SPECIAL

Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance

Deliverable D2.6

Formal representation of the legislation V2

Document version: V1.0

SPECIAL DELIVERABLE

Name, title and organisation of the scientific representative of the project's coordinator:

Ms Jessica Michel t: +33 4 92 38 50 89 f: +33 4 92 38 78 22 e: jessica.michel@ercim.eu

GEIE ERCIM, 2004, route des Lucioles, Sophia Antipolis, 06410 Biot, France

Project website address: http://www.specialprivacy.eu/

Project	
Grant Agreement number	731601
Project acronym:	SPECIAL
Project title:	Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance
Funding Scheme:	Research & Innovation Action (RIA)
Date of latest version of DoW against which the assessment will be made:	17/10/2016
Document	
Period covered:	M1-M23
Deliverable number:	D2.6
Deliverable title	Formal representation of the legislation V2
Contractual Date of Delivery:	30-09-2018
Actual Date of Delivery:	31-12-2018
Editor (s):	Piero Bonatti (CeRICT), Sabrina Kiranne (WU)
Author (s):	P.A. Bonatti, S. Kirrane, I. Petrova, L. Sauro, C. Kerschbaum, Eliska Pirkova
Reviewer (s):	Eva Schlehahn (ULD), Rigo Wenning (ERCIM/W3C)
Participant(s):	WU, CeRICT, ERCIM, ULD
Work package no.:	2
Work package title:	Policy and Transparency Framework
Work package leader:	WU
Distribution:	PU
Version/Revision:	1.0
Draft/Final:	Final
Total number of pages (including cover):	40

Disclaimer

This document contains description of the SPECIAL project work and findings.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any responsibility for actions that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of the SPECIAL consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 28 Member States of the Union. It is based on the European Communities and the Member States cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu/).

SPECIAL has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731601.

Contents

List of Figures

1 Summary

One of the primary goals of the SPECIAL H2020 ICT-18-2016 project is to automatically check if personal data processing and sharing performed by data controllers and processors complies with both the obligations set forth in the General Data Protection Regulation (GDPR) and usage constraints specified by data subjects. A necessary first step is to develop a policy language that can be used to represent legislative obligations in a machine readable format. In order to better understand the constructs that are required to represent the GDPR in a machine readable manner it is first necessary to do a structural analysis of the GDPR from a rules perspective. This deliverable presents the results of our initial analysis of the types of rules that are required and a discussion as to the structure of the policy language.

It is worth noting that the work described herein is purely a technical analysis and does not at this stage include a legal interpretation of the underlying concepts of the data protection laws, such as fundamental rights protection, or the perspective of the data subject. These are overarching things that do not have relation only to the GDPR, but must be seen in a wider context, which cannot be covered here (e.g. context of other laws, society, moral concepts, etc.). Considering the iterative nature of the SPECIAL project, such considerations will be included in later deliverables.

Later versions will also include a distinction between the rules and principles set forth in the GDPR. Rules are definitive requirements. A rule can only be fulfilled or not fulfilled. Principles can be understood as optimisation commandments. They should be fulfilled to the highest degree depending on the legal and factual possibilities and circumstances. The difference between those two kinds of norms plays a role when there are conflicts between different norms. For example, if you have two colliding norms with conflicting requirements, it becomes important to know whether these norms are rules or principles. Conflicts between rules can be resolved if there is an exemption from a rule, or if one of the two rules is invalid. Conflicts between principles are resolved differently since they can be fulfilled to varying degrees. Therefore, principles are scalable. If principles collide, one principle might be regarded as more important than the other, while no exemption or invalidness is needed.

Additionally, this deliverable does not take into account the specific circumstances and inherent risks of individual processing operations. In the case of the SPECIAL project this contextual information will be provided by the pilot leaders. In relation to the aforementioned principles, this has an effect on which principles may be regarded more or less important, depending on the context. For example, it makes a difference if a company processes personal data based on valid consent from the data subject, or e.g. a police force collects and processes data for a criminal investigation. If you compare these two very different scenarios, you may imagine that for the company, the principle of transparency is quite important to obtain informed and free (and thus legally valid) consent, while for a police force, this principle is less important, since they may not want the suspect to know that an investigation is ongoing. In the case of the police scenario, other principles might play a much bigger role due to the specific risks of the processing. For instance, the need of data integrity is incredibly high since literally lives depend on the information being correct.

What is in this deliverable

Chapter 1 summarises the results of our structural analysis of the text of the GDPR. The primary goal being to derive a set of must-have structural requirements, that are necessary in order to

represent the GDPR in a machine readable format. *Chapter* 2 in turn details our progress on the the formal representation of the GDPR, which will be iteratively refined throughout the course of the project and completed with the release of the final versions of the pilots' business policies and GDPR's formalization (as part of the implementation packages WP3 and WP5).

What is *not* in this deliverable

The goal of the initial analysis presented in this deliverable is to gain a better understanding of the expressivity needed in order to model the relevant subsets of the GDPR in a machine-understandable way.

Therefore, the initial analysis focused a structural and text analysis of the GDPR. Next steps include: the isolation of rules than can be checked automatically; the breakdown of said rules into discrete components that can be compared against both business policies and processing and sharing events; and the incorporation of legal interpretations into the analysis and the policy language.

As per *Deliverable D2.5 Policy Language V2*, additional joint work with the pilot leaders and the members with legal expertise is needed in order to detail the vocabularies for the policy language and the legal assessment. All examples currently included in the paper are generic and as such are only meant to illustrate the policies' structure.

Additionally, further iterations of GDPR's formalization will take on-board the results of SPECIALs standardisation activities, that include also aspects specifically related to GDPR's concepts. Further details on SPECIALs standardisation activities can be found in *Deliverable D6.3 Plan for community group and standardisation contribution*.

Chapter 1

Analysis of the GDPR

1 GDPR Analysis

Laws on Data protection follow a known technique from German public law since the seventies. While the default in a democratic society is that everything is allowed unless it is prohibited, the order is reversed by Data protection laws. Already the Directive 95/46EC contained an Article 7 saying: `Member states shall provide that personal data may be processed` **`only`** `if...`. Article 6 (1) of the GDPR contains the same general prohibition of the processing of personal data by stating: `Processing shall be lawful` **`only`** `if...`. This general prohibition of the processing of personal data is then accompanied with large sweeping clauses containing permissions. In this document, those permissions are modelled as *dispensations*. It may be discussed further whether dispensations and permissions have the same functionality. In the normal course of action, the dispensation of the general prohibition will contain further rules of all types. If one of those rules is unsatisfied the chain of permissions or dispensations will collapse leading to a fallback of the general prohibition stated in Article 6 (1).

1.1 Approach

The primary goals of our analysis were to identify the articles from the GDPR that related to consent and transparency and to determine the type of rules that are required in order to record relative legislative obligations. Before describing the results of our analysis we first describe our approach, which can be divided into the following concrete tasks:

1. The first task was to identify GDPR articles that relate to consent and/or transparency either directly or indirectly. This task was guided by a review of the legal literature that directly relates to the GDPR. The goal of the literature review being to identify relevant articles and/or paragraphs within the regulation that need to be analysed in detail.

2. Prior to commencing with the analysis of the GDPR we first devised a controlled vocabulary that could be used to annotate the text of the GDPR (i.e. a dictionary *Table* 1.1). The objective being to identify a set of expressions that can be used to accurately describe legal requirements in a way that can later be translated into machine processable rules. The initial dictionary, was composed of `Obligations` (used to describe obligations that must be fulfilled by companies), `Constraints` (used to constrain the obligation)

Table 1.1: GDPR Annotation Dictionary

Type	Annotation	Description
Prohibition	`P`	you must not (i.e. equivalent to negative obligation)
Obligations	`O`	you must
Dispensation	`D`	exemption from the rule (dispensation condition for processing in a legal sense)
Constraints	`+C` `-C`	a limitation or restriction (e.g. its allowed if) a limitation or restriction (e.g. its allowed if you don't)
Definitions	`Def`	explains the meaning of a certain term or defines how an obligation or a constraint must be understood
References	`eRef[]` `tRef[]`	an article contains an explicit reference (e.g. eRef[Art. 89 (1)]) an article contains a reference related to a certain term (e.g. tRef[consent])
Dispositions	`Disp`	an example/best practice/suggestion
Opening Clause	`OC`	indicates a need to consult other legislation (National or European)

and `Definitions` (used to define meaning). However, the dictionary was subsequently extended to include both positive and negative `Obligations` and `Constraints`. In addition, we added both explicit (denoted using an eRef) and implicit references (denoted using an tRef), `Dispensations` (used to record exemptions), `Dispositions` (used to highlight best practices/suggestions) and `Temporal` (used to identify temporal requirements). Finally, subjectives terms that may be dependent on further interpretation or on judicial decisions were marked using red typeface.

