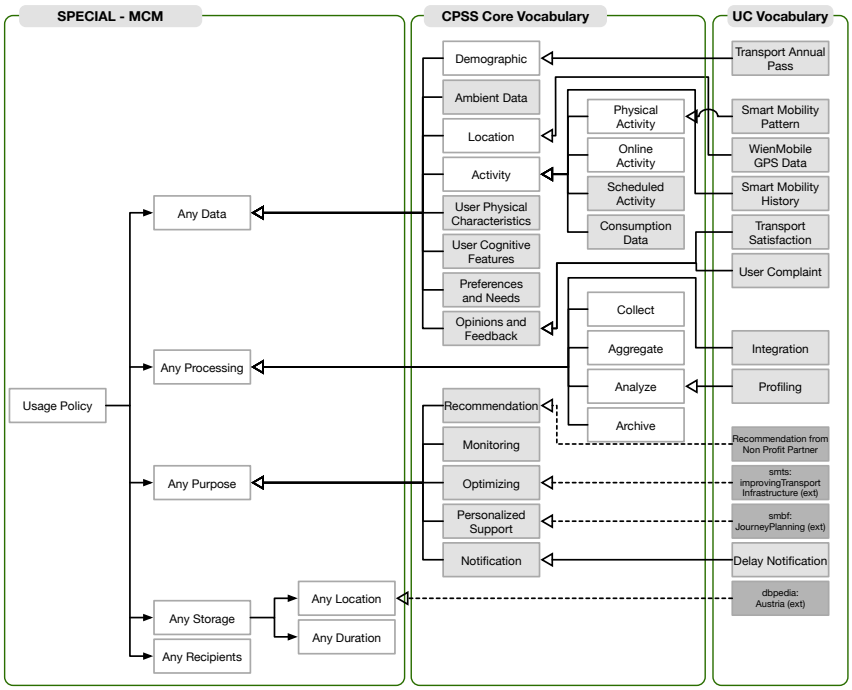

Fig. 5 Layered vocabulary: SPECIAL-MCM, SPECIAL-CPSS core ontology and the use case specific extensions.

and *Recipients*) as (i) those can be seen as more general or domain-agnostic categories and (ii) they are already well-covered in SPECIAL. In any case, we provide specific examples of use-case based extensions in Section 6.

Extensions to SPECIAL Data. CPSS span the physical, the cyber and the social spaces; the data sets most often being used in CPSS describe either the physical or the social space of the system.

In terms of the physical space, *AmbientData* provides information about the surrounding environment such as weather conditions, air quality or temperature. Increasingly, such data is collected with smart sensors installed in the participants' personal sphere, and therefore can be subject to user consent. User *Location* is another frequently collected data category, for example, through smart phones' GPS sensors.

In terms of data sets that describe different aspects of the (human) participants in the CPSS, *Activity* data is collected in various ways:

1. *PhysicalActivity* details user actions in the physical space, for example, a user's *DrivingData*, *HomeActivity*, or *MobilityData*.
2. *OnlineActivity* captures activities in the online sphere, for example, various digital traces left by the user, such as *SearchLogs*.
3. *ScheduledActivity* refers to past or future activities that were scheduled, for example, by means of the user's calendar entries. For example, the concept

Table 2 CPSS Upper Level Ontology to describe Data Sources

Category	Description	SubClasses	Sources
AmbientData	Characteristics of the physical environment	AirQuality, Temperature	[14], [20]
Location			
ActivityData	General activity data.		
PhysicalActivity	Activities performed in the physical space.	DrivingActivity	
OnlineActivity	Activities performed in the online.	SearchLogs	
ScheduledActivity	Scheduled activities (past and future).	PlannedEventData	[40]
ConsumptionData	Measurement of resource consumption.	EnergyConsumption	[11]
UserPhysicalCharacteristics		HearthRate, BloodPreassure	
UserCognitiveFeatures		MemoryProblems	[33]
PreferencesAndNeeds	Needs and preferences to be taken account during recommendations or personalized support.	WalkingPreferences	[40]
OpinionsAndFeedback	User ratings or complaints.	ServiceSatisfaction, UserComplaint	[20]

PlannedEventData could be introduced to captures events collected from a user's calendar such as done in [40].

4. *ConsumptionData* captures consumption of some resources, for example, energy consumption as recorded by smart meters [11].

Several systems, especially those with applications in the health care domain, actively collect *UserPhysicalCharacteristics* including for example, their *HearthRate* or *BloodPreassure*. Similarly, *UserCognitiveFeatures* (e.g., their attention span) are needed in those CPSSs that aim to adjust a process to these user characteristics. For example, in the smart manufacturing domain, adaptive manufacturing systems aim to improve the working conditions for aging workers by improving the human-machine interaction [33]. To that end, both physical conditions (e.g., colour blindness, short-sightedness, hearing loss) as well as cognitive features (anxiety disorders, memory problems) are collected and used within the smart manufacturing CPSS.

PreferencesAndNeeds, such as the users *WalkingPreferences*, are often used in CPSSs that offer recommendations or personal support. For example, an intelligent parking assistant suggests parking places closer/further to a meeting's place depending on whether the driver prefers to have a shorter/longer walk from the parking place to the meeting's location [40].

Finally, *OpinionsAndFeedback* provided by users are important sources of information for CPSSs that aim to adjust recommendations according to the users' perception of some service (e.g., a restaurant). *ServiceSatisfaction* records as well as *UserComplaints* are typical types of data collected. For example, user ratings are used to recommend suitable airport services in [20].

Extensions to SPECIAL Purpose. We extend the SPECIAL *purpose* category with five broad purposes (see Table 3), which emerged from our overview of CPSS systems in various domains.

Table 3 CPSS Upper Level Ontology to describe Purpose.

