CITYSPIN

Cyber-Physical Social Systems for City-wide Infrastructures

Deliverable 6.1: Privacy policy formalization (v.1)

Authors	:	Javier D. Fernández
Dissemination Level	:	Public
Due date of deliverable	:	30.06.2018 (v1)
Actual submission date	:	15.07.2018 (v1)
Work Package	:	6. Secure Data Access & Privacy
Type	:	Report
Version	:	1.0

Abstract

The overall goal of WP6 is to harness enterprise linked data technologies to comply with the legal privacy & security requirements for handling sensitive information within CPSS. To this end, we follow a privacy-by-design approach, where legal and internal policies (in particular related to the new EU GDPR) need to be formally represented. These policies will guide automatic transparency and compliance mechanisms to document any processing of personal data and to verify that this is compliant with the defined policies. This deliverable reports on our current efforts towards such formalization of policies in a CPSS scenario.

FFG

Project Funded by FFG – IKT der Zukunft Programme

Project Number: 861213 **Start date: 01.10.2017** **Duration: 30 months**

History

Version	Date	Reason	Revisited by
0.3	23.06.2018	Draft of structure and policy formalisation	JF
0.4	1.07.2018	Inclusion of CitySPIN use cases	JF
0.5	4.07.2018	Adaptation of examples to new use cases	JF
0.6	6.07.2018	Initial identification of vocabularies	JF
0.8	10.07.2018	Draft of the practical scenario	JF
0.9	11.07.2018	Inclusion of outcomes from D2.1	JF
1.0	13.07.2018	Minor corrections	JF

Author List

Project Partner	Name(Initial)	Contact Information
WU Wien	Javier D. Fernández (JF)	javier.fernandez@wu.ac.at

Executive Summary

One of the main characteristics of *Cyber-Physical Social Systems* (CPSSs) is the focus on the social component. Although this fact is crucial to both generate relevant data and make more informed decisions, it comes at the cost of managing a complex engine often fuelled by personal data. In this context, privacy and security requirements emerge as key but challenging concerns in CPSSs.

WP6 addresses these concerns using the new EU General Data Protection Regulation (GDPR) as the baseline, providing methodologies and tools for privacy-aware data access in CPSSs. In particular, WP6 is aimed at using Linked Data technologies to establish a CPSS-tailored transparency and compliance architecture that can (i) formalize the relevant data policies in the system, (ii) document in an accountable and immutable manner any processing of personal data and (iii) enable automated compliance checking to assure that the CPSS company complies with the requirements of the GDPR, all of them (iv) under a secure environment through access control and encryption.

This deliverable primarily focuses on the first goal, i.e. how to represent data subject's consent and data usage policies in formal terms, in a CPSS scenario. To that extend, we first motivate the problem with two use case scenarios, and we review existing policy languages. Then, we decide to follow the policy language developed within the EU H2020 SPECIAL project, which is aimed at GDPR transparency and compliance. We analyze the SPECIAL's usage policy language and its suitability for CPSSs. Overall, we conclude that this policy formalization fits most of the needs of CPSSs, but further vocabulary extensions might be needed to cope with use case specific requirements. In order to help in this process, this deliverable introduces a methodology to guide companies to establish CPSS data subjects' consent and data usage policies. In addition, we provide concrete examples taking into account the outcomes of our CPSS blueprint in D2.1, such as the main CPSS domains, collected social data, goals, activities, etc.

The results of this deliverable will guide our future steps in WP6 towards GDPR transparency and compliance in CPSSs, but also the scalable Linked Data platform in WP4 and the proof of concepts in WP7 (as they have to integrate the proposed policies), as well as the analysis and monitoring in WP5 (subject to privacy-aware data access).

Table of Content

List of Figures

List of Tables

1 Introduction

Adequate privacy protection is a fundamental requirement in the context of *Cyber-Physical Social Systems* (CPSSs) [26], which often make use of and integrate sensitive information from various sources.

The goal of **WP6** of the CitySPIN project, entitled "Secure Data Access & Privacy", is to provide methodologies and tools for privacy-aware data access in CPSSs. In particular, we focus on the new EU General Data Protection Regulation (GDPR) as the baseline, given that this is one of the most restrictive regulations, hence our work can be extrapolated to a more general scenario.

Implementing a privacy-by-design, GDPR-compliant CPSS infrastructure is a challenge per se. In this scenario, CPSS companies must provide transparency with respect to personal data processing and sharing within and between organisations. Additionally companies need to demonstrate that their systems and business processes comply with usage constraints specified by data subjects. At the core of any transparency and compliance architecture is the logging of data processing and sharing events in a manner than can be used to verify compliance with relevant policies.

Thus, the main objective of WP6 is to harness Linked Data architectures to enable these documentation needs, being able to represent CPSS data usage policies and the log of data processing and sharing events, including the consent provided by the data subject and subsequent changes to or revocation of said consent. The resulting CPSS platform shall further enable automated compliance checking by exploiting the formal semantics of policies and checking them against traces of data processing stored in the log. Furthermore, the CPSS platform must be secured, implementing performant access-control and encryption mechanisms, in conformance with the identified CitySPIN use case requirements (such as data volume and change frequency).

1.1 Deliverable Goal

The goal of Deliverable D6.1 is to establish a policy language syntax and semantics that can cope with CPSS requirements. In particular, we focus on using Linked Data technologies to express both the data subjects' consent and the data usage policies of CPSS data controllers in formal terms. This allows for automatically verifying that the usage of personal data complies with data subjects' consent.

1.1.1 What is in this deliverable

This deliverable inspects current approaches to formalize policies. Based on current state of the art, we follow the policy language developed within the EU H2020 SPECIAL[1] project. We show how the SPECIAL policy language [3] can be adapted to cope with general CPSS needs, with particular attention to the considerations emerging from the initial release of the CitySPIN use cases. Note that, given the wide spectrum of CPSS approaches, this deliverable only contains a first draft of such adaptation; an improved version is being developed jointly with use case partners based on their needs.

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

1.1.2 What is *not* in this deliverable

This deliverable does not address the description of business policies and GDPR obligations, which we consider complementary. The description of the policy log is deferred to deliverable *D6.3 Transparency framework*, and the full description of the compliance algorithm will be provided in deliverable *D6.4 Compliance checking components*. In addition, it is worth mentioning that more joint work with the partners of the CitySPIN use cases is needed to stablish the vocabularies to be considered in the policy specifications. Thus, this deliverable considers generic CPSS scenarios to illustrate the potential of our privacy-by-design approach.