3. The analysis of the relevant articles from the GDPR was first conducted on paper and consisted of two steps: dividing the paragraphs into segments according to their meaning; and annotating each segment according to the annotation vocabulary.

4. On completion of the desktop based analysis of the GDPR, the identified segments and corresponding annotations were recorded in an excel spreadsheet. Both colour and segmentation were used to make the analysis easier digest. Additionally, all annotations were checked by a second person with a view to uncovering errors and omissions. The inconsistencies identified were discussed between both parties until agreement was reached with respect to annotation.

5. Once the final output is available the initial analysis will be reviewed by at least two other team members and again inconsistencies will be highlighted and discussed until a consensus with respect to the most appropriate annotation is reached between the various parties.

Table 1.2: GDPR Rule Structure

Type	Rule Structure
Prohibitions	P w/wo C OR C w/wo P
Obligations	O w/wo C OR C w/wo O
Dispensations	D w/wo C OR C w/wo D
Obligations and Prohibition	O w/wo C OR C w/wo O Followed By P w/wo C OR C w/wo P
Obligations and Dispensation	O w/wo C OR C w/wo O Followed By D w/wo C OR C w/wo D
Obligations, Prohibition and Dispensation	O w/wo C OR C w/wo O Followed By P w/wo C OR C w/wo P Followed By D w/wo C OR C w/wo D
Opening Clause	OC

1.2 GDPR Structure Analysis

This section presents a summary of the results of our initial text analysis of the GDPR. A high level overview of the different rule structures is presented in *Table* 1.2[1]. While, some concrete examples of the different types of rules found in the GDPR is presented below. It is worth noting that the objective of the initial analysis is not to decide on the granularity of the rules but rather to determine the type of rules that will be required and the general structure of said rules.

Prohibition A prohibition in the legal sense is essentially a rule or law that forbids something (i.e. you must not). The GDPR contains a general prohibition in Article 6, as was pointed out in the introduction.

Article 6 relates to the *lawfulness of processing* and contains the general prohibition of the processing of personal data, which is derived from the general freedom of action in a democratic society. The Article does not use the word *prohibition*. It uses the word *lawful only* with the implicit understanding that unlawful processing is subject to administrative fines according to Article 83. The prohibition is not in itself total as the law itself is scoped excluding the areas enumerated in Article 2 (2 & 3). This general prohibition is complemented by dispensations, obligations and constraints that typically contain the more detailed rules.

(P start) Processing **(P end)** shall be lawful only (R1)

Article 9 reiterates the general prohibition of the processing of personal data and adds additional rules for the *processing of special categories of personal data*. Following the same approach as Article 6, Article 9 starts with a prohibition and then enumerates dispensations. By stating certain categories of personal data Article 9 defines its scope in relation to Article

[1]The shorthand notation w/wo is used to denote with or without

6. It then subsequently indicates several dispensations to this prohibition that contain special constraints compared to the dispensations in Article 6. Paragraph 1 states that the *processing of personal data* with the following `constraint` *revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*, and continues by stating the `prohibition` (*see* R2).

```
(P start) Processing of personal data                    (R2)
(C) revealing racial or ethnic origin, political
opinions, religious or philosophical beliefs, or trade
union membership, and the processing of genetic data,
biometric data for the purpose of uniquely identifying a
natural person, data concerning health or data concerning
a natural person's sex life or sexual orientation
(P end) shall be prohibited.
```

Article 21, which relates to the *right to object*, paragraph 3 states *where the data subject objects to processing for direct marketing purposes the personal data shall no longer be processed for such purposes*. Here there is a `constraint` indicating where the data subject objects to processing for direct marketing purpose and a corresponding `prohibition` indicating that it is not permissible to continue processing the data form direct marketing purposes (*see* R3). Because of the general prohibition by Article 6, the prohibition here is the end of a dispensation (*no longer*).

```
(C) Where the data subject objects to processing for     (R3)
direct marketing purposes,
(P) the personal data shall no longer be processed for
such purposes.
```

Dispensations A dispensation is essentially an exemption to a rule. Like obligations and prohibitions, dispensations are generally either preceded or followed by one or more constraints.

Article 6 (1) does not only contain the general prohibition of the processing of personal data, but also general dispensations in paragraphs a – f. One of those general dispensations has to apply to make the processing of personal data lawful. The cases specified in the points presented in Article 6 are then further detailed in other Articles or even in laws outside the GDPR. Article 6 (1) provides for a dispensation from the general prohibition if *the data subject has given consent to the processing of his or her personal data for one or more specific purposes;*. Here there is a `dispensation` with respect to processing of personal data if a `constraint` indicating that *the data subject has given consent to the processing of his or her personal data for one or more*

specific purposes is satisfied (*see* R4).

(D) `Processing shall be lawful only if and to the extent that at least one of the following applies:`
(C) `the data subject has given consent to the processing of his or her personal data for one or more specific purposes;` (R4)

Article 11, which relates to *processing which does not require identification*, paragraph 1 states *if the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.* Here there is a `constraint` indicating *the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller,* and a corresponding `dispensation` *the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation* (*see* R5).

(C) `If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller,`
(D) `the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.` (R5)

Obligations An obligation in the legal sense is a duty that must be fulfilled (i.e. you must). Although it is possible to have standalone obligations generally speaking obligations are either preceded or followed by one or more constraints.

Article 5, which denotes the *principles relating to processing of personal data*, paragraph 1 starts by stating *personal data shall be:* (and continues under point (a) stating) *processed lawfully, fairly and in a transparent manner in relation to the data subject*. Here there is an `obligation` that personal data is *processed lawfully, fairly and in a transparent manner* and a `constraint` that the data to be processed is *in relation to the data subject* (*see* R6). Although personal data by definition relates to a data subject, for this initial analysis we would like to stay as close as possible to the actual text of the GDPR, therefore we model the constraint nonetheless (*see* R6).

(O) `Personal data shall be: processed lawfully, fairly and in a transparent manner`
(C) `in relation to the data subject` (R6)

Article 7, which relates to the *conditions for consent*, paragraph 1 starts by stating *where processing is based on consent* and continues by stating *the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data*. Here there is a `constraint` when the processing is based on consent and a corresponding `obligation` that the controller is able to demonstrate that they have consent for the processing (*see* R7).

(C) `Where processing is based on consent,` (R7)
(O) `the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.`

Obligation and Dispensation Article 19, which denotes *notification obligation regarding rectification or erasure of personal data or restriction of processing*, indicates that *the controller shall communicate any rectification or erasure of personal data or restriction of processing carried out in accordance with Article 16, Article 17(1) and Article 18 to each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort*. Here there is an `obligation` to *communicate any rectification or erasure of personal data or restriction of processing*, however there is a `dispensation` in case *this proves impossible or involves disproportionate effort* (*see* R8).

(O) `The controller shall communicate any rectification or erasure of personal data or restriction of processing carried out in accordance with Article 16, Article 17(1) and Article 18 to each recipient to whom the personal data have been disclosed,` (R8)
(D) `unless this proves impossible or involves disproportionate effort`

Article 12, which relates to *transparent information, communication and modalities for the exercise of the rights of the data subject*, paragraph 4 states *if the controller does not take action on the request of the data subject, the controller shall inform the data subject without delay and at the latest within one month of receipt of the request of the reasons for not taking action and on the possibility of lodging a complaint with a supervisory authority and seeking a judicial remedy*. Here there is an implicit dispensation stating *if the controller does not take action on the request of the data subject* and an corresponding `obligation` indicating information that

must be provided to the data subject (*see* R9).

```
(D) If the controller does not take action on the request      (R9)
of the data subject
(O)the controller shall inform the data subject without
delay and at the latest within one month of receipt of
the request of the reasons for not taking action and on
the possibility of lodging a complaint with a supervisory
authority and seeking a judicial remedy.
```

Obligation, Prohibition and Dispensation Article 5, which denotes the *principles relating to processing of personal data*, paragraph 1 starts by stating *Personal data shall be:* (and continues under point (b) stating) *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes*. Here there is an `obligation` that personal data is *collected for specified, explicit and legitimate purposes* and a `prohibition` that the data is not *further processed* with a `constraint` *in a manner that is incompatible with those purposes*. Finally there is a `dispensation` which states that *further processing shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes;* with a `constraint` *for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes* (*see* R10).

```
(O) collected for specified, explicit and legitimate          (R10)
purposes
(P) and not further processed
(C) in a manner that is incompatible with those purposes;
(D start) further processing
(C) for archiving purposes in the public interest,
scientific or historical research purposes or statistical
purposes
(D end) shall, in accordance with Article 89(1), not be
considered to be incompatible with the initial purposes;
```

Opening Clause Opening Clauses permit legislators to further refine via other National or European legislation. Article 9 paragraph 4 indicates that when it comes to *the processing of genetic data, biometric data or data concerning health* it is necessary to consult relevant

Table 1.3: GDPR Temporal Expressions

Annotation	Expression
at any time	the right to withdraw his or her consent **at any time**
before	processing based on consent **before** its withdrawal
prior to	**prior to** giving consent
at the time	**at the time** when personal data are obtained
within one month	**within one month** of receipt of the request
at the latest	**at the latest** at the time of the first communication
without undue delay	**without undue delay** the rectification of inaccurate personal data concerning him or her
for a period	**for a period** enabling the controller to verify the accuracy of the personal data
no longer	the personal data shall **no longer** be processed

member state legislation (*see* R11).