Category	Description	SubClasses	Sources
(Personalized) Support	User is guided during a process to achieve goals in the best possible ways while taking into account real-time conditions.	DrivingSupport, NavigationSupport, ParkingSupport, JourneyPlanning	[44], [40]
Monitoring		HealthMonitoring, MonitoringTraffic	
Optimizing	Optimizing a process or a service by adjusting it in order to achieve efficiency or effectiveness.	OptimizingEnergyConsumption, OptimizingManufact.Process, ReducingCommutingTime, SharingResources	[11]
Recommendation	Suggest an object, event or other entity based on the user's constraints/profile and ambient conditions.	EventRecommendation	[44]
Notification	Feedback provided to the system users in diverse situations, ranging from messages to alerts.	EmergencyResponse, HealthWarnings, AnomalyDetection	[14]

(Personalized) Support is the purpose of those CPSSs in which a user is guided during a process (e.g., driving, parking etc.) to achieve goals in the best possible way while taking into account real-time conditions (e.g., traffic congestion). *DrivingSupport*, *ParkingSupport* [40] , *NavigationSupport* (e.g., for visually impaired), *PersonalizedManufacturing*, *JourneyPlanning* are a few examples of more specific purposes in this category.

The purpose of *Monitoring* a process or the state of the environment is common among CPSSs, mostly as a pre-requisite to enable other purposes such as optimization or recommendation. Examples are *ManufacturingProcessTracking* and *HealthMonitoring*.

The *Optimizing* purpose is common among CPSSs. Indeed, many CPSSs have a feedback loop into their environment that allows the systems to modify the environment in ways that lead to optimization. Optimization can focus on a process or a service and it can aim at adjusting it in order to achieve efficiency or effectiveness. These adjustments often respond to changing conditions in the system's environment.

CPSSs that provide *Recommendation* services suggest an object, event or other entity based on the user's constraints/profile and ambient conditions. *EventRecommendation* is, for example, the purpose of the system presented in [44], which supports visually impaired students to find and attend suitable events on the university campus.

Notifications consist in feedback provided to the system users in diverse situations. Depending on the level of risk and danger in these situations, notifications can range from informative notes and messages to warnings (e.g., *HealthWarning*) and alerts. E.g., health warnings are provided to asthma patients depending on registered levels of pollen and air pollution in [14].

In the following section, we describe a practical workflow to define CPSS data subjects' consent and data usage policies.

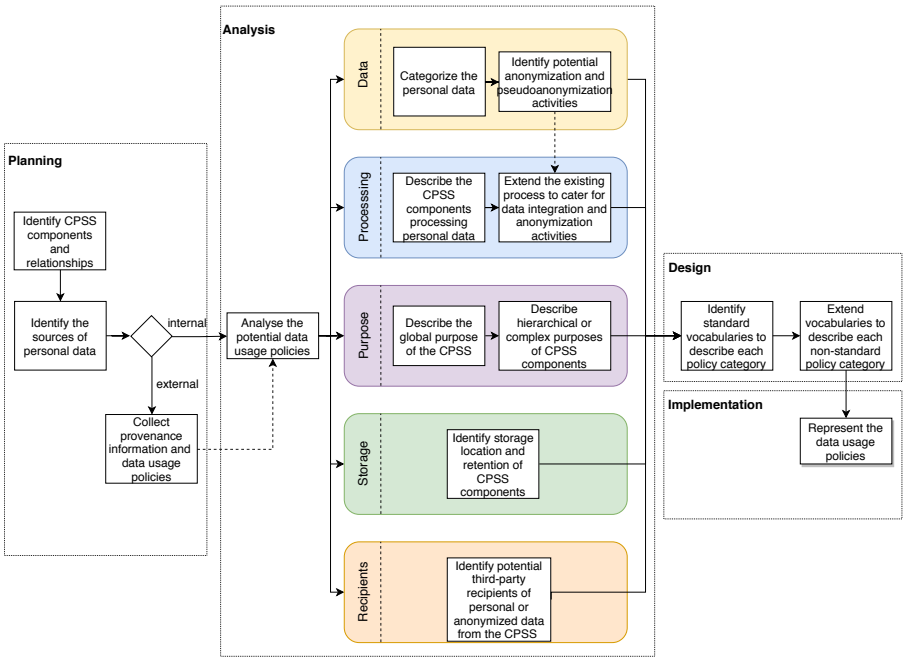

Fig. 6 Practical workflow to define CPSS data subjects' consent and data usage policies.

6 A Practical Workflow for Conceptualizing CPSS Data Usage Policies

Our practical workflow, depicted in Figure 6, is aimed at supporting CPSS owners to analyze their CPSSs and to establish the terms that will be used to represent the CPSS data usage policies. These policies (i) are used to ask for data subjects' consent, and (ii) they shall be integrated in those CPSS components processing personal data in order to facilitate transparency and compliance. The workflow consists of a sequence of steps that take into account the SPECIAL usage policy model (cf. Section 3.2) and the general guidelines of the privacy by design [10] philosophy. Note that we group the series of steps following a typical planning, analysis, design and implementation life cycle, described in the following.

6.1 Planning

In a first phase, we identify the CPSS components, relationships and sources of data that will guide the rest of the process.

6.1.1 Identify CPSS components and relationships

The first step aims at identifying all CPSS components that manage data, as well as the different relationships among them. CPSSs are often complex sys-

tems composed of components of diverse nature [45], from physical world entities (e.g. sensors, vehicles, robots, smart meters, etc.) to socio-technical systems (crowdsourcing, collective intelligence systems, etc.) and cyber components (recommenders, decision support, etc). Thus, this introductory phase must clearly reveal and describe the components and the expected flow of data. Special attention should be paid to the description of inputs and user feedback loops, a key aspect in CPSSs that will be reflected in the MCM components (e.g., in the processing and purpose categories).

6.1.2 Identify the sources of personal data

Once the components and their relationships are clearly described, this step aims to identify all sources of personal data. The concept of *personal data* is defined in the GDPR as *"any information relating to an identified or identifiable natural person ('data subject')"*. This step is of particular importance given that most CPSSs integrate different data sources, with a strong social component.