1.2 Relation to other Work Packages

WP6 cooperates with most of the other work packages, as its underneath privacy-by-design philosophy can be seen as orthogonal to most of the components developed in CitySPIN. First, the study performed in WP2 (*Designing Cyberphysical Social Systems*), and in particular the social actor model (task T2.2) will be a core input to understand the human role in CPSSs and how the proposed tools can cope with inherent security and privacy concerns. As highlighted above, given the diverse range of CPSS applications and systems, a continuous feedback from our use case requirements in WP3 (*Requirements Elicitation*) will assure that our approach is specific enough to deal with real-world CPSSs. In turn, the proposed tools should be extensible in order to cover the different CPSS scenarios considered in the proof of concepts of WP7 (*PoC Technology Stack Implementation and Evaluation*). Obviously, the scalable Linked Data platform in WP4 (*Scalable Linked Data Integration*) has to integrate the transparency and compliance tools developed in this workpackage, such that all processes, including the analysis and monitoring of real world information in WP5 (*Process Mining and Monitoring on Linked Data*), are subject to our secure secure/privacy-aware data access.

1.3 Deliverable Structure

This deliverable is structured as follows. Section 2 first motivates our work on transparency and compliance in CPSSs, presenting use case scenarios. Then, we review current alternative policy languages and we present the basic components of the SPECIAL framework, the most prominent GDPR-based transparency and compliance architecture. Section 3 analyzes the SPECIAL's usage policy language, which is then extended to cope with practical CPSS needs, in Section 4. Finally, Section 5 concludes and discusses future work of WP6.

2 Personal Data Processing, Transparency and Compliance

The European General Data Protection Regulation (GDPR) defines a set of obligations for controllers and processors of personal data. Primary obligations include obtaining explicit consent from the data subject for the processing of personal data and providing full transparency with respect to processing and sharing.

With the recent application of the GDPR in May 2018, several tools [12, 18, 20] have recently been released to assist companies in assessing their GDPR compliance. However,

such tools are targeted at self assessment (i.e. companies complete standard questionnaires in the form of a privacy impact assessment) and cannot be used to automatically check compliance with usage constraints.

In the following, we first motivate the need of an infrastructure to cope with transparency and compliance in CPSSs (Section 2.1). Then, we present our analysis of current alternative policy languages, logging mechanisms and vocabularies, and GDPR compliance tools (Section 2.2). Finally, we describe the main components of the SPECIAL architecture (Section 2.3), which will be adapted and extended to cope with CPSSs requirements, discussed in the next sections.

2.1 Motivating Use Case Scenarios

In order to exemplify potential GDPR-based transparency and compliance requirements in CPSSs, we present two general scenarios, the *Smart Mobility* and the *Wien Energie* use cases.

Disclaimer: *The following motivating scenarios are made-up, exemplifying use-case descriptions for the project and they do not necessarily reflect the reality of the companies as well as current or future exploitation plans.*

The *Smart Mobility* use case. Doris installs the *WienMobil* APP to look for real-time information about public transport in Vienna. In its first execution, Doris is presented with an *informed consent request* associated to a data usage policy. The policy says that, in order to optimize the public transport infrastructure, the APP will record the history of lookups. These data will be stored in the company servers in EU. Additionally, it asks if the activity history and location data can be integrated with other sources (social media, environment data, traffic congestions) to create an anonymous profile to optimize the public transport infrastructure. Doris accepts this option and starts using the APP. Some time later, Doris updates the APP and she receives a message announcing a new functionality: The user can now provide some real-time feedback, e.g. mentioning if the waiting time was adequate or the wagon was too crowded. To use this functionality, she signs a new consent, stating that this feedback information is aggregated and treated anonymously.

At some point Doris can connect to the WienMobil dashboard and find all the information gathered for her as well as all processing events, verifying that everything was compliant with the given consent, which can be updated or revoked with immediate effects. In turn, the company is able to integrate the provided user data with several data sources and analyse the flow of passengers to further optimize the public transport infrastructure.

The *Wien Energie* use case. Eva is a data scientist at Wien Energie. In her experience, she realized that the network has room for improvement as some stations are underused. Eva is working in a prediction model which can consider multiple sources of data, from sensor and environmental data to social data (city events, demographic data, etc.). Thus, she collects and integrates all the information from multiple sources and partners, together with the associated usage policy, in a semantic data lake. For example, Eva can integrate and create profiles from anonymous location data coming from a third-party partner, but only for the purpose of optimizing the infrastructure (e.g.

marketing is avoided) and only for a period of 1 year after the data is collected. When it comes to running the prediction model and some simulations, the system automatically checks that no usage policy is violated and keeps record of the usage. In addition, Wien Energie can easily prove that all the process complies with the GDPR.

2.2 Formal Policy Languages, GDPR Transparency and Compliance

The previous use cases exemplify that, first and foremost, there is a need of a consent (to process personal data) signed by the data subject and the CPSS data controller. Traditionally, this consent is obtained via a human readable description (i.e. a *contract*, or *terms and conditions*) in some general terms, which prevents automatic consent processing. Instead, formal policy languages are designed to unambiguously represent usage policies, which allows for automatically verifying that the data subjects' consent is enforced. In the following, we briefly summarize the most relevant formal policy languages. Then, we review current proposals on two main tasks mentioned in the use cases (together with consent management): *transparency*, i.e. the CPSS data controller 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.

Formal Policy Languages. When it comes to the formal representation of usage policies there are several potential candidates including semantic policy languages [25, 13, 5, 14] and standard based policy languages [8, 11]. KAoS [25] is a general policy language which adopts a pure ontological approach, whereas Rei [13] and Protune [5] 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. [14] 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. While, the Platform for Privacy Preferences (P3P)[2], is a W3C recommendation, which enables websites to express their privacy preferences in a machine readable format. An more recent W3C recommendation known as the Open Digital Rights Language (ODRL)[3], which was released earlier this year, is a general rights language, which can be used to define rights to 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 more recent standard OWL2, and select language constructs carefully in order to achieve an optimal tradeoff between expressiveness and computational complexity. This is the main objective of the SPECIAL policy language [3], developed within the EU H2020 SPECIAL project. An analysis of the policy language and the adaptation to CPSS needs is provided in the next section.

Transparency. As for transparency with respect to data processing, relevant work primarily relates to the re-purposing of existing logging mechanisms as the basis for personal data processing transparency and compliance [4]. Many of the works use a secret key signing scheme based on Message Authentication Codes (MACs) together with a hashing

[2]P3P,http://www.w3.org/TR/P3P/
[3]ODRL,https://www.w3.org/TR/odrl-model/

algorithm to generate chains of log records that are in turn used to ensure log confidentiality and integrity [2] (cf. [4] for a summary of existing approaches). MACs are themselves symmetric keys that are generated and verified using collision-resistant secure cryptographic hash functions. However, only a few works [22, 23] 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. [28]. 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 [24] propose an extension of blockchain technology with *Linked Data* to create tamper-proof audit logs and non-repudiation. Nonetheless, very little research has been conducted into transparency requirements and performance/scalability issues of such blockchain-based solutions.

The description of the transparency mechanisms for CPSSs is deferred to deliverable *D6.3 Transparency framework*, hence it is out of scope of this deliverable.

