Analysis of Ontologies and Policy Languages to Represent Information Flows in GDPR

Beatriz Esteves [a,*], Víctor Rodríguez-Doncel [a]
[a] *Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*
E-mail: beatriz.gesteves@upm.es

Abstract. This article surveys existing vocabularies, ontologies and policy languages that can be used to represent informational items referenced in GDPR rights and obligations, such as the 'notification of a data breach', the 'controller's identity' or a 'DPIA'. Rights and obligations in GDPR are analyzed in terms of information flows between different stakeholders, and a complete collection of 57 different informational items that are mentioned by GDPR is described. 12 privacy-related policy languages and 9 data protection vocabularies and ontologies are studied in relation to this list of informational items. ODRL emerges as the language that can partially represent the highest number of rights and obligations in GDPR if complemented with DPV and GDPRtEXT, since 39 out of the 57 informational items can be modelled. Online supplementary material is provided, including a simple search application and a taxonomy of the identified entities.

Keywords: privacy policy languages, data protection ontologies, GDPR, rights, obligations

1. Introduction

Westin [1] shaped the way we define online privacy before the web existed at all. One of his two major postulates was that individuals should be able to determine to what extent information about them is communicated to others. The second of these postulates was that technological artifacts could be used to achieve this goal. His books in the late sixties and the seventies exerted significant influence on the privacy legislation that was enacted in the following years, and even today, the European General Data Protection Regulation (GDPR), which came into full effect on May 25th of 2018, owes much to his work. Any information system has data representation needs, and privacy and data protection related information systems will have to represent ideas such as 'consent' or 'the right to erasure'. If these applications are to interoperate, then the need for standard formats is clear, and the adoption of semantic-web enabled technologies that facilitate privacy-related data exchange is advantageous such as in data portability.

Machine readable policy languages have been on the scene for some decades. Policy languages allow to represent the will of an individual or organization to grant access to a certain resource, and they govern the operation of actual systems over actual data. They seem perfectly aligned with Alan Westin's dreams and indeed several privacy-related policy languages have been defined and used in real scenarios. On the other hand, computers can also help in other privacy and data protection tasks different from enforcing access to personal data, and policy languages are not enough to cover every representation need. Thus, in the last few years, vocabularies and computer ontologies have appeared to formalize concepts and rules in the domain that can be used either to simply represent information as RDF, or to govern ontology-based information systems. Not all of them, however, had the GDPR specifically as the reference framework.

This paper surveys existing policy languages, vocabularies and ontologies in the domain of privacy and data protection, and it analyses their adequacy to support GDPR-related applications. These GDPR-related applications may either support individuals to manage their personal information or to support data controllers, data processors and other stakeholders to bet-

*Corresponding author. E-mail: beatriz.gesteves@upm.es.

1570-0844/0-1900/$35.00 © 0 – IOS Press and the authors. All rights reserved

ter manage compliance with the GDPR. This joint analysis of needs (individual-oriented and company-oriented) is based on the claim that these tools may converge in a near future, and that having common vocabulary elements and common data models to refer to GDPR rights and obligations and to denote specific GDPR concepts would permit heterogeneous applications to speak in the same terms and interoperate.

The main contributions of this paper are: (i) a study of GDPR in terms of flows of information in different deontic modalities and (ii) a survey of 21 existing vocabularies, ontologies and policy languages and their analysis in relation to that informational model. As supplementary material, a machine-readable extract of this analysis is provided[1].

The paper is organized as follows: Section 2 describes in detail the types of information that have to be shared between data subjects, controllers and other interested parties, as well as the main rights and obligations found in the GDPR that may be represented. Section 3 systematically reviews the existing privacy-related policy languages first, and then the most salient vocabularies and ontologies in the domain. Section 4 provides an analysis of the solutions in the light of GDPR, following a systematic comparison framework, and the description of the supplementary webpage which has been published with additional resources about the reviewed solutions, a REST API service to look for specific concepts and a vocabulary with the concepts identified in Section 2. Finally, the last section synthesizes our conclusions, explicitly identifying the recommendations and possible representational needs that have to be covered.

2. Information flows in the GDPR

In the light of the established GDPR rights and obligations, a set of information flows, related to the information that needs to be exchanged between stakeholders, can be identified. These stakeholders can be classified as a (DS) data subject, a (DC) data controller, a (DP) data processor, a (Rp) recipient, a (SA) supervisory authority or a (DPO) data protection officer.