Application in the CitySPIN smart mobility use case. As a practical example of the first two steps, we identified the CPSS components and relationships as well the internal/external data for the CitySPIN use case (Section 4). The result of the analysis is shown in Figure 7.

Our component identification process consists of the following steps: First, we extract terms (e.g. *'collect'*,*'integrate'*, *'aggregate'*, etc.) from the use case as CPSS component candidates. Next, we filter out duplicates and unnecessary terms, and finally we classify the remaining terms according to the minimum core model for usage policies (cf. 2) and data subjects.

We identify two types of data subjects: WienMobil users (e.g., Doris), and WStW planners (e.g., Eva). We focus on the first type, WienMobil users, as they are the source of personal data in the use case. In particular, they share up to five types of data: *routing requests, search history, location data, event attendance*, and *real-time feedback*. In addition, we identify other non-personal data sources, external (city events, weather data, etc.) and internal (e.g. public transport data). These data are gathered in a first stage (Figure 7.1).

The collection of personal data will go through a personal data collection process (Figure 7.2) to keep track of data provenance. From this point on, the data might be processed further to create profiles (Figure 7.3) and used directly within the integration process with external data (Figure 7.4), depending on the user consent. The profiles may also be stored in the profile storage. The main processing component of the use case is the analysis (Figure 7.5). This component is responsible for producing analysis results for various data recipients, e.g., delay notifications (Figure 7.6.1) and different kinds of recommendations, i.e. personalized planning (Figure 7.6.2), for WienMobil users. In addition, the feedback of users for such recommendations (if consented) will be processed by WStW planners.

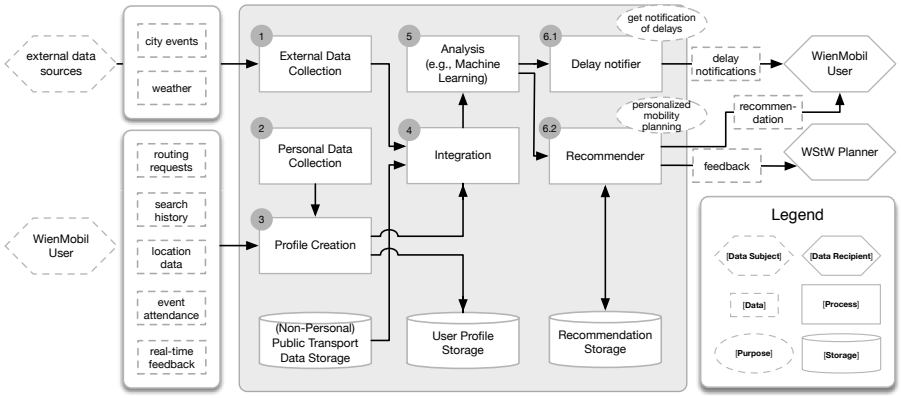

Fig. 7 CitySPIN CPSS privacy-related components.

6.1.3 Collect provenance information and data usage policies

At this point, after the first two steps, we must categorize the sources of personal data in two categories, *external* and *internal*. On the one hand, external personal data refers to personal data that is not generated in the CPSS. Note that processing personal data gathered from public sources (e.g. open data) or third-party companies is also subject to the GDPR, as the company behind the CPSS should be able to demonstrate that the data was collected and managed in compliance with the GDPR. This aspect is covered in the following phase. On the other hand, internal personal data refers to personal data generated within the CPSS. In this case, the usage policies should be represented (described below) and the appropriate data subjects' consent associated to the data should be obtained.

In this step, we first focus on the external personal data, where all provenance information (data sources, third-party contracts and terms, etc.) and data usage policies are collected. In this case, the CPSS company should take care of linking the provenance information with the policies and the concrete data that adhere to such policies. This process of linking provenance information, external policies and actual data is out of scope of this paper. In the future, we plan to extend SPECIAL to consider this aspect, implementing the concept commonly referred to as "sticky policies" [29].

6.2 Analysis and design of the data usage policies

As depicted in the workflow in Figure 6, (a) the internal data usage policies should be represented, which will then provide the basis to ask data subjects' for their consent to manage such data, or (b) the external policies and provenance information should be collected, in order to link it to the actual data

and keep track of the process. Note that the second case can be simplified (and be limited to the linkage of data and policies) if the usage policy is already provided in a standard format by the data source provider, e.g. using the Resource Description Format (RDF) [39].

When it comes to representing CPSS data usage policies using the SPECIAL model, a first step involves a deeper analysis of the potential data usage policies for the use case. This implies to analyze the concrete terms that we need to specify in each of the five elements (data, processing, purpose, storage and recipients) of the SPECIAL minimum core model (cf. see Section 3.2), summarized below. Then, a second step consists of (i) identifying standard vocabularies to represent such terms (within the SPECIAL auxiliary vocabularies or others existing and reusable vocabularies) or (ii) extending the SPECIAL auxiliary vocabularies to cover the CPSS use case needs. Examples of existing and reusable vocabularies related to CPSS are the taxonomy for planning and designing smart mobility services [12], the Road Traffic Management ontology [3] and other smart city ontologies [15], to name but a few.

In the following, we describe the analysis and design for each of the five elements (data, processing, purpose, storage and recipients) of the SPECIAL minimum core model.