(OC) Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health. (R11)

1.3 GDPR Text Analysis

In addition to analysing the structure of rules, our initial analysis included the annotation of temporal references, both explicit and implicit references to other articles and paragraphs, and terms that we deemed subjective.

1.3.1 Temporal Annotations

In this section, we examine the different temporal conditions found in the text of the GDPR. A high level overview of the different temporal expressions is presented in *Table* 1.3. While, some more detailed examples are presented below.

At any time, before & prior Article 7 (*Conditions for consent*), paragraph 3 uses terms such as *at any time, before* and *prior to* (*see* R12).

The data subject shall have the right to withdraw his or her consent **at any time.** The withdrawal of consent shall not affect the lawfulness of processing based on consent **before** its withdrawal. **Prior to** giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent. (R12)

At the time Article 13 (*Information to be provided where personal data are collected from the data subject*), paragraphs 1 and 2 refers to *at the time* when personal data are obtained (*see* R13 and R14).

> Where personal data relating to a data subject are collected from the data subject, the controller shall, **at the time** when personal data are obtained, provide the data subject with all of the following information: (R13)

> In addition to the information referred to in paragraph 1, the controller shall, **at the time** when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing: (R14)

Article 21 (*right to object*), paragraph 4 refers to *at the time* (*see* R15).

> **At the latest at the time** of the first communication with the data subject, the right referred to in paragraphs 1 and 2 shall be explicitly brought to the attention of the data subject and shall be presented clearly and separately from any other information. (R15)

Within a reasonable period after, within one month & at the latest Article 14 (*Information to be provided where personal data have not been obtained from the data subject*), paragraphs 3 refers to *within a reasonable period after*, *within one month* and *at the latest* (*see* R16).

> The controller shall provide the information referred to in paragraphs 1 and 2: (a) **within a reasonable period after** obtaining the personal data, but at the latest **within one month**, having regard to the specific circumstances in which the personal data are processed; (b) if the personal data are to be used for communication with the data subject, at the latest **at the time** of the first communication to that data subject; or (c) if a disclosure to another recipient is envisaged, **at the latest** when the personal data are first disclosed. (R16)

Without undue delay Article 16 (*Right to rectification*), refers to *without undue delay* (*see* R17).

```
The data subject shall have the right to obtain from
the controller without undue delay the rectification of
inaccurate personal data concerning him or her.
```
(R17)

For a period Article 18 (*Right to restriction of processing*), paragraph 1 refers to *for a period* (*see* R18).

```
The data subject shall have the right to obtain from the
controller restriction of processing where one of the
following applies: (a) the accuracy of the personal data
is contested by the data subject, for a period enabling
the controller to verify the accuracy of the personal
data;
```
(R18)

No longer Article 21 (*Right to object*), paragraph 3 is an example of the use of *no longer* with respect to the processing of personal data (*see* R19).

```
Where the data subject objects to processing for direct
marketing purposes, the personal data shall no longer be
processed for such purposes.
```
(R19)

1.3.2 Explicit and Implicit References

In this section, we present the different types of references found in the GDPR and give some concrete examples based on Article 5, which denotes the *principles relating to processing of personal data*.

Implicit Reference Article 5, paragraph 1 starts by stating *personal data shall be:* (and continues under point (a) stating) *processed lawfully, fairly and in a transparent manner in relation to the data subject* (*see* R20). Here there is an implicit reference to Article 6 *Lawfulness of Processing* (which we denote using a tRef annotation: **tRef** [Art 6]) and Article 12 *Transparent information, communication and modalities for the exercise of the rights of the data subject* (which we denote using a tRef annotation: **tRef** [Art 12]).

```
Personal data shall be: (a) processed lawfully, fairly
and in a transparent manner in relation to the data
subject ('lawfulness, fairness and transparency');
```
(R20)

Explicit Reference Article 5, paragraph 1 starts by stating *personal data shall be:* (and continues under point (b) stating) *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation')* (*see* R21). Here there is an explicit reference to Article 89 paragraph 1 (which we denote using an eRef annotation: **eRef** [Art 89(1)]).

```
Personal data shall be: (b) collected for specified,
explicit and legitimate purposes and not further
processed in a manner that is incompatible with those
purposes; further processing for archiving purposes in
the public interest, scientific or historical research
purposes or statistical purposes shall, in accordance
with Article 89(1), not be considered to be incompatible
with the initial purposes ('purpose limitation');
```
(R21)

A second type of explicit reference can be seen in Article 5 paragraph 2, which simply has a textual reference to paragraph 1 (*see* R22). Here there is an explicit reference to paragraph 1 (which we denote using an eRef annotation: **eRef** [Art 5 (1)]).

```
The controller shall be responsible for, and be
able to demonstrate compliance with, paragraph 1
('accountability').
```
(R22)

1.3.3 Subjective Terms

The open textured nature of legal texts is a highly desirable feature, as it leaves room for interpretation on a case by case basis, however when such ambiguity poses challenges for automatic compliance checking.

Article 7, which relates to the *conditions for consent*, paragraph 2 provides guidelines for consent *in the context of a written declaration which also concerns other matters*. In the context of this article terms such as *clearly distinguishable, intelligible, easily accessible, using clear and plain language* are open to human interpretation and as such cannot be verified automatically.

```
If the data subject's consent is given in the context of
a written declaration which also concerns other matters,
the request for consent shall be presented in a manner
which is clearly distinguishable from the other matters,
in an intelligible and easily accessible form, using
clear and plain language. Any part of such a declaration
which constitutes an infringement of this Regulation
shall not be binding.
```
(R23)

1.4 External Assessment of the Analysis

The primary goal of the research team is to determine what aspects of the GDPR could potentially be checked automatically, with a view to determining what is the best way to formally represent these obligations such that it is possible to (semi-)automatically (possibly with the help of external sources such as court proceedings) to verify if personal data processing and sharing performed by data controllers and processors complies with obligations in GDPR. The main purpose of the external survey was to assess the systematic analysis of the GDPR text, which is being conducted within the framework of the SPECIAL Project.

1.4.1 Objectives

The survey consisted of five tasks that contain several sub-questions and sub-exercises. The survey aims to provide for the following outcomes:

- First, to identify the essential legal provisions within the GDPR that should be primarily applied when the legality of data processing needs to be assessed;
- Second, to clarify the systematic order of how the individual GDPR provisions should be invoked; and
- Third, to evaluate the annotation of the GDPR text completed by the project team members, with the special focus on the implicit references (iRef), explicit references (eRef) and subjective terms (S).

Three independent legal experts on the data protection (who are not connected to the SPECIAL project), with extensive legal research experience, were approached to complete the evaluation survey. In order to ensure maximum objectivity, their identity is anonymised. In the text that follows respondents are referred to as `respondent A`, `B` and `C`.

1.4.2 Preliminary Questions

All three respondents indicated in the survey that they have legal background and that they actively work in the field of data protection. When asked to rate their familiarity with the GDPR text on the scale ranging from 1 to 5 (5 being the highest), `Respondents A` and `C` selected the highest rate of 5, while `Respondent B` opted for 3.

The objective of the survey was to identify the scope of the GDPR legal provisions, which should be applied when automatically checking compliance with GDPR obligations governing personal data processing and sharing performed by data controllers and processors. Based on the use case scenario given to the respondents at the beginning of the evaluation exercise, they were asked to list Articles of the GDPR, from which they would start their legal analysis and consequently, to clearly indicate the most relevant Articles applicable to the given use case scenario.

`Respondant A` started their legal assessment with Article 4 GDPR that contains the legal definitions, in order to determine whether the use case concerns a data controller or a data processor. According to `Respondant A`, it is crucial for the use case resolution to establish if the use case falls under the category of a data processor or a data controller. As for the most relevant measures of GDPR for the use case, `Respondant A` indicated the whole Chapter 4 of GDPR, from Article 24 to 43, which stipulates the legal obligations for both, data controllers

and data processors. The broadness of Respondant A's response may be due to the lack of information provided in the use case scenario. It also underlines the need to determine the exact knowledge of what kind of activity is at stake: data processing or data controlling.