In this context, an information flow refers to the information that has to be transmitted from one stakeholder to another so that a right or obligation can be invoked and granted. For instance, if a data subject invokes its right to erasure, along with the request, there is the need to represent information related to the ground on which the request is based, and the controller needs to transmit this information to the other controllers processing the same personal data.

Figure 1 shows a diagram of the information flows that represent the transfer of information foreseen by GDPR's rights and obligations regarding data subjects, controllers and other stakeholders.

Therefore, in this section, GDPR rights and obligations are studied with the purpose of assessing which informational elements need to be represented in order to support the flow of information between GDPR stakeholders. In particular, we shall emphasize the need to support Articles 13 and 14 of the GDPR, which describe the so-called *'right to be informed'*. According to these articles, whether personal data is collected directly from the data subject or obtained through other data sources, data controllers need to inform data subjects about any processing of personal data so that their activities are legal, fair and transparent. These articles, and the others that make up Chapter III of the GDPR, are studied here in order to understand what information data subjects are entitled to receive in the exercise of their rights and, correspondingly, what information data controllers need to disclose to be compatible with the GDPR. Sections 2.1 and 2.2 briefly describe these rights, as well as the informational items that may need to be represented.

The rights and obligations of controllers and processors, described in GDPR's Chapter IV, are also analyzed here for the same purpose of identifying which pieces of information need to be represented in order for these stakeholders to be in compliance with the GDPR. Section 2.3 details the informational elements and respective rights and obligations that may need to be modeled.

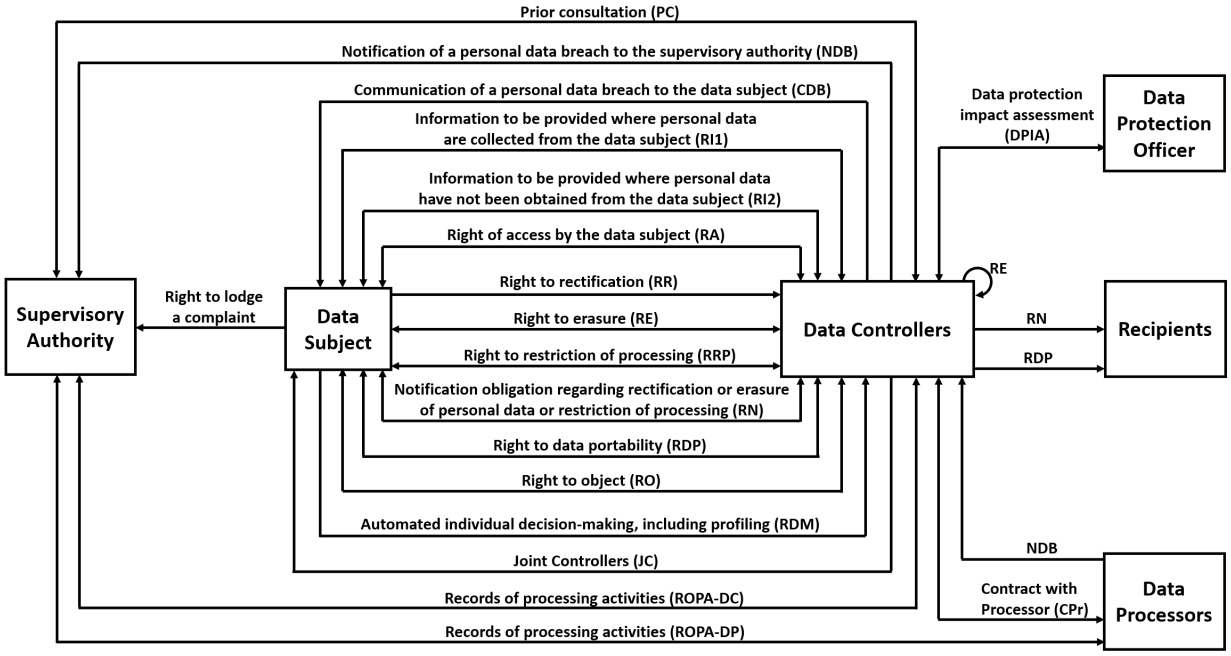
2.1. The Right to be Informed

Chapter III of the GDPR establishes nine fundamental rights of the data subject when it comes to the lawful processing of their personal data.

In particular, Articles 13 and 14 detail the *'Information to be provided where personal data are collected from the data subject'* (RI1) and the *'Information to be provided where personal data have not been obtained from the data subject'* (RI2), respectively. According to them, for the processing of personal data to be lawful, fair and transparent, a certain set of informational

[1] https://protect.oeg.fi.upm.es/sota/

Fig. 1. GDPR's rights and obligations as information flows. The bidirectional arrows represent a right or obligation in which a request for information and respective response is expected, while the unidirectional arrows represent only a request or notification and no reply is expected.



items must be provided, namely items I1 to I19 described in Table 1.