- The element '*Data*' describes the personal data collected from a data subject. First, the already identified CPSS elements and data sources must be analyzed further to categorize such data. In this step, rather than the actual data, the category of the data and the potential *skeleton* (i.e., structure) of typical data items should be identified.
  In a second step, following the privacy by design [10] philosophy, potential anonymization and pseudoanonymization activities shall be identified. This step plays an important role as the GDPR does not concern the processing of anonymous information, i.e., *"information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable"*. However, if the CPSS company is in charge of the collection of personal data and their anonymization, then the appropriate data subject's consent should be obtained. Thus, this step is the place where we can identify whether anonymization, pseudonymization and aggregation of data (e.g., applying techniques such as k-anonymity and l-diversity [18,1]) is applied or can be applied in the CPSS. The strength of the anonymization techniques and possible attacker models based on combining anonymous data and other knowledge is out of scope for this paper. In our CitySPIN use case, we focus on consent, given the description and the activities in Figure 7 and the fact that anonymization impacts utility.
  Then, the category of the final relevant personal data in the CPSS (i.e., data that cannot be further anonymized) needs to be represented in the SPECIAL policy language. Thus, we must identify standard vocabularies to describe the personal data categories, extending or providing new vo-

cabularies if needed. As mentioned before, this step completely depends on the particular scenario, hence it is expected that the initial auxiliary vocabulary presented in SPECIAL (cf. see Table 1) needs to be extended with use case specific ontologies.

Application in the CitySPIN smart mobility use case. In our previous *smartMobility* example, following from the description and the activities in Figure 7, the analysis would reveal that the CPSS needs to store (i) location data, consisting of (potentially real-time) GPS locations of the user, (ii) routing requests, including source and target destination at a particular moment in time, (iii) a history of lookups in the wien Mobile APP, which is basically a log of user queries to the APP, (iv) event attendance, which is a particular case of a search of a route request to attend a specific event, and (v) real-time feedback of alternative routes. In the following, we show the analysis and vocabulary selection/creation for each of them. Note that we use the `wm` prefix to denote the use-case specific `Wien Mobile` namespace.

* Location data, i.e., GPS locations of the user, can be directly represented with the existing `svd:location` in SPECIAL. Nonetheless, note that the company might decide to have a more informative consent, stating that the data is collected from the GPS location of the APP. In such case, a specific `wm:WienMobileGPSData` category could be created, as a subclass of `svd:location`. We consider this as an instantaneous location when searching for a route, while we further refer to `wm:WienMobileGPSDataRealTime` as a continuous location stream.
* Event data, i.e., event records for a particular user and time, are underrepresented in SPECIAL, hence a particular extension is needed. In this case, we can make use of the proposed CPSS ontology to describe data source (see Table 2), e.g., using the `svd-cpss:ScheduledActivity` data category.
* Route requests and history of lookups, i.e., user's lookups in the APP. This data category could potentially be covered by the categories defined in the SPECIAL project, `svd:activity`, which represents data concerning user's activities, and its subcategory `svd:online-activity`, considering *data describing online activities such as browsing, liking on social networks, posting, etc.* [5]. Although these categories should cover several scenarios, fine-grained, company-specific categories can be preferred. For instance, in our example, we decide to create the class `wm:SmartMobilityHistory` extending `svd:Activity` to represent the history of lookups in the WienMobil APP. In addition, given that a mobility pattern can be extrapolated from the data, which will be part of the profile as specified in our diagram in Figure 7, we also create the `wm:SmartMobilityPattern` category, extending `svd:PhysicalActivity`.
* Service ratings, i.e. information to reflect the user's satisfaction with the Wien Mobile service. In this case, the SPECIAL auxiliary vocabulary provides the general category `svd:preference`, which stands for *data about an individual's likes and dislikes - such as favorite color or musical*

tastes [5]. An organization designing a CPSS may need to provide further details on the information collected. In that case, the `svd:preference` class should be extended to cope with the respective needs. In our particular scenario, we can make use of our aforementioned CPSS extensions (Table 2). Thus, we define the novel `wm:TransportSatisfaction` and `wm:UserComplaint` categories as subclasses of `svd-cpss:OpinionsAndFeedback`.

- *Processing*. The element '*Processing*' specifies the operations that are performed on the personal data. Given the inherent complexity of CPSSs, where multiple components are often organized in a 'pipeline' architecture, the first step is to analyze the information flow and the components identified in the planning phase, and to describe the CPSS components processing personal data. Given that the identification of components could be incomplete, as some processing activities could be implicit (e.g., a machine learning component can have several input sources and an implicit data integration process could be required), a second step considers to extend the previous analysis of components to identify and describe such potential implicit processes. Special attention shall be paid to describing (i) potential data anonymization and pseudoanonymization activities emerging from the previous 'data' phase (as represented with a dashed arrow in Figure 6), and (ii) integration activities, which are often present while combining different data sources in a CPSS scenario.
  Once all components and activities have been identified, similarly to the previous case, standard vocabularies to represent the concrete CPSS processing must be identified, or new concepts must be provided if needed. Note that, given the broad spectrum of CPSS applications and components, CPSS processing could cover all potential processing activities of an information system. SPECIAL provides a set of processing concepts (summarized in Table 1) that are more closely related to data protection, such as `svpr:Aggregate`, `svpr:Analyze`, `svpr:Collect`, etc.

Application in the CitySPIN smart mobility use case. In the following, we review the most important CPSS stages/activities emerging from our CitySPIN use case and the description and the activities in Figure 7:

* Data collection - can be directly mapped to the SPECIAL `svpr:Collect` concept.
* Profile creation - is not directly present in the SPECIAL auxiliary vocabularies. However, it is explicitly recommended to provide a use case specific concept as a subclass of `svpr:Analyze` [5]. In our case, we created the `wm:Profiling` concept.
* Data integration - can be (partially) mapped to the SPECIAL `svpr:Derive` and `svpr:Aggregate` concepts. Given that the mapping could be inaccurate, a new concept extending the general `spl:AnyProcessing` class could be provided. In our case, we use `wm:Integration`.
* Data analysis - can be directly mapped to the SPECIAL `svpr:Analyze` concept.

  * Proactive recommendations and notifications. In this case, SPECIAL considers such cases rather a 'purpose' (described below). Then, the 'processing' leading to the concrete recommendation and notification could be seen as a result of the previous steps, in particular `svpr:Analyze`. In our CitySPIN example, we follow this approach. Note that other use cases could need further processing steps, which could be included as processing if needed.