On the other hand, Respondant B solely focused on the actual material and territorial scope stipulated within Article 2 and 3 GDPR, in order to determine whether the GDPR is actually applicable to the given case scenario. If the use case does not concern data processing, Respondant B suggests to consult the provision dealing with data protection by design and by default (Article 25). Overall, Respondant B listed as relevant provisions Article 2 (material scope), Article 3 (territorial scope), and similarly to Respondant A Article 4 (definitions).

Respondant C starts the legal analysis by using Article 6 (Lawfulness of data processing), emphasising the importance of clarifying the purpose and the nature of collected data. If consent is needed, Articles 7 and 8 should be considered as well. When the kind of data to be collected is known, Article 9 (Processing of special categories of personal data) should be invoked. Thus, the exact kind of processing activity is crucial for Respondant C. Regarding the concrete Articles, Respondant C considers Articles 6, 7, 8 and 9 as necessary if the use case concerns only the lawfulness of data processing. However, Respondant C also raises the issue of lack of information about whether the use case deals with data processing or data controlling. In the case of data processing, Articles 28 (Processor), Article 29 (Processing under the authority of the controller or processor), Article 30 (Records of processing activities), Article 31 (Cooperation with the supervisory authority) and Article 33 (Notification of a personal data breach to the supervisory authority) will be equally relevant to the use case.

1.4.3 Process/Approach

The second part of the survey aimed to clarify the systematic application of the individual GDPR provisions in the given case scenario. The respondents were asked to first, indicate the exact starting point of their legal assessment within the GDPR framework and second, to determine the exact order of their analysis.

The starting point for Respondant A is clearly Article 4 GDPR (Definitions), in order to know what kind of duties and responsibilities have to be met. Given that the use case deals with a data controller, Respondant A would then continue with Article 24 (Responsibility of the controller), Article 25 (Data protection by design and by default), Article 32 (Security of processing) and with other important GDPR provisions governing the role and duties of data controllers in Chapter 4 of the GDPR. Only afterwards would Respondant A would turn to Article 5 (Principles relating to processing of personal data), Article 6, 7, 8, 9, 10 (Processing of personal data relating to criminal convictions and offences) and Article 11 (Processing which does not require identification).

Respondant B starts the legal assessment by determining the scope and the applicability of GDPR framework (Articles 2 and 3). If GDPR is applicable in this particular case, Respondant B would then carry out the assessment by looking at or creating a data element inventory in line with Article 30 (Records of processing activities) as well as conducting a privacy impact assessment (Article 35). Regarding the order of the legal analysis, Respondant B's starting point is Articles 2 and 3, continuing with Article 6, Article 5, obligations of data controller and data processor as stipulated in Chapter 4, The Rights of Data Subject (Article 12) and finally, other applicable legal requirements depending on the contextual information about the use case.

For Respondant C, it is very important to know whether the use case is about data processing or data controlling. Thus, Article 4 GDRP should be applied before any other legal assessment is carried out. If the use case indeed concerns a data processor, Article 28 should be invoked next and consequently, Article 29 (Processing under the authority of the controller or processor), Article 30, Article 32 and Article 33(3) (Notification of a personal data breach to the supervisory authority) should follow. Respondant C concludes the legal assessment with Article 82(1-2) that regulates the right to compensation and liability.

1.4.4 Multiple Interpretations

In this part of the Survey, the annotation of iRef and eRef within the text of GDPR is being evaluated by the Respondents. All three Respondents agreed that some provisions of GDPR require further information in order to achieve their proper legal interpretation. When asked about what kind of legal authority or sources they would consult, Respondant A underlined the need for a precise CourtsâĂŹ jurisprudence defining the concept of a data controller. However, Respondant A did not specify further, whether he refers to the national Courts or the Court of Justice of the European Union. Respondant B indicated that many Articles of the GDPR would require further elaboration or sector specific information. As an example of sources, Respondant B mentions the Article 29 Working Party, Specific EU or national legal frameworks and court judgements. Although Respondant C agrees that other legal sources need to be consulted, it seems that Respondant C′s answer concerns only the given use case scenario and not the general interpretation of the GDPR text. Respondant C suggests that initial Articles of GDPR, such as the definitions, the principles, the legal basis for processing and the function of consent, might need further consultation. However, it seems from the answer's formulation that the further consultation is needed in relation to the particular case. Thus, it is not clear whether Respondant C had in mind the iRef and eRef within the text of GDPR, when answering this question.

1.4.5 Discussion

Overall the first evaluation showed that even though approaches differ to a certain extent, several provisions were mentioned by all evaluators. The articles mentioned match our initial expectation of Articles that must be covered by the analysis to enable compliance checking in general. The difference in approaches may be due to the intentionally vague case description that let room for various interpretations of the situation. According to our own assessment and the answers of the evaluators we can now identify articles that, so to say, build the core of any GDPR compliance assessment. Necessary provisions include:

- Applicability of the GDPR
- Principles of processing
- Lawfulness of processing

These provisions will have to be assessed every time, accompanied by more specific provisions according to the individual use case.

Now that we have confirmed the Articles that need to be checked, the next step will be to take a very close look on these core provisions. In order to find out what can be checked

(semi-) automatically, we must ensure that every term has a clearly defined meaning. That is because our system will not allow us to assess undefined terms. The main issue here is that most provisions of the GDPR require a great amount of interpretation that has to be somehow done in advance so that the system can be filled with legal knowledge. For example, it is relatively easy to have the system check whether a person lives in the EU or not. If there is enough information about the person provided this is just a Yes or No decision that a machine can make. However, it is much harder to have the system check things like consent. It is not possible to check whether consent was given for specific purposes if the computer does not know what the word specific means in this context. A human being can always base their judgement on the circumstances of the individual case compared with materials and jurisprudence, but a machine can not. That is the reason why it will be essential to find every term that requires human interpretation before setting up the system. For the purpose of finding and resolving those terms we distinguish between references and subjective terms. References can either refer to another piece of legislation [eRef] or to implicit knowledge about the law [iRef] (e.g. âĂIJscope of Union LawâĂİ). Subjective terms are single words or parts of a sentence that can be interpreted in various ways, possibly giving a provision multiple meanings. As a starting point for this task we focus on Chapter II of the GDPR, and especially on Art 5 and Art 6 where we already identified references and subjective terms. Subsequently with the next phase of the evaluation we will target the completeness of the reference and subjectivity assessment.

Acknowledgements

We thank Eliska Pirkova from the faculty of law at the University of Helsinki (`eliska.pirkova@helsinki.fi`) for contributions to *Chapter* 1.

Chapter 2

Partial GDPR compliance in OWL 2

1 Business Policies as Partial Business Process Descriptions

The GDPR sets obligations that apply to the data controller's internal organization and processes. Here are two examples:

- whenever the data controller operates on personal data, it must *acquire explicit consent* from the involved data subjects, unless the purpose of data processing falls within a set of exceptional cases (e.g. the processing is required by law);
- whenever data are transferred to a branch of the data controller residing in a country whose data protection regulations do not match the EU requirements, guarantees must be provided in the form of company regulations; the GDPR calls them *binding corporate rules*.

Moreover, and differently from the above examples, the GDPR sets obligations that are not directly related to the controller's business processes, such as the requirement that data subjects can *access, rectify, and delete* their personal data. In order to fulfil such obligations, data controllers have to set up suitable processes.

The above observations show that checking compliance with the GDPR requires as an input a description of the data controllers' internal processes. For automated compliance checking such description should be adequately formalized in a machine-understandable way; moreover, the formalization should represent accurately the real processes, in order to make the automated compliance verification reliable.

Introducing Business Policies and their Relations with Business Processes

In SPECIAL, we address a concrete setting – suggested by one of the industrial partners – in which a partial and abstract description of processes is available. Each process description is shaped like a *formalized business policy* consisting of the following set of features:

- the file(s) to be processed;
- the software that carries out the processing;
- the purpose of the processing;

- the entities that can access the results of the processing;
- the details of where the results are stored and for how long;
- *the obligations that are fulfilled while (or before) carrying out the processing.*

It is not hard to see that the first five elements in the above list match the Minimal Core Model (MCM) implemented by SPECIAL's *usage policy language* (UPL) introduced in D2.1 and extended in D2.5. As far as the above elements are concerned, the only difference between UPL expressions and a business policy is the granularity of attribute values. For example, the involved data (specified in first element in the above list) are not expressed as a general, content-oriented category, but rather as a concrete set of data sources or data items. Such objects can be modelled as instances or subclasses of the general data categories illustrated in D2.1, thereby creating a link between digital artifacts and usage policies. Similar considerations hold for the other attributes:

- processing is not necessarily described in the abstract terms adopted by the processing vocabulary introduced in D2.1; in a business policy, this can be specified by naming concrete software procedures;
- the purpose of data processing may be directly related to the data controller's mission and products;
- recipients may consist of a concrete list of legal and/or physical persons, as opposed to general categories such as `Ours` or `ThirdParty`;
- storage may be specified by a list of specific data repositories, at the level of files and hosts.