Compliance. As for GDPR compliance, recently the Information Commissioner's Office (ICO) in the UK [12], Microsoft [18], and Nymity [20] have developed compliance tools that enable companies to assess the compliance of their applications and business processes by completing a predefined questionnaire. Recent works also look at the challenges of representing GDPR concepts and obligations [21] as well as informed consent [9]. The management of events for business process compliance monitoring and process mining [16] can be seen as orthogonal work.

In contrast to existing approaches, in this deliverable 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. The concrete compliance algorithm will be addressed in deliverable *D6.4 Compliance checking components*.

2.3 SPECIAL Consent, Transparency and Compliance Framework

When it comes to personal data processing and GDPR, CitySPIN follows the SPECIAL consent, transparency and compliance framework, adapting and extending their components to CPSSs. 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 usage policy. CitySPIN currently considers minor modifications to this component, which can be required in the proof of concepts of WP7 (*PoC Technology Stack Implementation and Evaluation*).

(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 usage control policies. As discussed above, this deliverable focuses on the policy language formalization, while the transparency framework and the compliance algorithms are deferred to future deliverables.

(iii) the *SPECIAL Middleware* includes sub-components that connect the SPECIAL primary components with existing Line of Business access control mechanisms and business logic, and middleware that enables companies to perform policy aware business

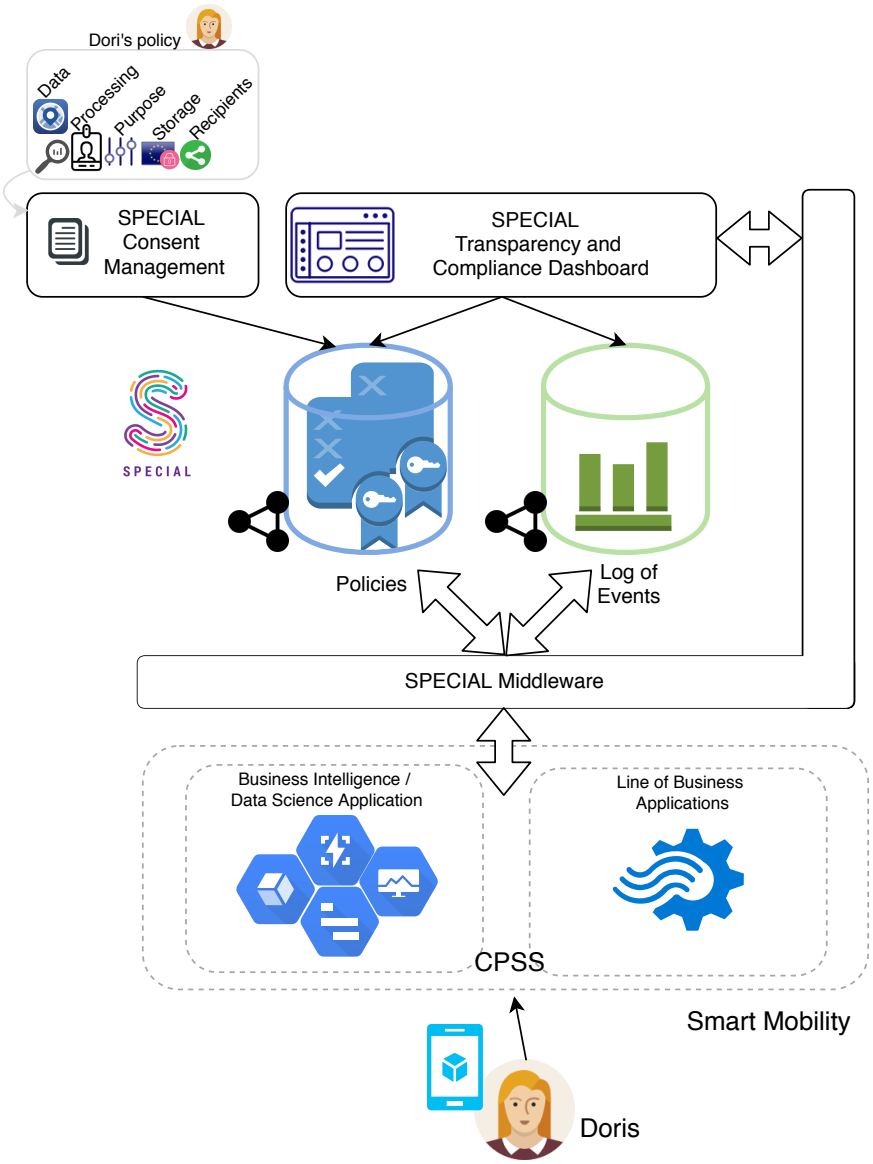

Figure 1: The SPECIAL Consent, Transparency and Compliance framework

intelligence and data science. Specific CPSS-based components are required in the course of the scalable Linked Data platform in WP4 and the proof of concepts of WP7.

As shown in Figure 1, in addition to existing data sources that support business operations (i.e. *Line of Business Applications*), and strategic decision making (i.e. *Business Intelligence / Data Science Applications*), SPECIAL proposes two additional data sources referred to as *Policies*, used to store the consent, regulatory and business policies, and *Events*, used to store (i.e. log) the data processing or sharing events. In this deliverable, we focus specifically on the former, providing additional details on the required adaptations of the SPECIAL policy language and auxiliary vocabularies in the next Section.

3 The SPECIAL's Usage Policy Language

In this section we provide a comprehensive review of the SPECIAL usage policy language, which will be analyzed and extended in a practical CPSS scenario in the next section.

SPECIAL usage policies are encoded in OWL 2 [19]. In the examples[4] that follow, the `spl` prefix represents `http://www.specialprivacy.eu/langs/usage-policy#`. Additional details, including the full grammar of policy expressions in Backus normal form (BNF), can be found in the SPECIAL documentation [3].

3.1 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, and whether or not the data will be shared with others. Usage policies can consist then of the following five elements, referred to as the *minimum core model* (MCM), depicted in Figure 2.

- "Data" describes the personal data collected from the data subject. In order to describe which categories of data are collected, an ontology of *personal data* is needed to cover the most common data categories. It is envisaged that the ontology will be extended with suitable profiles and/or integrated with further use case specific ontologies.
- "Processing" describes the operations that are performed on the personal data. Data processing should be described through a suitable ontology of data operations.
- "Purpose" specifies the objective that is associated with data processing. Objectives such as marketing, service optimisation and personalisation, scientific research, are pervasive across a variety of contexts. Purpose descriptions are part of most usage policy languages developed so far (e.g. P3P [8] and ODRL [11]).
- "Storage" specifies where data are stored and for how long. Note that the GDPR requires that storage is strictly bound to the service needs. This implies storage minimisation, hence the need to express *upper bounds* to storage duration, that may be expressed either in terms of the duration of the service that the data have been collected for, or in absolute terms.
- "Recipients" specifies who is going to receive the results of data processing and, as a special case, whom data are shared with. The GDPR does not clearly state to which level of detail this information has to be specified, and there are potentially conflicting needs, such as the companies' desire to keep some of their business relations confidential, and the data subjects' right to trace the flow of their personal information.