- *Purpose.* The element '*Purpose*' specifies the objective that is associated with data processing. In CPSSs, we could establish a two-phase identification of (a) describing the global purpose of the CPSS, and (b) analyzing and describing a hierarchical structure of the identified purposes of the CPSS components. The rationale behind this approach is that, whereas the final purpose can be almost extracted from the textual description of the use case, CPSSs often involve complex components and relationships that might be re-purposed for a specific goal, hence further analysis is required. Once these purposes are identified, standard vocabularies, or extensions, to describe CPSS purposes must be put in place.

  Application in the CitySPIN use case. In the following, for exemplary purposes, we review the purposes identified in our CitySPIN use case:

  * Notifying of delays is under represented in the SPECIAL auxiliary vocabularies. Note, however, that it would be possible to make use of the `svpu:Current` concept (i.e., completion and support of activity for which data was provided), as a general concept if the main goal of the data collection and the CPSS is to provide such notifications to the user. In our CitySPIN use case, we create a specific `wm:DelayNotification` extending the proposed `svd-cpss:Notification` in the CPSS ontology to describe purposes (see Table 3).
  * Recommendations are only implicitly represented in SPECIAL, as part of marketing purposes (`svpu:Marketing`). Thus, proactive recommendations (`svd-cpss:Recommendation`) are specifically considered in the CPSS ontology (see Table 3). In our use case, additionally, we also reuse the existing `smbf:JourneyPlanning` category and we define a specific recommendation for non-profit partners (`wm:RecommendationNonProfitPartner`, which extends the proposed `svd-cpss:Recommendation`).
  * Providing feedback, for our last scenario, can be represented with the SPECIAL `svpu:Feedback` concept. In addition, given that the final objective is to optimize the transport infrastructure, we can consider the CPSS ontology to describe an optimization purpose (`svd-cpss:Optimizing`). In particular, in our CitySPIN use case, we make use of the existing `smts:improvingTransportInfrastructure` category.

- *Storage.* The element '*Storage*' specifies the location and temporal retention policy for the CPSS data. In the particular case of a CPSS, and given its potential distribution, this implies to identify the storage location and the required data retention of the individual CPSS components. Data retention

periods can be then simply represented as a numeric range in the SPECIAL policy language (cf. Section 3.2). In turn, storage locations can be listed with the SPECIAL auxiliary vocabulary (e.g., using concepts such as `svl:EU` or `svl:ThirdParty`) or be extended if finer details are needed by the use cases. Note that the former should cover most CPSS use cases as the SPECIAL vocabulary for locations is designed to cover the GDPR requirements of specifying (i) whether the information is stored in the EU or in countries with similar data protection legislation, and (ii) whether the information is kept by the data controller or stored outside its boundaries [5].

Application in the CitySPIN use case. In our CitySPIN use case, we only need to specify that data are stored on the company servers in Austria. Thus, we make use of both the SPECIAL `svl:OurServers` concept and the well-established `dbpedia:Austria` term. As for temporal retention, we just need to specify the number of days, from a single day up to 2 years, depending on the scenario.

- *Recipients.* Finally, the element '*Recipients*' specifies who can receive the results of the CPSS personal data processing. In this case, potential third-party recipients of personal data from the CPSS should be identified. Given the inherent complexity of CPSSs, this step may involve careful inspection of all (potentially distributed) CPSS components, involved partners and stakeholders. Then, as in previous elements, standard vocabularies to describe CPSS recipients must be analyzed, and extended where needed. Similarly to the *storage* element, SPECIAL auxiliary vocabularies (cf. see Table 1) should cover most of the CPSS use cases, while specific fine-grained descriptions may need some extensions, e.g., using the FOAF [9] and PROV [26] vocabularies.

  Application in the CitySPIN use case. In the particular case of CitySPIN, no recipients are needed, hence the use of the SPECIAL `svr:Ours` term.

6.3 Implementation: Representing the data usage policies

As a last phase, the final data usage policies should be represented using the SPECIAL policy language, using the selected terms in the previous phase. Thus, each concrete scenario should be reviewed carefully, and each component of the SPECIAL MCM model should be represented in a simple but complete way, aiming to reflect the scenario (i.e. the textual policy) precisely. Obviously, the process can reveal some gaps that should be filled (e.g., if the data retention time has not been identified), which could require to repeat some of the previous steps of the proposed workflow.

The final policies in the CitySPIN use case are presented in the next section.

7 Validation: User Policy Representation in CitySPIN Use Cases

This section presents the results of the practical application of the workflow to establish CPSS data subjects' consent for specific use cases (described in Section 6) to our CitySPIN smart mobility use case, (shown in Section 4). Once the main CPSS components and personal data sources have been identified (see Figure 7), and we have carefully selected or extended vocabularies (see a summary of the use case specific extensions in Figure 5) for each of the components of the SPECIAL MCM model (data, processing, purpose, storage and recipients), we then proceed to represent the data usage policies.

In the following, we summarize the final policies for each of the "personal data" scenarios in the aforementioned CitySPIN smart mobility use case. Note that we do not specify a policy for scenario 3, as it is built upon the previously defined policies to exemplify the use of the privacy dashboard, providing transparency to data subjects.

7.1 Scenario 1: A personalized mobility planning

The study and analysis of the first scenario of the *Wien mobile* use case (as shown in Section 4) result in the following textual policy: *"The history of transport routing data and GPS location data (at the moment of the search) can be integrated with other non-personal data sources (city events, environment data, traffic congestions) and analyzed to create a mobility profile, in order to recommend best routes and notify about delays in the future. These profiles are stored for two years on the company servers in Austria"*. This policy is formalized in Listing 4.