With this level of granularity, specific access control authorizations can be derived from the business policy, for example:

> *The indicated software procedure can read the indicated data sources. The results can be written in the specified repositories. The specified recipients can read the repositories...*

This methodology for generating authorizations fosters a close correpondence between the business policy and the actual behavior of the data controller's systems and processes.

The last attribute of a business policy (that specifies obligations) is not part of usage policies. It plays a dual role:

- on the one hand, it represents a precondition to the authorizations specified by the business policy, e.g. if the obligation is something like `getValidConsent` then the derived authorizations is a *rule* like *the specified software can read the data sources if consent has been given*;
- on the other hand, the list of obligations witnesses that the data controller has set up *processes for fulfilling the indicated obligations* – e.g. a process to obtain consent from the data subjects – which is relevant to checking compliance with the GDPR.

In this chapter, we show how to leverage the partial business process descriptions encoded in business policies to check compliance with some of the GDPR's articles. There are two aspects in this plan: (i) how to encode business policies in a machine understandable way, and (ii) how to encode the relevant parts of the GDPR in a machine undestandable way. Once these two goals have been achieved, it will be possible to verify automatically whether a set of business policies is compliant using the reasoning tools for subsumption checking, that have been outlined in D2.1 and specialized in D2.4 and D2.8.

In the next two sections we recall the first version of the business policy language and GDPR encoding. Then, in Section 4 we introduce some extensions to business policies that are useful in modelling a larger set of requirements posed by the GDPR.

2 Business Policies in OWL 2

A basic business policy is simply a usage policy (as in D2.1) extended with zero or more obligations,[1] encoded with attribute `hasDuty`, as in

```
ObjectIntersectionOf(                                                    (2.1)
   ObjectSomeValuesFrom(<spl:hasData> SomeDataCategory)
   ObjectSomeValuesFrom(<spl:hasProcessing> SomeProcessing)
   ObjectSomeValuesFrom(<spl:hasPurpose> SomePurpose)
   ObjectSomeValuesFrom(<spl:hasRecipient> SomeRecipient)
   ObjectSomeValuesFrom(<spl:hasStorage> SomeStorage)
   ObjectSomeValuesFrom(<sbpl:hasDuty> SomeDuty)
)
```

Multiple obligations are expressed by replicating the `hasDuty` expression, for example the following policy associates the collection of personal demographic information to the obligations to get consent and let the data subject exercise her rights:

```
ObjectIntersectionOf(
   ObjectSomeValuesFrom(spl:hasData svd:Demographic)
   ObjectSomeValuesFrom(spl:hasProcessing svpr:Collect)
   ObjectSomeValuesFrom(spl:hasPurpose svpu:Account)
   ObjectSomeValuesFrom(spl:hasRecipient svr:Ours)
   ObjectSomeValuesFrom(spl:hasStorage
      ObjectIntersectionOf(
         spl:hasLocation svl:OurServers
         spl:hasDuration svdu:Indefinitely
      )
   )
   ObjectSomeValuesFrom(sbpl:hasDuty getValidConsent)
   ObjectSomeValuesFrom(sbpl:hasDuty getAccessReqs)
   ObjectSomeValuesFrom(sbpl:hasDuty getRectifyReqs)
   ObjectSomeValuesFrom(sbpl:hasDuty getDeleteReqs)
)
```

[1] Eventually, business policies will be extended with further attributes to encode the policy's version, its validity period, and any further metadata is needed to manage the policy. Such additional attributes are being jointly identified with the pilot leaders and will be included in the next versions of the deliverable and of the business policy language.

Figure 2.1: SPECIAL's Business Policy Language Grammar

BusinessPolicy := **BasicBP** | 'ObjectUnionOf' '(' **BasicBP BasicBP** { **BasicBP** } ')'

BasicBP :='ObjectIntersectionOf' '(' **Data Purpose Processing Recipients Storage {Duty}** ')'

Data := *see D2.1*

Purpose := *see D2.1*

Processing := *see D2.1*

Recipients := *see D2.1*

Storage := *see D2.1*

Duty :='ObjectSomeValuesFrom' '(' 'sbpl:hasDuty' **DutyExpression** ')'

DutyExpression :='sbpl:AnyDuty' | **DutyVocabExpression**

DutyVocabExpression := *to be specified in WP3 and WP5*

Note the difference between the attribute of usage policies and `hasDuty`: the former are functional (i.e. each attribute of a single authorization has one value), while `hasDuty` is *not* functional and each policy may be associated to multiple obligations. The possible values for the `sbpl:hasDuty` attribute will be illustrated later on.

Similarly to usage policies, *general* business policies can be composed by enclosing several basic business policies inside the `ObjectUnionOf` operator of OWL 2.

Full syntax and the logical semantics of SPECIAL's Business Policy Language are specified in Figure 2.1 and Appendix 1, respectively.

3 Formalizing Selected Parts of the GDPR in OWL 2

Here we recall the first formalization of the GDPR and its application, as they were introduced in D2.2. Accordingly, here we only deal with the features that belong to the first version of business policies. A more advanced view is introduced in Section 4.

3.1 Getting Consent for Processing Personal Data

The GDPR requires that personal data be processed only after consent has been appropriately obtained from the data subjects. This is not the only legal ground for processing (some alternatives will be discussed later), however *if* valid consent is available, *then* processing is legal, even if other legal grounds are applicable as well.[2]

Legally valid consent is defined in fuzzy and rather subjective terms (cf. D1.3). Therefore, it is impossible to verify automatically that consent has been properly obtained. Accordingly, in our formalization, when we refer to consent we implicitly mean "legally valid consent" – from our axiomatic perspective, *illegal consent is no consent at all*. The legal validity of consent requests shall be certified by humans (preferably with a specific legal background).

[2]It should be mentioned that there are also special cases where consent does not suffice, e.g. when data are about criminal convictions, cf. Art. 10. In this case we regard consent as *not* valid.

Anonymous data are *not* regarded as personal data, so the corresponding GDPR's restrictions do not apply. We will refer to anonymous data with the term `svd:Anonymous`.

> Note that `svd:Anonymous` does not occur among the data categories illustrated in D2.2. The apparently similar term `svd:Anonymized` refers to the output of anonymisation algorithms, while `svd:Anonymous` refers to data that can be regarded as anonymous *in legal terms*, i.e. it should be *impossible* to re-identify a data subject. No state-of-the-art algorithm can guarantee this. *If* future regulations will relax this definition by explicitly stating that certain anonymity guarantees (e.g. ϵ-differential privacy, for a specified ϵ) are equivalent to *legal* (i.e. perfect) anonymity, then the data that satisfy such anonymity guarantee shall be classified under `svd:Anonymous`. Until that day, the output of those anonymization algorithms shall be classified under the (different) term `svd:Anonymized`.

Some exceptions to the need for consent are related to the *purpose* of data processing, e.g. consent – or further consent – is not necessary if data are processed because law requires it, or for purpose recognized as unequivocally compatible with the original purpose (e.g. archiving or scientific research). Archiving is captured by the class `svpu:Historical` of the purpose vocabulary, while law requirements call for an additional class `svpu:Law`.

All the aforementioned restrictions on consent can be expressed as follows in OWL 2:[3]

```
ObjectUnionOf(
    ObjectSomeValueFrom( sbpl:hasDuty getValidConsent )          (2.2)
    ObjectSomeValueFrom( spl:hasData svd:Anonymous )             (2.3)
    ObjectSomeValueFrom( spl:hasPurpose svpu:Historical )        (2.4)
    ObjectSomeValueFrom( spl:hasPurpose svpu:Law )               (2.5)
)
```

The above compound class contains all the business policies (processes) that ask for consent (2.2), plus those that operate on anonymous data only (2.3), plus those that process data for archiving purposes (2.4) and those that process data because it is required by the law (2.5). The analysis of the GDPR reveals further alternatives to consent request, that shall be included in the above union (this is done in the next section). Note that even if the above policy is incomplete, it addresses "correctly" the goal of automated compliance in the following sense: it accepts *only* (althoug not *all*) policies that obviously comply with the areas of the GDPR selected for automated verification (e.g. consent) and flag the other policies so that a human expert can evaluate them thoroughly.

If we call C the resulting class (formalizing the selected part of the GDPR), then a business policy (process) P complies with the GDPR's obligations about consent collection if it is a subclass of C (but the converse is not always true, due to the partial nature of compliance checking). In OWL 2 terms, we shall ask the inference engine whether the following inclusion holds:

$$\texttt{SubClassOf(}\ P\ C\ \texttt{)}.$$

[3] Please consider that this is only a partial, simplified example – for illustrative purposes – that does not take into account all the possible legal grounds of the GDPR, but only those mentioned so far. The complete definition is given in the next section.