Table 1 provides a high level overview of the initial vocabularies that are necessary to represent the elements of the MCM. All namespaces share the `S` which represents `http://www.specialprivacy.eu/`. Note that these vocabularies can be extended to support CPSS scenarios and, in particular, the CitySPIN use cases.

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

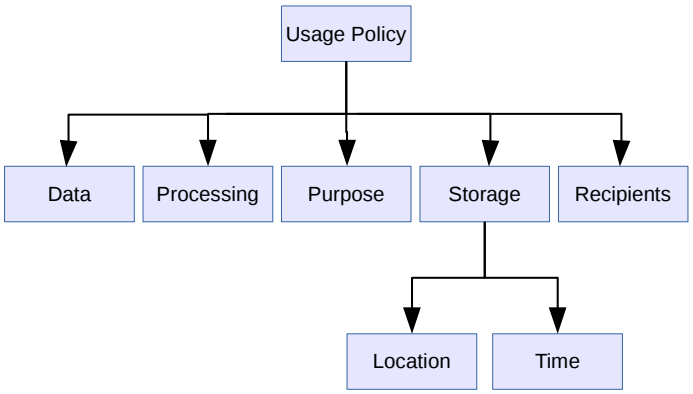

Figure 2: The minimum core model (MCM) for usage policies.

Table 1: SPECIAL auxiliary vocabularies for usage policies.

Category	Namespace	#Classes	Examples	Superclass
Data	svd:=(S)/vocabs/data	27	svd:Activity, svd:Anonymized, svd:Financial, svd:Health, svd:Location, svd:Navigation, svd:Preference, svd:Profile, etc.	spl:AnyData
Processing	svpr:=(S)/vocabs/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)/vocabs/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)/vocabs/recipients	6	svr:Delivery, svr:OtherRecipient, svr:Ours, svr:Public, svr:Same, svr:Unrelated	spl:AnyRecipient
Storage location	svl:=(S)/vocabs/locations	7	svl:ControllerServer, svl:EU, svl:EULike, svl:ThirdCountries, svl:OurServers, svl:ProcessorServers, svl:ThirdParty	spl:AnyLocation
Storage duration	svdu:=(S)/vocabs/duration	4	svdu:BusinessPractices, svdu:Indefinitely, svdu:LegalRequirement, svdu:StatedPurpose	spl:AnyDuration

3.2 Basic Usage Policies

A usage policy is composed of one or more *basic usage policies*, each of which is an OWL 2 expression of the form:

```
ObjectIntersectionOf(                                              (1)
   ObjectSomeValuesFrom(spl:hasData SomeDataCategory)
   ObjectSomeValuesFrom(spl:hasProcessing SomeProcessing)
   ObjectSomeValuesFrom(spl:hasPurpose SomePurpose)
   ObjectSomeValuesFrom(spl:hasRecipient SomeRecipient)
   ObjectSomeValuesFrom(spl:hasStorage SomeStorage)
)
```

The important parts in this expression are the policy's attributes highlighted in bold. The policy author needs to decide for each of them a suitable range, that in the above text is highlighted in italics. The example presented authorizes all operations that:

1. fall within the specified *SomeProcessing* category,
2. operate only on data that belong to *SomeDataCategory*,
3. have any purpose covered by the *SomePurpose* category,
4. disclose the results to any member(s) of the *SomeRecipient* category, and

5. store the results in any place belonging to the *SomeStorage* category.

Therefore, policy (1) encodes the set of all authorizations that have (at least) the specified attributes, which match the minimum core model (MCM), introduced in the previous section. Although SPECIAL defines auxiliary vocabularies providing a set of classes for *SomeDataCategory, SomeProcessing, SomePurpose, SomeRecipient*, it should be noted that it is not possible to enumerate over all possible classes and as such the policy language and by extension the vocabularies were designed to be extensible. CitySPIN builds upon this extensibility.

3.3 General Usage Policies

A general usage policy may contain a union of any number of basic policies, each of them of the form (1). The resulting policy is conceptually the *union* of all the authorizations supported by the basic policies, that is, an operation is authorized by the general policy if and only if the operation is authorized by *at least one* of its basic policies.

For instance, the following *general usage policy* states that personal data can only be used for non-commercial purposes and shall neither be stored nor disclosed to third parties, while pseudonymised data can be used freely (where auxiliary vocabularies define the terms *PersonalData, NonCommercial, PseudonymizedData*):

```
ObjectUnionOf(                                                    (2)
  ObjectIntersectionOf(
    ObjectSomeValuesFrom(spl:hasData PersonalData)
    ObjectSomeValuesFrom(spl:hasProcessing spl:AnyProcessing)
    ObjectSomeValuesFrom(spl:hasPurpose NonCommercial)
    ObjectSomeValuesFrom(spl:hasRecipient spl:Null)
    ObjectSomeValuesFrom(spl:hasStorage spl:Null)
  )
  ObjectIntersectionOf(
    ObjectSomeValuesFrom(spl:hasData PseudonymizedData)
    ObjectSomeValuesFrom(spl:hasProcessing spl:AnyProcessing)
    ObjectSomeValuesFrom(spl:hasPurpose spl:AnyPurpose)
    ObjectSomeValuesFrom(spl:hasRecipient spl:AnyRecipient)
    ObjectSomeValuesFrom(spl:hasStorage spl:AnyStorage)
  )
)
```

The `hasStorage` policy attribute is a structured object itself, with two attributes, and is specified as follows:

```
ObjectIntersectionOf(                                             (3)
   ObjectSomeValuesFrom(spl:hasLocation SomeLocation)
   ObjectSomeValuesFrom(spl:hasDuration SomeDuration)
   DataSomeValuesFrom(spl:durationInDays Interval)
)
```

where *SomeLocation* and *SomeDuration* are expressed in terms of the corresponding storage location and duration auxiliary vocabularies. It is not necessary to include in the policy both `hasDuration` and `durationInDays`. However at least one of them should be specified. The *Interval* limiting storage duration in days is expressed with the integer *facets* of OWL 2, that is:

```
DatatypeRestriction( xsd:integer                                  (4)
   xsd:minInclusive min duration                    (optional)
   xsd:maxInclusive max duration                    (optional)
)
```

4 Practical use in a CPSS Scenario

The SPECIAL policy language presented in Section 3 can be extended to cope with practical CPSS scenarios. In this section we illustrate some examples, we identify extension points and we provide a methodology to approach the formalization of data subjects' consent and data usage policies in a CPSS scenario.

4.1 Using the Policy Language to Express Policies

Following from the use case scenarios presented in Section 2.1, we show, in Example 4.1, how Dori's policy would look like in the SPECIAL policy language.

