SPECIAL

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

Deliverable D2.2

Formal representation of the legislation V1

Document version: V1.0

SPECIAL DELIVERABLE

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

Mr Philippe Rohou t: +33 4 97 15 53 06 f: +33 4 92 38 78 22 e: philippe.rohou@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-M12
Deliverable number:	D2.2
Deliverable title	Formal representation of the legislation V1
Contractual Date of Delivery:	31/12/2017
Actual Date of Delivery:	27/12/2017
Editor (s):	Piero Bonatti (CeRICT), Sabrina Kiranne (WU)
Participant (s):	P.A. Bonatti, S. Kirrane, I. Petrova, L. Sauro, C. Kerschbaum, E. Schlehahn, R. Wenning
Reviewer (s):	Eva Schlehahn (ULD), Rigo Wenning (ERCIM/W3C)
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):	26

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 future versions of this deliverable.

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 as 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 initial thoughts on the the formal representation of the GDPR, which will be iteratively refined throughout the course of the project.

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 D2.1 Policy Language V1*, additional joint work with the pilot leaders is needed in order to detail the vocabularies for the policy language and the contextual information required for legal interpretation. All examples currently included in the paper are generic and as such are only meant to illustrate the policies' structure.

Additionally, further iterations of the policy language will take on-board the results of SPECIALs standardisation activities, especially concerning vocabularies to model privacy policies, regulations, and the involved (business) processes. 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 the subset of articles from that GDPR 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

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	a limitation or restriction (e.g. its allowed if)
	−C	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[]	an article contains an explicit reference (e.g. eRef[Art. 89 (1)])
	tRef[]	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)

that must be fulfilled by companies), Constraints (used to constrain the obligation) 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:` (R4)
(C) `the data subject has given consent to the processing of his or her personal data for one or more specific purposes;`

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,` (R5)
(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.`

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` (R6)
(C) `in relation to the data subject`

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       (R8)
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,
(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          (R11)
conditions, including limitations, with regard to the
processing of genetic data, biometric data or data
concerning health.
```

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      (R12)
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.
```

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)

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 ad-hoc 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 Miimal Core Model (MCM) implemented by SPECIAL's *usage policy language* (UPL) introduced in D2.1. In this respect, 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 described in the abstract terms adopted by the processing vocabulary introduced in D2.1; in a business policy, this is specified by naming concrete software procedures;
- the purpose of data processing may be directly related to the data controller's mission and products;
- recipients 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 `GetConsent` 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 usage policies, that have been outlined in D2.1.

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 getConsent)
    ObjectSomeValuesFrom(sbpl:hasDuty getAccessReqs)
    ObjectSomeValuesFrom(sbpl:hasDuty getRectifyReqs)
    ObjectSomeValuesFrom(sbpl:hasDuty getDeleteReqs)
)
```

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.

[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 the next versions of the deliverable*

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

This section is being written while the GDPR is analysed for extracting the list of obligations that controllers shall fulfil. Since this analysis is not available at the moment of writing, we are not going to formalize all the aspects of the GDPR that will eventually be checked for compliance by SPECIAL's components. We shall only show how some selected requirements of the GDPR can be encoded in OWL 2. We are confident that similar approaches can be extended smoothly to the other obligations of the GDPR subject to automated checking.

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 always mandatory (there are exceptions, discussed later), however *if* valid consent is available, *then* processing is legal, even if consent were useless in the particular case at hand.

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).

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 is not necessary if data are processed because law requires it, or for archiving purposes. The latter is captured by the class `svpu:Historical` of the purpose vocabulary, while the former calls for an additional class `svpu:Law`.

All these restrictions on consent can be expressed as follows in OWL 2:

```
ObjectUnionOf(
    ObjectSomeValueFrom( sbpl:hasDuty getConsent )                (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). If the analysis of the GDPR will reveal further exceptions to consent request, those exceptions shall be included in the above union. This addresses correctly the goal of automated compliance, namely, accepting the 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

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

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 (at least) two such possible guarantees:

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 T the above OWL 2 union class. 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. In order to guarantee this kind of "correctness" of the compliance verification procedure, it is essential to classify correctly the data sources (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.

Chapter 3

Appendix

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
```