3.2 Restrictions on International Transfer of Data

We use international data transfer as an additional example of our formalization methodology. The GDPR states that data can be transferred across different countries as long as they belong to the EU. Data can be transferred to a non-EU country only if there are appropriate data protection guarantees. The GDPR mentions several options, including for example the following:

1. the regulations of the country to which data is transferred provide sufficient protection; we denote the class of such countries with the OWL 2 class `svl:EULike` (see the Location vocabulary in D1.1);

2. data remain within the data controller's boundaries (although in a branch residing in a different country) and the data controller adopts the binding corporate rules mentioned at the beginning of this section.

These two restrictions can be encoded in OWL 2 similarly to the restrictions related to consent. The OWL 2 class that contains all the business policies that comply with the above rules is the following:

```
ObjectUnionOf(
  ObjectIntersectionOf(
      ObjectSomeValueFrom( spl:hasStorage spl:Null )
      ObjectSomeValueFrom( spl:hasRecipient spl:Null )                    (2.6)
  )
  ObjectSomeValueFrom( spl:hasStorage
      ObjectSomeValueFrom( spl:hasLocation svl:EU )                       (2.7)
  )
  ObjectSomeValueFrom( spl:hasStorage
      ObjectSomeValueFrom( spl:hasLocation svl:EULike )                   (2.8)
  )
  ObjectIntersectionOf(
      ObjectSomeValueFrom( spl:hasRecipient svr:Ours )
      ObjectSomeValueFrom( sbpl:hasDuty BindingCorpRules )                (2.9)
  )
```

In particular, the above class contains all the business policies (processes) that satisfy some of the following conditions:

- data is neither stored nor made accessible to any recipient (2.6);
- data remains within the EU or in countries with comparable data protection guarantees (2.7) (2.8);
- data remains within the controller's boundaries and is protected by binding corporate rules (2.6).

Let us call the above OWL 2 union class T. In order to check whether a business policy P complies with the above rules it suffices to check whether

$$\texttt{SubClassOf(}\ P\ T\ \texttt{)}$$

is entailed. In all the other cases, the intervention of a human is required to check whether P complies with the GDPR.

The compliance checks illustrated in Sec. 3, if successful, should guarantee that the given business policy is compliant with the GDPR. Failed checks may either indicate a real problems or be due to the intrinsic limitations of automated verification (cf. D1.3). This is why, in general, human intervention is required after a failure is notified. In order to guarantee this kind of "correctness" of the compliance verification procedure, it is essential to classify the data sources correctly (by specifying which data categories they contain), the software (i.e. which category of processing it performs), the purpose, the storage, and the recipients, because – obviously – an incorrect description of the business policy makes verification unreliable.

4 The current approach to GDPR formalization and compliance checking

This section has been written while SPECIAL's formalization of the GDPR is being refined and assessed. For this reason, we are not presenting the full axiomatization, nor are we covering all the aspects of the GDPR that will eventually be checked for compliance by SPECIAL's components. We shall only outline

- the scope and nature of the automated compliance check that is the ultimate goal of the formalization of the GDPR (i.e. what should be expected from the compliance tool);
- the prerequisites of automated compliance checks (i.e. what kind of preliminar activities are required in order to instantiate the compliance tool in a novel application domain);
- the methodology being followed to identify the relevant parts of the GDPR and formalize them (that should convey the flavor of the final axiomatization).

We include some illustrative examples along the way. Please consider that the vocabularies adopted in these examples are not final and are still subject to refinements and assessment.

4.1 New properties for business policies and the activity record (Art. 30)

Recall that business policies extend usage policies with property **`hasDuty`** that associates the specified processing activity with the obligations prescribed by the GDPR. This is not the only useful additional property that should be encoded in business policies.

Article 30 of the GDPR requires controllers and processors to keep a record of the processing activities under their responsibility. Such records significantly overlap the old notion of business policies, as they should contain:

- The information contained in usage policies, cf. points (b)–(f) of Art. 30(1) and points (b) and (c) of Art. 30(2).

The additional properties required by Art. 30 comprise:

- Further annotations, such as: (i) data about the controllers, the processors, their representatives, and the data protection officer (when applicable), cf. point (a) of Art. 30, paragraphs 1 and 2.

- In case of transfers regulated by Art. 49(1), the documentation of the appropriate safeguards in place.
- If possible, a general description of the security measures referred to in Art. 32(1).

All of the above properties can be encoded in OWL by attaching additional object properties to business policies. The range of these properties – and in particular the granularity of their representation – will be the subject of future work, that will adopt the criteria illustrated in Section 4.3.2 below. The options range from textual annotation to a taxonomy of safeguards and security measures. The description of controllers, processors, and officers may leverage some existing proposals of vocabularies for representing organizations.

> Once these properties are added to the business policy vocabulary, the development of business policies can also address the obligation to keep a record of the processing activities as per Article 30.

Moreover, the ongoing discussions with the pilot leaders and within the W3C community group[4] are highlighting the usefulness of including in the business policy the legal basis for the processing. We encode this information by attaching to business policies a new object property called **`legalBasis`** . This property ranges over terms that directly relate to the appropriate articles in the GDPR, or to specific laws of the Union or Member States. For example, a business policy of the form:

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( hasData ...  )
   ...
   ObjectSomeValueFrom( legalBasis  Art6_1_a_Consent )
   ObjectSomeValueFrom( hasDuty  getValidConsent )
)
```
(2.10)

asserts that processing is carried out under Art. 6(1).a, that is, "*the data subject has given consent to the processing of his or her personal data for one or more specific purposes*".

4.2 Scope and limitations of GDPR's formalization and related tools

Caveats and limitations:

- *The extended business policies outlined in the previous section shall be authored by humans, who have the responsibility of assessing the correspondence between the formal policy and the real processing*, including related obligations. For example, in policy (2.10), the clause