Example 4.1. The following policy: *"(...)the activity history and location data can be integrated with other sources (social media, environment data, traffic congestions) to create an anonymous profile to optimize the public transport infrastructure"* can be formalized with a factorized general policy, which also considers that profiles are stored indefinitely in the EU only by the data controller, as follows:

```
ObjectIntersectionOf(
  ObjectSomeValueFrom( spl:hasData
    ObjectUnionOf(
      smartMobility:History svd:Location ))
  ObjectSomeValueFrom( spl:hasProcessing
    ObjectUnionOf(
      smartMobility:Profiling svpr:Anonymize smartMobility:Integration ))
  ObjectSomeValueFrom( spl:hasPurpose smartMobility:Optimization )
  ObjectSomeValueFrom( spl:hasStorage
    ObjectIntersectionOf(
      ObjectSomeValuesFrom( spl:hasLocation
        ObjectIntersectionOf( svl:OurServers svl:EU ))
      DataSomeValuesFrom(   spl:durationInDays
        DatatypeRestriction( xsd:integer
          xsd:mininclusive "0"^^xsd:integer ))))
  ObjectSomeValueFrom( spl:hasRecipient svr:Ours )
)
```

□

In this example, the auxiliary SPECIAL vocabularies need to be extended with four new classes:

- `smartMobility:History`, i.e. the class representing the history of lookups in the WienMobil APP, which extends the general SPECIAL class `svd:Activity`;
- `smartMobility:Profiling`, i.e. the class that represents all processes to create a user profile by analysing the location and history of lookups of the user, which extends the SPECIAL class `svpr:Analyze`;
- `smartMobility:Integration`, i.e the class representing the integration of user profiles with additional data sources, which is a subclass of the general SPECIAL class `spl:AnyProcessing`;
- `smartMobility:Optimization`, i.e. the class representing the optimization of the public transport infrastructure, which extends the SPECIAL class `svpu:Develop`.

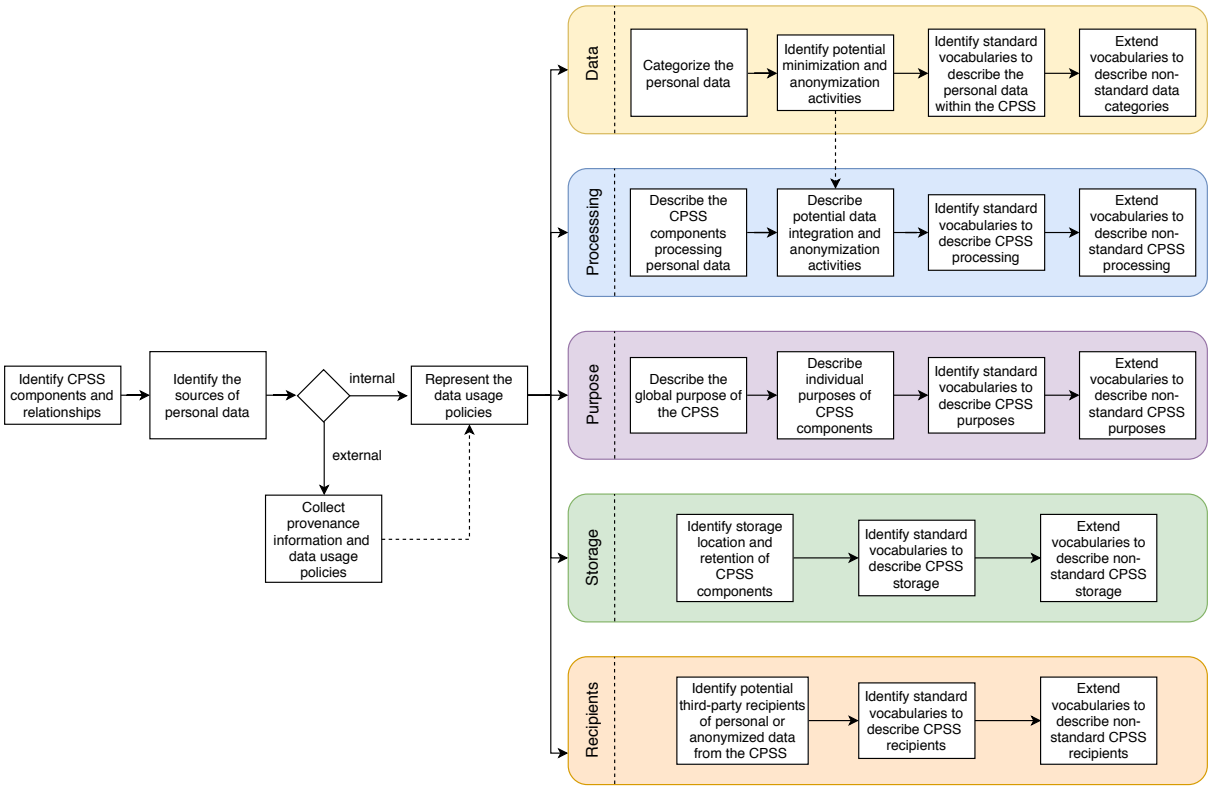

Figure 3: Practical methodology to define CPSS data subjects' consent and data usage policies.

As illustrated in the previous example, the SPECIAL usage policy and, in particular, its auxiliary vocabularies have to be extended to cope with the particular needs of CPSSs. Given the nature of CPSSs, defining these extensions and the concrete usage policies could involve different activities and CPSS "components". In the following, we describe a practical methodology to define CPSS data subjects' consent and data usage policies.

4.2 Practical methodology to define CPSS data subjects' and data usage policies

Figure 3 shows the main steps of the methodology we establish in CitySPIN in order to analyze a CPSS and to establish CPSS data usage policies. Then, these policies (i) set the basis to ask for data subjects' consents, and (ii) they shall be integrated in those CPSS components processing personal data in order to assure automatic compliance. The sequence of steps, described below, are designed taking into account the SPECIAL usage policy model (cf. see Section 3) and the general guidelines of the privacy by design [7] philosophy.

- *Identify CPSS components and relationships.* In a first phase, all CPSS components managing data should be identified, as well as the different relationships among them. CPSS are often complex systems composed of components of diverse nature [27], 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 must be paid to the description of inputs and user feedback loops, a key aspect in CPSSs.

- *Identify the sources of personal data.* Once the components and their relationships are clearly described, this phase regards the identification of all sources of personal data. It is worth mentioning that the concept of *personal data* shall mean, according to GDPR, *"any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"*. This phase is of particular importance given that most CPSSs consider different input sources, with a strong human/social component.

  At this point, we must categorize the sources of personal data in two categories, *external* or *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 of the current 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 must be represented (described below) and the appropriate data subjects' consent associated to the data must be obtained.

- *Collect provenance information and data usage policies.* When it comes to external personal data, the CPSS should be able to keep track of the origin of the data and the associated data usage policy. In this phase, all provenance information (data sources, third-party contracts and terms, etc.) and data usage policies are collected. Ideally, these data usage policies are already represented using the SPECIAL policy language. In this case, the CPSS should link the provenance information with the policies and the concrete data that adhere to such policies. To the best of our knowledge, there is a lack of standard vocabularies and guidelines to perform these links using SPECIAL usage policies. We show a practical illustration of a potential modelling in Example 4.2. The formalization of this process for CPSSs swill be addressed in collaboration with CitySPIN partners, and it is deferred to future work.