Listing 4 Final policy of the CitySPIN scenario 1 - personalized mobility planning

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( spl:hasData
      ObjectUnionOf(
         wm:SmartMobilityHistory wm:MobilityPattern wm:WienMobileGPSData ))
   ObjectSomeValueFrom( spl:hasProcessing
      ObjectIntersectionOf(
         wm:Profiling wm:Integration svpr:Collect ))
   ObjectSomeValueFrom( spl:hasPurpose
      ObjectUnionOf(
         svd-cpss:Recommendation smbf:JourneyPlanning wm:DelayNotification ))
   ObjectSomeValueFrom( spl:hasStorage
      ObjectIntersectionOf(
         ObjectSomeValuesFrom( spl:hasLocation
            ObjectIntersectionOf( svl:OurServers dbpedia:Austria ))
         DataSomeValuesFrom( spl:durationInDays
            DatatypeRestriction( xsd:integer
              xsd:maxInclusive "730"^^xsd:integer ))))
   ObjectSomeValueFrom( spl:hasRecipient svr:Ours ) )
```

Thanks to the previous steps of the workflow, the formalization of the policy is almost straightforward. First, the data category can be represented with a union (`ObjectUnionOf`) of three use-case specific terms (`wm:SmartMobilityHistory`,

wm:MobilityPattern and wm:WienMobileGPSData) that accurately reflect the personal data involved in the scenario. The type of processing, also revealed during the identification of CPSS components, is restricted to profiling and integration, both represented with specific use-case terms (wm:Profiling and wm:Integration). Note that we also include the data collection process (svpr:Collect) although it was not explicit in the policy, as we assume the company is the responsible for collecting the data. As for the purpose, together with the general recommendation (svd-cpss:Recommendation), we consider the delay notification goal (wm:DelayNotification), and the planning of the journey purpose (smbf:JourneyPlanning). Finally, the storage (location and duration) and recipient (just ours) directly follow from the use case description and are encoded according to the SPECIAL policy language (e.g. the two year period is represented with a xsd:maxinclusive restriction).

7.2 Policies of scenario 2: Event Partnership

The analysis of the second scenario results in two different policies. The first one extends scenario 1 adding the processing of demographic data (from the transport annual pass) to receive partnership recommendations for non-profit cultural and heritage events, related to user's mobility patterns. Listing 5 shows the formalization of this extended policy.

Listing 5 Final policy of the CitySPIN scenario 2 - Event Partnership

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( spl:hasData
      ObjectUnionOf(
         wm:SmartMobilityHistory wm:MobilityPattern wm:WienMobileGPSData
         wm:AnnualPass ))
   ObjectSomeValueFrom( spl:hasProcessing
      ObjectIntersectionOf(
         wm:Profiling wm:Integration svpr:Collect ))
   ObjectSomeValueFrom( spl:hasPurpose
      ObjectUnionOf(
         wm:RecommendationNonProfitPartner smbf:JourneyPlanning
         wm:DelayNotification ))
   ObjectSomeValueFrom( spl:hasStorage
      ObjectIntersectionOf(
         ObjectSomeValuesFrom( spl:hasLocation
            ObjectIntersectionOf( svl:OurServers dbpedia:Austria ))
         DataSomeValuesFrom( spl:durationInDays
            DatatypeRestriction( xsd:integer
              xsd:maxInclusive "730"^^xsd:integer ))))
   ObjectSomeValueFrom( spl:hasRecipient svr:Ours ) )
```

Similarly to the previous case, the representation of the policy follows from all previous steps. In this case, the only modifications are the inclusion of the demographic data (wm:annualPass) and the new partnership purpose (wm:RecommendationNonProfitPartner). Note that both concepts are represented using specific use-case terms, hence they should include a human-readable comprehensive definition of the actual type of data considered in each case (e.g. cultural and heritage events in the case of the partner recommendations).

Then, a second policy extends the previous one considering live updates and recommendations based on real-time GPS location for attended events. In our scenario, the user only consents for a period of one day. This policy is shown in Listing 6.

Listing 6 Final policy of the CitySPIN scenario 2 - real time recommendations

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( spl:hasData
      ObjectUnionOf(
         svd-cpss:PlannedEventData
         wm:SmartMobilityHistory wm:MobilityPattern
         wm:WienMobileGPSDataRealTime ))
   ObjectSomeValueFrom( spl:hasProcessing
      ObjectIntersectionOf(
         wm:Profiling wm:Integration svpr:Collect ))
   ObjectSomeValueFrom( spl:hasPurpose
      ObjectUnionOf(
         wm:RecommendationNonProfitPartner smbf:JourneyPlanning
         wm:DelayNotification ))
   ObjectSomeValueFrom( spl:hasStorage
      ObjectIntersectionOf(
         ObjectSomeValuesFrom( spl:hasLocation
            ObjectIntersectionOf( svl:OurServers dbpedia:Austria ))
         DataSomeValuesFrom( spl:durationInDays
            DatatypeRestriction( xsd:integer
              xsd:maxInclusive "1"^^xsd:integer ))))
   ObjectSomeValueFrom( spl:hasRecipient svr:Ours ) )
```

Similarly to the previous cases, the policy represents the data with a union (`ObjectUnionOf`) of terms to capture the real time use-case location data history (`WienMobileGPSDataRealTime`) and the planned event data (in this case using the CPSS upper level ontology via `svd-cpss:PlannedEventData`). The processing, purpose, storage and recipients follow the previous examples.

7.3 Policy of Scenario 4: Decision support for WStW planners

Finally, the last scenario builds upon scenario 1 and considers feedback (potentially in real-time) on personalized alternative routes. In this case, the storage is limited to one year after collection. This policy is formalized in Listing 7.

In this case, the personal data include the use-case specific feedback, which is represented with use-case specific terms (`wm:TransportSatisfaction` and `wm:UserComplaint`). The processing and recipients are similar to the previous case, whereas we reduce the storage to 1 year and, in this case, we adapt the purpose to the general provision of feedback (via the existing `svpu:Feedback`). To be more transparent, we also include the implicit use-case specific purpose of improving the transport infrastructure (`smts:improvingTransportInfrastructure`).

Listing 7 Final policy of the CitySPIN scenario 4 - Feedback for decision support

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( spl:hasData
      ObjectUnionOf(
         wm:TransportSatisfaction wm:UserComplaint wm:SmartMobilityHistory
         wm:MobilityPattern wm:WienMobileGPSData ))
    ObjectSomeValueFrom( spl:hasProcessing
      ObjectIntersectionOf(
         wm:Profiling wm:Integration svpr:Collect ))
   ObjectSomeValueFrom( spl:hasPurpose
      ObjectUnionOf(
         svpu:Feedback smts:improvingTransportInfrastructure ))
   ObjectSomeValueFrom( spl:hasStorage
      ObjectIntersectionOf(
         ObjectSomeValuesFrom( spl:hasLocation
            ObjectIntersectionOf( svl:OurServers svl:EU ))
         DataSomeValuesFrom( spl:durationInDays
            DatatypeRestriction( xsd:integer
              xsd:maxInclusive} "365"^^xsd:integer ))))
   ObjectSomeValueFrom( spl:hasRecipient svr:Ours ) )
```