  ```
  ObjectSomeValueFrom( hasDuty  getValidConsent )
  ```

  requires the policy author(s) to verify that consent is actually acquired in a lawful manner, satisfying all the requirements laid out by the GDPR including those in Art. 4(11) (i.e. consent should be freely given, specific, informed, unambiguous, and so on). Clearly such

[4] https://www.w3.org/community/dpvcg/

requirements cannot be assessed by a machine so they lie outside the scope of GDPR's axiomatization and related tools.[5]

Another example: if the policy fails to mention part of the data involved in the processing, then an automated compliance check may erroneously label the policy as compliant. For instance, if the policy does not mention that processing involves sensitive data (cf. Art. 9), then the reasoner may fail to notice that the stronger conditions introduced in Art. 9(2) should be applied. It is not possible to automatically infer the complete set of data categories involved in the processing; tools such as the personal data catalogue being developed by TenForce support humans in finding personal data in a complex system – thereby reducing the probability of overlooking important data – but they can give no 100% guarantee that some personal data is missed, just like any AI algorithm.

In summary, the reliablility of the compliance checking tools depends on the correctness of the given business policy, that can only be assessed by a human. As pointed out in the previous section, the effort of producing correct business policies is also required by Article 30.

- For similar reasons, *an automated reasoning tool cannot reliably infer the legal basis of processing from the other properties* (such as `hasData`, `hasPurpose`, and so on). Consequently, the responsibility of choosing a correct legal basis ultimately relies on policy authors, and the **`legalBasis`** property of business policies is only meant to document what the controller claims to be the legal basis of the processing.[6]

What can be done with SPECIAL's formalization of the GDPR:

- It is possible to verify that the different properties of the given business policy are coherent with each other, thereby preventing some possible human errors. For example,
  - if the legal basis is consent, as in policy (2.10), then the policy should contain the corresponding duty `getValidConsent`;
  - if data are personally identifiable, then the policy must contain the duties associated to data subjects' rights, such as those in Art. 12–22 and Art. 33–34;
  - if data are sensitive or involve criminal records, then the legal basis should be based on Art. 9 or Art. 10, respectively;
  - if storage or recipients are located in a third country that does not occur in the list as per Art. 45, and the recipient is an organization that does not occur in that same list, then ordinary consent does not suffice and additional requirements apply, see for example Art. 46 of the GDPR.

  In summary, GDPR's formalization helps in verifying the internal coherency of the business policies and can spot some human errors.

[5] What can be formalized is metadata about consent, such as when it was given, who gave it, et cetera. Some of these properties are already supported by the SPLog vocabulary for expressing the entries of the transparency ledger. More properties may be identified by the ongoing discussion in the W3C community group.

[6] Actually the **`legalBasis`** properties is also used to trigger appropriate verifications of compatibility with the other policy properties, such as data categories, purposes and so on. This is something that can be automated to some extent.

- *The compliance checker can also spot missing information*, that has not been provided by policy authors. For example if the data involved in the processing are classified neither as sensitive nor as non-sensitive, and the legal basis does not refer to Art. 9(2), then the compliance check fails, because it cannot verify whether the requirements of Art. 9 are fulfilled. Then the attention of the policy authors is drawn to the missing classification in terms of sensitivity, which leads them to filling in the gaps in the domain-specific ontology or in the policy.[7]
- A related benefit is that *once the data have been correctly classified, new processing activities on the same data re-use the classification*. For example, when data are re-purposed, knowing that the data does or does not fall within the special categories referred to by articles 9 and 10 automatically triggers on or off the verification of the additional requirements prescribed by those articles.

Summarizing and discussing the above points: When SPECIAL's framework is instantiated in a new application domain for the first time, it may be necessary to extend SPECIAL's vocabularies by introducing subclasses of some of their terms (such as domain-specific data categories, purposes and processes, that are special cases of the general terms occurring in SPECIAL's vocabularies). Care must be placed in avoiding misclassifications, especially in terms of anonymity and of the special data categories referred to by Art. 9 and 10, because misclassifications may lead to erroneous compliance assessments. On the other hand, if the vocabulary extension is missing relevant information, then the formalization is conceived in such a way that the compliance check fails and draws the attention of policy authors to the gap in the vocabulary. That is, the tools for checking compliance with the GDPR are sensitive to misclassification, but robust with respect to incomplete classification. The effort required for extending the vocabularies is needed only once; the extended vocabularies are re-used each time a business policy is updated or a new business policy is introduced. The initial classification effort specializes once and for all the legal bases applicable to those data, thereby reducing the possibility of future human errors.

- There is another way of using the compliance checker and its explanation system. *One can submit an incomplete business policy to the checker and get suggestions on how it could be completed to become compliant*. Then policy authors can see whether some of the suggested options match the real scenario and assess their viability. This may be useful to support and speed up the design of new business activities. For example, suppose that a controller C is planning to process location data contained in a repository of device profiles, and suppose that such data are denoted by an application-specific term `DevProfile_C` that, after a careful analysis, has been classified as a subclass of `Anonymous`, `NonSensitive`, and `NonCriminal`. The incomplete business policy specifying the above information, that is

  `ObjectSomeValueFrom( hasData DevProfile_C )` (2.11)

  is then fed to the system. The answer is that (2.11) is compliant, due to the anonymous (hence non-personal) nature of the data that places this kind of activity outside the scope

[7] In turn such correction actions may trigger further revisions of the controller's activity, possibly including reformulations of the legal basis of data processing.

of the GDPR. This means that the data about device profiles can be lawfully processed (with respect to the GDPR) no matter how the missing properties of business policies are filled in (i.e. purposes, processing, recipients, storage); moreover, processing is subject to no obligations. Next suppose that the same data is classified as `PersonalData` instead of `Anonymous`. Then compliance fails. The diagnostic engine notifies the authors that compliance needs to be supported by one of the legal basis specified by Art. 6. Since data are classified as `NonSensitive` and `NonCriminal`, the diagnostic engine does not mention the refined legal basis of Art. 9 and 10. Suppose that after this first round, policy authors decide to select consent as a legal basis. Under the hood, this amounts to submitting to the checker a slightly extended incomplete policy, obtained by adding the legal basis to (2.11); we also assume that policy authors remember that a corresponding duty is required and insert it in the incomplete policy:

```
ObjectIntersectionOf(
   ObjectSomeValueFrom( hasData  DevProfile_C )
   ObjectSomeValueFrom( legalBasis  Art6_1_a_Consent )
   ObjectSomeValueFrom( hasDuty  getValidConsent )
).
```
(2.12)

Then the diagnostic engine notifies the authors that several required duties are not reported in the policy (including the support of user rights as per Chapter 3, and the notification, protection, and risk assessment obligations required by Chapter 4 of the GDPR). By repeating similar iterations policy authors are guided in the compilation of the new business policy.

Note that this process is not equivalent to a simple decision tree since it uses the ontology that extends the vocabularies to specialize decisions based on the classification of domain-specific terms. A possible way of phrasing this is: the ontology automatically "selects" an ad-hoc decision tree from a potentially infinite family of trees.

Of course policy authors are not meant to see the internal encoding of business policies illustrated in (2.11) and (2.12). What they should see is a GUI where they can associate a vocabulary term (taken from a domain-specific taxonomy) to the policy properties of their choice (so as to be able to compile an incomplete business policy). Then what they see after each iteration is either the message "the policy is compliant" or a highlighting of the missing parts. The set of options for the missing properties is automatically restricted according to the applicable values identified by the diagnostic engine based on the available information (e.g. data, purpose, legal basis, just to name a few), and based on the domain-specific ontology.

Another advantage of the above knowledge-based approach is that as the Union and Member States add new laws, and as the GDPR evolves, it suffices to add or update clauses in the formalization to adapt the entire system to the new legal requirements (because no legal knowledge is cast into the code of the reasoner and of the diagnostic engine).

It should not be forgotten that after each iteration, the choice of an available option must be subject to human assessment to guarantee that it matches both the given application scenario and all the legal requirements that cannot be automated. The system can help in this phase by producing a list of pointers to the applicable articles and laws that must be taken into account by the human responsible (cf. the use of OWL annotations discussed in the following).

4.3 The methodology for formalizing the GDPR

As in the previous version, reported in Section 3, we are encoding the GDPR in OWL in such a way that a business policy `BP` complies with the GDPR's encoding `ENC` if and only if:

1. `BP` is not internally contradictory (i.e. not a subclass of `owl:Noting`), and
2. `SubClassOf(BP ENC)` holds, that is, `BP` is a subclass of `ENC`.

4.3.1 Scope

The formalization covers only the aspects that concern controllers and processors, in terms of requirements on processing and obligations. The obligations and guidelines for the Union and Member States, and for the data protection authorities lie outside the scope of the formalization. This is because the tools for checking compliance with the GDPR are being designed primarily for controllers and processors, as a support to business process design and validation.

Nonetheless the requirements in Chapter 3 of the GDPR (that concern the rights of the data subjects) are taken into account since handling user rights requires controllers and processors to set up suitable business processes to meet the response times prescribed by law.

4.3.2 Granularity

We do not articulate in detail those requirements that cannot be checked, derived, or analyzed automatically, or that are always dealt with together, in a uniform way.

For example, the requirements on consent reported in Art. 4(11), that are expressed in vague terms, are not articulated and are lumped together with the obligation to obtain consent encoded by the class `getValidConsent`.

For documentation purposes, the articles that are not encoded in OWL's logic are associated to the classes they affect as human-readable annotations. For example, OWL2 allows to declare annotations with statements like

```
AnnotationProperty( checkBeforeUse )
```

that can be used in classes and properties to refer policy authors to the relevant articles, as in the following example

```
Declaration(
   Annotation( checkBeforeUse  "Art.4(11), Art.7, Art.8" )
   Class(getValidConsent)
).
```

As a second example, consider the obligation to support user rights introduced in Chapter 3 of the GDPR. Articles 12–22 are collectively represented by a single class

```
Art12-22_SubjectRights
```

since either data are not personal (and those rights need not be supported), or the controller is obliged to support them all. So it would be useless to introduce a separate class for each of them (access, rectification, erasure, and so on). Again, the meaning of the concept can be documented by means of annotations, as in the following example:

```
Declaration(
   Annotation( checkBeforeUse  "Use this class in hasDuty only if
     all the data subject rights as per Art.12-22 are supported")
   Class(Art12-22_SubjectRights)
).
```

In other cases, separate classes have been adopted in order to record important details of the business policies. For example, we followed this approach with the legal basis of processing. Recall that one of the requirements stemming from the discussions with the pilot leaders and with the W3C community group is that processing activities should be labelled with their legal basis for future reference and monitoring; accordingly we introduced a property **`legalBasis`** whose use has been illustrated with policy (2.10). In order to record the exact legal basis for the processing, we use a set of classes corresponding to letters *a–f* of Art. 6(1):

```
Art6_1_a_Consent
Art6_1_b_Contract
Art6_1_c_LegalObligation
...
```

where each of the above classes is annotated with the definition provided in Art. 6. A similar approach has been followed in formalizing Art. 9.

4.3.3 GDPR-related vocabulary terms

The GDPR places stronger constraints on some kinds of processing, that involve particular categories of data or particular purposes (cf. Art. 9 and 10, for example). Then it is useful to introduce classes that formalize such data categories and purposes. For example, we introduce the new data categories:

```
SensitiveData, NonSensitiveData,
CriminalData, NonCriminalData.
```
(2.13)

These classes can be used in conjunction with classes like `Demographic` and `Location` in order to fully describe the nature and contents of the data. For example