  Example 4.2. Listing 1 shows an example (in RDF/Turtle syntax) of linking data usage policies and personal data with provenance information (using the well-known PROV [15] vocabulary), managed in different named graphs. Following from the use case scenario presented in Section 2.1, smartMobility gathers city events from a third-party partner, referred to as `socialMediaCompany`. SmartMobility keeps track of data collection processes in the named graph `smartMobility:collectExternalData`. The metadata of these processes (e.g. `smartMobility:collection-01062018`) (i) links to the collected data (in this case, `socialMediaCompany:cityEvents-01062018`), which can be retrieved in the corresponding named graph, and (ii) points to the data usage policy under which the data was collected (e.g., `smartMobility:UsagePolicy123`). Data usage policies can be then specified separately using the SPECIAL language.

  Note that the link between data collection processes and data usage policies is performed with a non-standard `smartMobility:policy` property. Thus, as mentioned

Listing 1: Example of linking external personal data with data usage policies using provenance information in named graphs

```
# The default graph may include metadata about the graphs

socialMediaCompany:cityEvents prov:agent socialMediaCompany:Us .
smartMobility:collectExternalData prov:agent smartMobility:Us .
smartMobility:policies prov:agent smartMobility:Us .

# The following graph encodes city events information gathered from socialMediaCompany

socialMediaCompany:cityEvents {

  socialMediaCompany:cityEvents-01062018 a svd:Location, socialMediaCompany:events ;
     prov:wasAssociatedWith socialMediaCompany:anonymousUser ;
     prov:atTime  "2018-06-01T14:32:05Z"^^xsd:dateTimeStamp;
     rdfs:label   "City events registered in social media..." ;
     # ... other descriptions ...
}

# This graph encodes the collection of the socialMediaCompany data by smartMobility

smartMobility:collectExternalData{

  smartMobility:collection-01062018 a smartMobility:DataCollection;
     skos:member socialMediaCompany:cityEvents-01062018;
     prov:agent smartMobility:Us;
     smartMobility:policy smartMobility:UsagePolicy123 ;
     # ... other metadata ...
}

# This graph encodes the usage policies of smartMobility

smartMobility:policies{

 smartMobility:UsagePolicy123 a spl:Authorization;
    # ... Description of the usage policy using the SPECIAL vocabulary ...

}
```

above, there is a need of standard guidelines and vocabularies to bind concrete collected data to data usage policies, which is deferred to future work. □

- *Represent the data usage policies.* As depicted in the methodology workflow in Figure 3, the data usage policies must be represented following from (a) the internal personal data generated by the CPSS, which will be then the basis to ask for the appropriate data subjects' consent to manage such data, or (b) the external personal data, 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 formal terms by the data source provider.

  When it comes to representing CPSS data usage policies using the SPECIAL model, Figure 3 depicts different processes that can help in the identification and representation of the five elements (data, processing, purpose, storage and recipients) of the minimum core model (cf. see Section 3), summarized below.

  - The element '*Data*' describes the personal data collected from the data subject. First, the already identified CPSS elements and data sources must be further analyzed to categorize such data. In this step, rather than the actual data, the domain of the data and the potential *skeleton* (i.e. structure) of typical data items should be identified. For example, in our previous *smartMobility* example, the analysis would reveal that the CPSS needs to store location data, consisting of GPS locations of the user, and a history of lookups in the APP, which is basically a log of user queries to the API.

    In a second step, following the privacy by design [7] philosophy, potential min-

imization and anonymization activities shall be identified. Data minimization plays an important role in GDPR given that companies must limit personal data collection, storage, and usage to data that is absolutely necessary for carrying out the purpose for which the data is processed. In turn, 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"*. Thus, this step is the place where we can identify (i) whether data minimization is applied or can be applied in the CPSS, reducing the amount of detail linking to a particular person, and (ii) whether anonymization, pseudonymization and aggregation of data (e.g. applying different techniques such as k-anonymity or l-diversity [10, 1]) is applied or can be applied in the CPSS.

Then, the category of the final relevant personal data in the CPSS (i.e. data that cannot be further minimized or 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 vocabularies if needed. As mentioned before, this step completely depends on the particular scenario, hence it is expected that the initial vocabulary presented in SPECIAL (cf. see Table 1) needs to be extended with use case specific ontologies.

Providing a full list of potential existing vocabularies is out of scope of this deliverable, as it would never be complete given the wide range of domains[5] and use cases present in CPSSs. In future work, we plan to analyze and extend vocabularies for the categories of personal data involved in the CitySPIN use cases. We also expect to complete the categories of personal data with the CitySPIN's own Social Actor Model (emerging from task T2.2 of WP2), which will capture key characteristics of the social actors involved in CPSSs.

For exemplary purposes, we recall and analyse the examples of collected social data in CPSSs emerging from our study in *Deliverable 2.1: Cyber-Physical Social Systems Blueprint (v.1)* [17]:

* Ambient condition, i.e. information about the surrounding environment such as traffic accidents and congestion, weather condition, air quality. This constitutes a good example of potential sources of anonymization: the CPSS company must study if the particular ambient condition absolutely needs to be linked to the subject for the service in place, or aggregated, anonymous data can be equally effective. In any case, if such data is directly collected from the subject, the consent must specify the category of the ambient condition data. Currently, this category is underrepresented in the auxiliary vocabularies of SPECIAL, hence this opens up potential extension points. Some examples of existing ontologies falling in this category are: *Road Traffic Management*[6] (an ontology to describe the management of the traffic), *Home Weather*[7] (smart home ontology for weather phenomena and exterior conditions), and *Home Activity*[8] (an ontology to detect activity in a smart home).

[5] Our analysis in D2.1 [17] lists the most cited CPSS domains, from smart transportation and smart building to smart city, smart factory and other dynamic social systems.