8 Summary and Future Work

Privacy protection is a fundamental but challenging requirement in the context of Cyber-Physical Social Systems (CPSSs), which, by definition collect and make use of user-specific data (from the "social" space). CPSS owners need to ensure compliance with user policies be transparent in terms of how users' data is being processed. While automated compliance checking and transparency can be achieved based on formally represented usage policies, existing policy languages that enable specifying user consent are domain-agnostic and require adaptation when used in concrete use cases. For example, in this paper we exemplify extending the SPECIAL domain-agnostic policy language for describing user policies in a smart mobility use case provided by Vienna's largest utility provider.

We relied on an approach which aims to support CPSS owners in general, and WStW in particular, in adapting the policy language for the needs of their own use cases in two ways: (1) by providing the SPECIAL-CPSS core-vocabulary that already extends the domain-agnostic SPECIAL terms towards the domain of CPSS; (2) by proposing a novel practical workflow that can be used to elicit vocabularies for defining CPSS data subjects' consent and data usage policies. We validate the resulting vocabularies (both core and use case specific) by demonstrating that they can be used successfully to construct usage policies according to the SPECIAL specification.

A current limitation of this work is that the SPECIAL-CPSS core vocabulary and the proposed workflow have been tested on a mobility use case only. Our ongoing work focuses on reusing and validating these two outcomes on another CPSS use case from WStW in the domain of smart energy grids.

Additionally, the coverage of the SPECIAL-CPSS core vocabulary was influenced by the selection of query keywords used in the mapping study by not considering related terms due to practical considerations of the study feasibility. Therefore, we also focus on further improvements of the SPECIAL-CPSS core vocabulary in terms of (1) aligning it with foundational ontologies; (2) grounding it in agency models that better reflect the social aspect of CPSS and (3) planning follow-up studies on related terms, such as "participatory sensing" to make it more comprehensive. In the future, we plan to extend our model with layers dedicated to concrete domains, e.g., smart grid, smart manufacturing, smart home. Finally, we plan to extend our SPECIAL-CPSS approach with the concept of sticky policies for those data coming from external sources.

References

1. Aggarwal, C.C., Philip, S.Y.: A general survey of privacy-preserving data mining models and algorithms. In: Privacy-preserving data mining, pp. 11–52. Springer (2008)
2. Bellare, M., Yee, B.: Forward integrity for secure audit logs. Tech. rep., Computer Science and Engineering Department, University of California at San Diego (1997)
3. Bermejo, A., Villadangos, J., Astrain, J.J., Cordoba, A.: Ontology based road traffic management. In: Proc. of Intelligent Distributed Computing, pp. 103–108. Springer (2013)
4. Bonatti, P., Kirrane, S., Petrova, I., Sauro, L., Kerschbaum, C., Pirkova, E.: Special deliverable 2.6: Formal representation of the legislation v2 (2018). URL `https://www.specialprivacy.eu/images/documents/SPECIAL_D26_M21_V10.pdf`
5. Bonatti, P., Kirrane, S., Petrova, I., Sauro, L., Schlehahn, E.: Special deliverable 2.1: Policy language v1 (2017). URL `https://www.specialprivacy.eu/images/documents/SPECIAL_D2.1_M12_V1.0.pdf`
6. Bonatti, P., Kirrane, S., Polleres, A., Wenning, R.: Transparent personal data processing: The road ahead. In: Proc. of TELERISE, pp. 337–349 (2017)
7. Bonatti, P.A., Coi, J.L.D., Olmedilla, D., Sauro, L.: A rule-based trust negotiation system. IEEE Trans. Knowl. Data Eng. **22**(11), 1507–1520 (2010)
8. Bonatti, P.A., Kirrane, S.: Big data and analytics in the age of the gdpr (2019)
9. Brickley, D., Miller, L.: Foaf vocabulary specification 0.91 (2010)
10. Cavoukian, A.: Privacy by design in law, policy and practice. A white paper for regulators, decision-makers and policy-makers (2011)
11. Chen, S., Liu, T., Gao, F., Ji, J., Xu, Z., Qian, B., Wu, H., Guan, X.: Butler, Not Servant: A Human-Centric Smart Home Energy Management System. IEEE Communications Magazine **55**(2), 27–33 (2017)
12. Cledou, G., Estevez, E., Barbosa, L.S.: A taxonomy for planning and designing smart mobility services. Government Information Quarterly **35**(1), 61–76 (2018)
13. Cranor, L.F.: Web privacy with P3P - the platform for privacy preferences. O'Reilly (2002)
14. Dao, M.S., Pongpaichet, S., Jalali, L., Kim, K., Jain, R., Zettsu, K.: A Real-time Complex Event Discovery Platform for Cyber-Physical-Social Systems. Proc. of ICMR pp. 201–208 (2014)
15. Espinoza-Arias, P., Poveda-Villalón, M., García-Castro, R., Corcho, O.: Ontological representation of smart city data: From devices to cities. Applied Sciences **9**(1), 32 (2019)
16. Falkvinge, R.: Airport: "we're tracking every single footstep you take and can connect it to your mail address, but your privacy is safe because we say so" (2017). URL `https://falkvinge.net/2017/04/15/schiphol-airport-tracking-every-single-footstep/`
17. Fatema, K., Hadziselimovic, E., Pandit, H.J., Debruyne, C., Lewis, D., O'Sullivan, D.: Compliance through informed consent: Semantic based consent permission and data management model. In: Proc of PrivOn (2017)
18. Ghinita, G., Karras, P., Kalnis, P., Mamoulis, N.: Fast data anonymization with low information loss. In: Proc. of VLDB, pp. 758–769. VLDB Endowment (2007)