```
ObjectIntersectionOf( SensitiveData Demographic )
```

denotes the category of demographic data that are sensitive as per Art. 9. Moreover, when the compliance checker is instantiated in a specific application domain – say, e-health – one can extend the vocabularies to state once and for all that particular data categories are sensitive. For example a domain-specific assertion

```
SubClassOf(
   ObjectIntersectionOf( Medical PersonalData )
   SensitiveData
)
```

would state that personally identifiable medical information is sensitive. Note that the above assertion would *not* make anonymous medical information sensitive. We are currently evaluating whether it is worth including classes of purposes with a similar role as the data categories (2.13).

4.4 Modelling top-level requirements

The GDPR is organized in chapters that deal with lawfulness of processing, data subject rights, obligations of controllers and processors, and so on. SPECIAL's formalization introduces a class for each of such "top-level" requirements that concerns the business policies of controllers and processors, and in particular for chapters 2, 3, 4, 5, and 9. So the class that formalizes the "relevant" part of the GDPR, called

`GDPR_Requirements`,

is defined as the intersection of the top-level requirements, unless data are not personal (in that case processing is not subject to the GDPR) or specific derogations apply as per Chapter 9.[8] Accordingly, `GDPR_Requirements` is asserted to be equivalent to:

```
ObjectUnionOf(
   OutOfGdprScope_as_per_Art2
   ObjectIntersectionOf(
     Chap2_LawfulProcessing
     Chap3_RightsOfDataSubjects
     Chap4_ControllerAndProcessorObligations
     Chap5_DataTransfer
   )
   Chap9_Derogations[a]
)
```

[a] Chapter 9 may also introduce restrictions, that are formalized with other terms. In other words, this class referso only to the part of Chapter 9 that deals with derogations.

Informally speaking this means that, in order to be compliant with the GDPR, a business policy must either involve a kind of processing that does not fall under the GDPR (e.g. involving non-personal data only), or it should satisfy all the requirements encoded in the classes that formalize chapters 2–5, or it should be subject to the exceptions formalized by `Chap9_Derogations`.[9]

All of the aforementioned classes have a suitable definition. Here we report only some of them in order to illustrate the formalization strategy.

The class `Chap2_LawfulProcessing` is meant to contain all the business policies that satisfy the articles in Chapter 2 of the GDPR with particular regard to legal bases (hence the name "lawful processing"). In turn, this class is defined in terms of classes that contain the policies that satisfy articles 6, 9, 10, or the opening clauses introduced in Chapter 9 (provisions related to specific processing situations).

[8] An example of such exceptions is provided by Art. 85 that allows Member states to relax the requirements of Art. 6 in order to reconcile data protection with freedom of expression

[9] Note that in OWL2 ObjectUnionOf and ObjectIntersectionOf play a role similar to logical 'or' and 'and', respectively.

In particular, each business policy should have a legal basis among those specified by Art. 6. Additionally, if the data involved in the processing are sensitive, then processing should be allowed by some of the legal bases in Art. 9. If criminal records are processed, then the additional restrictions of Art. 10 apply. This set of requirements is encoded in OWL2 by asserting that `Chap2_LawfulProcessing` is equivalent to:

```
ObjectIntersectionOf(
   Art6_LawfulProcessing
   Art9_SensitiveData
   Art10_CriminalData
)
```

In turn, `Art6_LawfulProcessing` is defined as

```
ObjectUnionOf( Art6_1 Art6_4 )
```

that is, either the fundamental legal bases of Art. 6(1) apply, or the processing is compatible with the original purpose for collecting the data as per Art. 6(4).

In order to capture this meaning, class `Art6_1` is defined as:

```
ObjectSomeValueFrom( legalBasis
   ObjectUnionOf(
     Art6_1_a_Consent
     Art6_1_b_Contract
     Art6_1_c_LegalObligation
     Art6_1_d_VitalInterest
     Art6_1_e_PublicInterest
     Art6_1_f_LegitimateInterest
   )
)
```

Roughly speaking, this definition means that a business policy satisfies the requirements of Art. 6(1) if it contains a clause

`ObjectSomeValueFrom( legalBasis` X `)`

where X is some of the above six classes corresponding to points *a–f* of Art. 6(1). More precisely, it suffices that the above clause be *inferred* from the available knowledge.

The policy author has the responsibility to verify that the chosen legal basis X actually applies to the real processing, by taking into account all the conditions that cannot be formalized in logic. For example, paragraphs 2 and 3 of Art. 6 (that concern legal bases (c) and (e)) are not formalizable. However we draw the attention of the policy author to those articles by means of annotations to the corresponding classes, for instance:

```
Declaration(
   Annotation( checkBeforeUse  "Art. 6(2), Art. 6(3), Art. 48")
   Class(Art6_1_c_LegalObligation)
).
```

Article 9 – when applicable (i.e. when data are classified as sensitive) – is dealt with by attaching to the business policy a second legal basis corresponding to some of the points *a–j* of Art. 9(2). The axiomatization of `Art9_SensitiveData` is similar to the definition of `Art6_1` (see above); it uses a new list of concepts to denote the new legal bases. In some cases such legal bases are special cases of the bases in Art. 6, for example the axiomatization comprises the assertion:

```
SubClassOf( Art9_2_a_Consent Art6_1_a_Consent )
```

that states that Art. 9(2).a is a special case of Art. 6(1).a and strengthens it (for example, Art. 9(2).a further requires consent to be *explicit*). In these cases, it suffices to attach to the business policy the legal basis of Art. 9, because the corresponding legal basis of Art. 6 is inferred automatically (as if it were automatically added to the business policy) .

Article 10, instead, does not provide new legal bases. It rather adds some further contraints when the processing involves criminal records – in particular the controller must be suitably supervised. The corresponding concept `Art_10_CriminalData` is defined as:

```
ObjectUnionOf(
   SomeValueFrom( hasData  NonCriminalData
   ObjectIntersectionOf(
       SomeValueFrom( hasDuty  Art10_Requirements
       RefinementsAsPerChap9
   )
).
```

Informally speaking, the above definition says that either the processing does not involve criminal data, or the controller must satisfy the additional requirements laid out by Article 10 and, moreover, any further restrictions introduced by Chapter 9.

Many articles introduce constraints over different properties of the business policy. Some of these constraints can be formalized and automatically checked. For example, Article 9(2).d is applicable as a legal basis only under some conditions, e.g. that personal data are not disclosed outside the data controller without the consent of the data subject. This can be formalized as follows with a class inclusion axiom:

```
SubClassOf( SomeValueFrom( legalBasis  Art9_2_d )
   ObjectIntersectionOf(
       SomeValueFrom( hasRecipient Us )
       SomeValueFrom( hasStorage
           SomeValueFrom( hasLocation ControllerServers ) )
   )
).
```

The effect of this axiom is that if a business policy has Article 9(2).d as its legal basis, and the encoded processing transfers or stores data beyond the data controller's boundaries, then the policy is flagged as non-compliant (because it becomes incoherent by the above axiom). If the data subject gives her consent to the transfer then the controller must introduce a separate business policy dedicated to the transfer and including the clause:

```
ObjectSomeValueFrom( legalBasis  Art9_2_a_Consent )
```

since the legal ground for transfer, in this case, is consent.

The other articles and chapters of the GDPR are addressed in a similar way, by analogy with one or more of the examples of encodings illustrated above. The complete formalization and the associated documentation will be published in year 3 under WP3 and WP5.

Chapter 3

Appendix

In this appendix we recall the first version of the ontology for business policies, introduced in D2.2. Please consider that the following definitions and axioms are not stable. We are currently turning the ontology into a stable artifact, and in this process we may change the namespaces, deprecate some of the terms, and add new terms and properties, along the lines described in this deliverable. None of these changes affects the language's *structure*.

1 The Business Policy Ontology V1

```
Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
Prefix(rdf:=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(spl:=<http://www.specialprivacy.eu/langs/usage-policy#>)
Prefix(xml:=<http://www.w3.org/XML/1998/namespace>)
Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
Prefix(rdfs:=<http://www.w3.org/2000/01/rdf-schema#>)
Prefix(sbpl:=<http://www.specialprivacy.eu/langs/business-policy#>)

Ontology(<http://www.specialprivacy.eu/langs/business-policy/>
<http://www.specialprivacy.eu/langs/business-policy/1.0>

Import(<http://www.specialprivacy.eu/langs/usage-policy/1.0>)

Declaration(Class(sbpl:AnyDuty))
Declaration(ObjectProperty(sbpl:hasDuty))

ObjectPropertyDomain(sbpl:hasDuty spl:Authorization)
ObjectPropertyRange(sbpl:hasDuty sbpl:AnyDuty)

)# end of ontology
```