[6] http://www.sensormeasurement.appspot.com/ont/transport/traffic

[7] https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl

[8] http://sensormeasurement.appspot.com/ont/home/homeActivity#

  * Service ratings, user's emotional and physiological state, i.e. information to reflect the user's satisfaction with a 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* [3]. A CPSS company might need to provide further details on the information collected. In this case, the `svd:preference` class should be extended to cope with the particular needs. Potential existing ontologies are: *Review Vocabulary*[9] (a vocabulary for expressing reviews and ratings) and *RECommendations Ontology*[10] (a vocabulary for representing preferences-as-constraints and preferences-as-ratings).
  * Online activities, i.e. user's activities on social networks like check-ins, posts (text, audio, video and pictures), tweets on Twitter and accepted social events. This data category is covered by the SPECIAL categories `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.* [3]. Although these categories should cover most scenarios, fine-grained, company-specific categories can be provided. For instance, in Example 4.1, the class `smartMobility:History` extended `svd:Activity` to represent the history of lookups in the WienMobil APP.
  * Physical behaviors and activities, i.e. information about human behavior during evacuation processes, workers activities in an assembly line and pedestrians' paths. This category can also be covered by the SPECIAL category `svd:activity`, and its subcategory `svd:physical-activity`, considering *data describing the activities of a user in the real world, e.g. travels, sports, concerts, etc.* [3]. Similarly to the previous case, fine-grained, company-specific categories could be provided. In addition, note that gathering physical activities might require other information such as location data, which should be specifically noted in the data usage policy.
  * Quantifiable information, i.e. complaints and requests from citizens to call centers, subway passenger records, driving speeds on the roads, etc. This example of collected social data in CPSSs is rather general and might overlap with previous considerations on service ratings, online and physical activities. Nonetheless, and depending on the use case, other SPECIAL categories can be considered, such as `svd:statistical` (statistical data and analytics which make use of the user data) and `svd:derived` (any additional data produced by processing/combining data directly provided by the user).

- *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 to describe the CPSS components processing personal data. In particular, and emerging from the previous 'data' phase (as represented with a dashed arrow in Figure 3), special attention shall be paid to describing potential data integration and anonymization activities, which could affect the specification of the data usage policies. Once all components and activities are identified, similarly to the previous case, standard vocabularies to

[9] http://purl.org/stuff/rev#
[10] http://purl.org/reco#

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 would potentially 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:Anonymize`, `svpr:Collect`, etc. In the following, we review the most important CPSS stages/activities emerging from our study in *Deliverable 2.1: Cyber-Physical Social Systems Blueprint (v.1)* [17], mapped to SPECIAL concepts when possible (where further fine-grained, application-specific extensions are always possible):

  * Data collection - can be directly mapped to the SPECIAL `svpr:Collect` concept.
  * Data analysis - can be directly mapped to the SPECIAL `svpr:Analyze` concept.
  * Proactive recommendations. In this case, SPECIAL considers that a recommendation is rather a 'purpose' (described below), and it can be seen as a subtype of `svpu:Marketing`. Then, the 'processing' leading to a recommendation could be seen as a subtype of `svpr:Analyze`.
  * 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 (as shown in Example 4.1).
  * Data storage - can be directly mapped to the SPECIAL `svpr:Archive` concept[11].

- *Purpose.* The element '*Purpose*' specifies the objective that is associated with data processing. In CPSSs, we could establish a two-phase identification of (a) the global purpose of the CPSS, and (b) individual purposes of the CPSS components. The rationale behind this approach is that CPSSs often involve complex components and relationships that might be re-purposed for a specific goal. Thus, we must clearly distinguish between the final objective of the CPSS for the data subject, and the potential individual purpose of CPSS components, which might also require consent by the data subject. Once these purposes are identified, standard vocabularies, or extensions, to describe CPSS purposes must be put in place. Similarly to previous elements, there is a wide spectrum of purposes in CPSSs, which depend on the particular scenarios. In the following, for exemplary purposes, we review the CPSS goals identified in *Deliverable 2.1: Cyber-Physical Social Systems Blueprint (v.1)* [17] and how these can be mapped to SPECIAL vocabularies:

  * Smart transportation - driving support: the goal is to provide the drivers with proactive recommendations on the road. In this case, we identify two potential candidate concepts in the SPECIAL vocabularies: (a) `svpu:Current` (i.e. completion and support of activity for which data was provided), which could be used as a general concept if the main goal of the data collection and the CPSS is the driving support, and (b) `svpu:Feedback` (i.e. responding to user), if the system is mainly aimed at responding to user input.

[11] Note that this concept is not explicitly present in the SPECIAL ontology, but the documentation [3] suggests that it is possible to use it in accordance with the ODRL [11] 'archive' operation.

In contrast, the concepts of *transportation* and *recommendation*[12] are underrepresented in SPECIAL, hence additional application-specific vocabularies could be needed if fine-grained details are required. In this context, it is worth mentioning that the P3P[13] vocabulary provides two interesting classes that could be used in this scenario: `individual-decision`, i.e. *to determine the habits, interests, or other characteristics of individuals and combine it with identified data to make a decision that directly affects that individual*, and `pseudo-decision`, with the same purpose but data will not be used to attempt to identify specific individuals. Note, however, that the SPECIAL vocabularies specifically discarded these concepts because they do not specify which purpose that decision is related to [17]. Thus, CPSS application-specific vocabularies could base on these P3P concepts, providing more specific concepts for each use case.

* Smart building - emergency evacuation: the CPSS is focused on effectively guiding the building occupants to safe ground in a timely fashion. Similarly to the previous case, the SPECIAL `svpu:Current` concept could be the general candidate if the data is mainly collected for this purpose. In any case, given the vital importance of the emergency use case, we identify this area as a potential extension point to represent data usage policies. Vocabularies such as `incident`[14] (vocabulary to describe incident response by emergency services) and `moac`[15] (Management of a Crisis Vocabulary) are potential candidates for the extension.
* Smart building - smart power management: it focuses on efficient use of energy. This scenario, which is similar to the motivational use case presented in Section 2.1, is also underrepresented in SPECIAL, with the exception of the aforementioned general `svpu:Current` concept. Several vocabularies, such as `reegle`[16] (Renewable Energy and Energy Efficiency) and `EEPSA`[17] (Energy Efficiency Prediction Semantic Assistant Ontology) are good candidates to represent such purposes.
* Smart factory - production automation, better consumer-enterprise and improving human-machine interacting: the focus is to reduce work demands and improve the customer's ability to reach the enterprises, to receive news and updates, and to receive advice on products. These broad goals can be addressed by multiple SPECIAL concepts, such as `svpu:Develop` (enhance, evaluate, or otherwise review the site, service, product, or market), `svpu:Tailoring` (the information is used to modify the content of the 'site' and not used for any kind of future customization) and `svpu:Custom` (customize the user's online experience as explicitly requested by the user), as well as the aforementioned `svpu:Current` and `svpu:Feedback`. These concepts can cover most of the potential use cases, while further extensions could be provided to focus on very specific smart factory applications.
* Smart city - crowdsensing, supporting mobility logistics and helping people

[12] Note that the concept of recommendation is partially covered under `svpu:Marketing`, but it does not fit the main idea behind driving recommendations.
[13] http://www.w3.org/TR/P3P11/#PURPOSE
[14] http://vocab.resc.info/incident
[15] http://www.observedchange.com/moac/ns#
[16] http://reegle.info/schema
[17] https://w3id.org/eepsa

with disabilities: multiple goals are identifies within the smart city area, such as improving the navigation for visually impaired people or for urban pedestrian flows, guidance in shopping malls and for evacuation, monitoring medical patients behaviour and current state, etc. Similarly to the previous case, smart city is a broad term that opens up a wide range of possibilities. Besides the already mentioned SPECIAL concepts in previous categories, we also highlight the SPECIAL concepts `svpu:Government` (online government services such as voter registration, vehicle registration, and citizen information services) and `svpu:Health` (offer the user products or services that relate to their physical and/or mental health).