19. Hildebrandt, M.: Smart technologies and the end (s) of law: novel entanglements of law and technology. Edward Elgar Publishing (2015)
20. Hussein, D., Park, S., Han, S.N., Crespi, N.: Dynamic Social Structure of Things: A Contextual Approach in CPSS. IEEE Internet Computing **19**(3), 12–20 (2015)
21. Iannella, R., Villata, S.: Odrl information model 2.2. W3C Recommendation (2018)
22. Information Commissioner's Office (ICO) UK: Getting ready for the GDPR (2017). URL https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment/getting-ready-for-the-gdpr/
23. Kagal, L., Finin, T.W., Joshi, A.: A policy language for a pervasive computing environment. In: Proc. of POLICY, pp. 63– (2003)
24. Kitchenham, B.A., Budgen, D., Pearl Brereton, O.: Using mapping studies as the basis for further research - A participant-observer case study. Information and Software Technology **53**(6), 638–651 (2011)
25. Kolovski, V., Hendler, J., Parsia, B.: Analyzing web access control policies. In: Proc. of WWW, pp. 677–686 (2007)
26. Lebo, T., Sahoo, S., McGuinness, D.: Prov-o: The prov ontology. W3C Recommendation, April (2013)
27. Ly, L.T., Maggi, F.M., Montali, M., Rinderle-Ma, S., van der Aalst, W.M.: Compliance monitoring in business processes: Functionalities, application, and tool-support. Information systems **54**, 209–234 (2015)
28. Microsoft Trust Center: Detailed GDPR Assessment (2017). URL http://aka.ms/gdprdetailedassessment
29. Mont, M.C., Pearson, S., Bramhall, P.: Towards accountable management of identity and privacy: Sticky policies and enforceable tracing services. In: Database and Expert Systems Applications, pp. 377–382. IEEE (2003)
30. Motik, B., Patel-Schneider, P.F., Parsia, B.: OWL 2 Web Ontology Language – Structural Specification and Functional-Style Syntax (Second Edition). W3C Recommendation (2012)
31. Nymity: GDPR Compliance Toolkit (2017). URL https://www.nymity.com/gdpr-toolkit.aspx
32. Pandit, H., Lewis, D.: Modelling provenance for gdpr compliance using linked open data vocabularies. In: Proc of PrivOn (2017)
33. Peruzzini, M., Pellicciari, M.: A framework to design a human-centred adaptive manufacturing system for aging workers. Advanced Engineering Informatics **33**, 330–349 (2017)
34. Pulls, T., Peeters, R., Wouters, K.: Distributed privacy-preserving transparency logging. In: Proc. of WPES (2013)
35. Sabou, M., Musil, A.: Cityspin deliverable 2.1: Cyber-physical social systems blueprint (v.1) (2018). URL http://cityspin.net/wp-content/uploads/2017/10/D2.1.pdf
36. Sabou, M., Musil, A., Musil, J., Biffl, S.: Protocol for: A Systematic Mapping Study of Cyber-Physical Social Systems. Tech. Rep. IFS-QSE 18-02, TU Wien, Austria (2018). URL http://qse.ifs.tuwien.ac.at/publication/IFS-QSE-18-02.pdf
37. Sackmann, S., Strüker, J., Accorsi, R.: Personalization in privacy-aware highly dynamic systems. Communications of the ACM **49**(9) (2006)
38. Scherp, A., Saathoff, C., Franz, T., Staab, S.: Designing core ontologies. Appl. Ontol. **6**(3), 177–221 (2011). URL http://dl.acm.org/citation.cfm?id=2351285.2351289
39. Schreiber, G., Raimond, Y.: Rdf 1.1 primer (2014)
40. Smirnov, A., Shilov, N., Gusikhin, O.: Socio-cyberphysical System for Proactive Driver Support - Approach and Case Study. Proc. of ICINCO pp. 289–295 (2015)
41. Sutton, A., Samavi, R.: Blockchain enabled privacy audit logs. In: Proc. of ISWC, pp. 645–660 (2017)
42. Uszok, A. and Bradshaw, J.M. and Jeffers, R. and Suri, N. and Hayes, P.J. and Breedy, M.R. and Bunch, L. and Johnson, M. and Kulkarni, S. and Lott, J.: KAoS policy and domain services: Towards a description-logic approach to policy representation, deconfliction, and enforcement. In: Proc. of POLICY, pp. 93–96 (2003)
43. Wang, F.Y.: The emergence of intelligent enterprises: From CPS to CPSS. IEEE Intelligent Systems **25**(4), 85–88 (2010)
44. Xiao, J., Joseph, S.L., Zhang, X., Li, B., Li, X., Zhang, J.: An Assistive Navigation Framework for the Visually Impaired. IEEE Transactions on Human-Machine Systems **45**(5), 635–640 (2015)

45. Xiong, G., Zhu, F., Liu, X., Dong, X., Huang, W., Chen, S., Zhao, K.: Cyber-physical-social system in intelligent transportation. IEEE/CAA Journal of Automatica Sinica **2**(3), 320–333 (2015)
46. Zyskind, G., Nathan, O., et al.: Decentralizing privacy: Using blockchain to protect personal data. In: Proc. of SPW, pp. 180–184 (2015)