  * Other dynamic social systems, multiple objects interaction, smart office, network navigability: the goal is to interact with multiple objects at the same time in order to combine their capabilities and reach a new user experience. In this case, we identify the interaction of multiple objects as the 'technique', being the new user experience the main goal from the point of view of the usage policy. Previous considerations on recommendation and feedback mechanisms apply in this scenario.

- *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. see Section 3). 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 [3]. In any case, the particular CPSS use case could extend such vocabulary with classes corresponding to countries and other geographic areas, e.g. using the *FAO Geopolitical Ontology*[18].
- *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 or anonymized data from the CPSS should be identified. Given the inherent complexity of CPSSs, this step might imply to carefully inspect all (potentially distributed) CPSS components, involved partners and stakeholders. Then, as in previous elements, standard vocabularies to describe CPSS recipients must be analyzed, extending them when 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 might need some extensions, e.g. using the FOAF [6] and PROV [15] vocabularies.

[18] http://www.fao.org/countryprofiles/geoinfo/en/

5 Summary

Privacy protection is a fundamental but challenging requirement in the context of Cyber-Physical Social Systems (CPSSs). This deliverable focuses on establishing a formal policy language to express data subjects' consent and data usage policies in CPSSs. In particular, we consider the new EU General Data Protection Regulation (GDPR) as the baseline, and we inspect the policy language developed within the EU H2020 SPECIAL project from the perspective of CPSSs. To that extend, we establish two use case scenarios and we propose a novel practical methodology to define CPSS data subjects' consent and data usage policies in formal terms. Finally, we provide concrete examples of the application of the methodology based on the initial outcomes emerging from our first CPSS blueprint.

Our ongoing work focuses on collaborating with the CitySPIN use case partners to further elaborate the identified policy language extensions in order to cope with CPSS scenarios. In addition, we plan to work on the definition of the log of data processing and sharing events that, together with the aforementioned policy language, will allow for automated compliance checking.

References

[1] Charu C Aggarwal and S Yu Philip. A general survey of privacy-preserving data mining models and algorithms. In *Privacy-preserving data mining*, pages 11–52. Springer, 2008.

[2] Mihir Bellare and Bennet Yee. Forward integrity for secure audit logs. Technical report, Computer Science and Engineering Department, University of California at San Diego, 1997.

[3] P.A. Bonatti, S. Kirrane, I. Petrova, L. Sauro, and E. Schlehahn. Special deliverable 2.1: Policy language v1, 2017.

[4] Piero Bonatti, Sabrina Kirrane, Axel Polleres, and Rigo Wenning. Transparent personal data processing: The road ahead. In *Proc. of TELERISE*, pages 337–349, 2017.

[5] Piero A. Bonatti, Juri Luca De Coi, Daniel Olmedilla, and Luigi Sauro. A rule-based trust negotiation system. *IEEE Trans. Knowl. Data Eng.*, 22(11):1507–1520, 2010.

[6] Dan Brickley and Libby Miller. Foaf vocabulary specification 0.91, 2010.

[7] Ann Cavoukian. Privacy by design in law, policy and practice. *A white paper for regulators, decision-makers and policy-makers*, 2011.

[8] Lorrie Faith Cranor. *Web privacy with P3P - the platform for privacy preferences*. O'Reilly, 2002.

[9] Kaniz Fatema, Ensar Hadziselimovic, Harshvardhan J. Pandit, Christophe Debruyne, Dave Lewis, and Declan O'Sullivan. Compliance through informed consent: Semantic based consent permission and data management model. In *Proc of PrivOn*, 2017.

[10] Gabriel Ghinita, Panagiotis Karras, Panos Kalnis, and Nikos Mamoulis. Fast data anonymization with low information loss. In *Proceedings of the 33rd international conference on Very large data bases*, pages 758–769. VLDB Endowment, 2007.

[11] Renato Iannella and Serena Villata. Odrl information model 2.2. W3C Recommendation, 2018.

[12] Information Commissioner's Office (ICO) UK. Getting ready for the GDPR, 2017.

[13] Lalana Kagal, Timothy W. Finin, and Anupam Joshi. A policy language for a pervasive computing environment. In *Proc. of POLICY*, pages 63–, 2003.

[14] Vladimir Kolovski, James Hendler, and Bijan Parsia. Analyzing web access control policies. In *Proc. of WWW*, pages 677–686, 2007.

[15] T. Lebo, S. Sahoo, and D. McGuinness. Prov-o: The prov ontology. *W3C Recommendation, April*, 2013.

[16] Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil MP van der Aalst. Compliance monitoring in business processes: Functionalities, application, and tool-support. *Information systems*, 54:209–234, 2015.

[17] Angelika Musil Marta Sabou. Cityspin deliverable 2.1: Cyber-physical social systems blueprint (v.1), 2018.

[18] Microsoft Trust Center. Detailed GDPR Assessment, 2017.

[19] Boris Motik, Peter F. Patel-Schneider, and Bijan Parsia. OWL 2 Web Ontology Language – Structural Specification and Functional-Style Syntax (Second Edition). W3C Recommendation, 2012.

[20] Nymity. GDPR Compliance Toolkit.

[21] H.J. Pandit and D Lewis. Modelling provenance for gdpr compliance using linked open data vocabularies. In *Proc of PrivOn*, 2017.

[22] Tobias Pulls, Roel Peeters, and Karel Wouters. Distributed privacy-preserving transparency logging. In *Proc. o WPES*, 2013.

[23] Stefan Sackmann, Jens Strüker, and Rafael Accorsi. Personalization in privacy-aware highly dynamic systems. *Communications of the ACM*, 49(9), 2006.

[24] Andrew Sutton and Reza Samavi. Blockchain enabled privacy audit logs. In *Proc. of ISWC*, pages 645–660, 2017.

[25] Andrzej Uszok, Jeffrey M. Bradshaw, Renia Jeffers, Niranjan Suri, Patrick J. Hayes, Maggie R. Breedy, Larry Bunch, Matt Johnson, Shriniwas Kulkarni, and James Lott. KAoS policy and domain services: Towards a description-logic approach to policy representation, deconfliction, and enforcement. In *Proc. of POLICY*, pages 93–96, 2003.

[26] Fei-Yue Wang. The emergence of intelligent enterprises: From cps to cpss. *IEEE Intelligent Systems*, 25(4):85–88, 2010.

[27] Gang Xiong, Fenghua Zhu, Xiwei Liu, Xisong Dong, Wuling Huang, Songhang Chen, and Kai Zhao. Cyber-physical-social system in intelligent transportation. *IEEE/-CAA Journal of Automatica Sinica*, 2(3):320–333, 2015.

[28] Guy Zyskind, Oz Nathan, et al. Decentralizing privacy: Using blockchain to protect personal data. In *Proc. of SPW*, pages 180–184, 2015.