This information, and any other communications provided in the context of the provision of data subjects' rights, should be given in a concise, transparent and clear language and in an easily accessible manner. This information may also be provided with standardized icons for a more visible and intelligible overview of the intended processing.

2.2. Other data subject's rights

The data controller has the obligation to support the exercise of the data subject's rights and needs to reply with information to any requests related to the exercising of such rights within a month upon receiving the request. This period can be extended by a further two months if the data subject's request is too complex or in the case of a large number of requests. The information should be freely provided and by electronic means, unless the data subject states otherwise.

Apart from the *'right to be informed'*, already described in the previous section, the data subject is entitled to the following rights:

(RA) the *'right of access'* to the personal data being processed: data subjects have the right to receive confirmation that their data is being processed and a copy of the data in a common electronic format, as well as information about the purposes for processing, categories of the concerned personal data, their source, if not directly collected from the data subject, the recipients, the storage period, the existence of the data subject's rights as well as the right to lodge a complaint with a DPA, details of the existence of automated decision making and the security measures applied where personal data is transferred to a third party.

(RR) the *'right to rectification'*: the data subject has the right to obtain from the data controller the amendment of inaccurate personal data and, where the data is incomplete, the right to have personal data completed.

(RE) the *'right to erasure'* or *'right to be forgotten'*: the data controller has the obligation to delete personal data when it is no longer needed for the purposes which it was collected; when the data subject withdraws consent and there is no other legal basis for the processing; when the data subject objects to the processing; when said processing in unlawful; when it has to be erased to comply with a legal obligation; or when the data was collected for the provision of information society services.

Table 1

Informational items to be represented and respective identifiers (I*), which will be used to specify the informational elements necessary for the management of each right and obligation represented in Figure 1. The GDPR articles that mention these items are also specified.

I*	informational items - **GDPR Article(s)**	I*	informational items - **GDPR Article(s)**
I1	Controller identity - **13.1(a), 14.1(a), 30.1(a), 30.2(a)**	I30	Grounds to not comply with right not to be subjected to decision making - **22.2**
I2	Controller contact details - **13.1(a), 14.1(a), 30.1(a), 30.2(a)**	I31	Joint controller identity - **26, 30.1(a)**
I3	Controller's representative identity - **13.1(a), 14.1(a), 30.1(a), 30.2(a)**	I32	Joint controller contact details - **26, 30.1(a)**
I4	Controller's representative contact details - **13.1(a), 14.1(a), 30.1(a), 30.2(a)**	I33	Responsibilities of joint controllers - **26, 36.3(a)**
I5	DPO contact details - **13.1(b), 14.1(b), 30.1(a), 30.2(a), 33.3(b), 34.2, 36.3(d)**	I34	Subject-matter of the processing - **28.3**
I6	Purposes of the processing - **13.1(c), 14.1(c), 15.1(a), 28.3, 30.1(b), 35.7(a), 36.3(b)**	I35	Duration of the processing - **28.3**
I7	Legal basis of the processing - **6.1, 9.2, 13.1(c), 14.1(c)**	I36	Categories of processing - **28.3, 30.2(b)**
I8	Legitimate interests - **6.1(f), 13.1(d), 14.2(b), 35.7(a)**	I37	Categories of data subjects - **28.3, 30.1(c), 33.3(a)**
I9	Recipients / categories of recipients - **13.1(e), 14.1(e), 15.1(c), 17.2, 19, 30.1(d)**	I38	Obligations of the controller - **28.3**
I10	Transfers to third countries - **13.1(f), 14.1(f), 30.1(e), 30.2(c), 46, 47, 49.1**	I39	Obligations of the processor - **28.3**
I11	Retention period - **13.2(a), 14.2(a), 15.1(d), 30.1(f)**	I40	DPO identity - **30.1(a), 30.2(a), 33.3(b), 34.2**
I12	Data subject's rights - **13.2(b), 14.2(c), 15.1(e), 28.3**	I41	Technical and organizational security measures - **30.1(g), 30.2(d), 32.1, 35.7(d), 36.3(c)**
I13	Right to withdraw consent - **6.1(a), 9.2(a), 13.2(c), 14.2(d)**	I42	Processor identity - **30.2(a)**
I14	Right to lodge a complaint - **13.2(d), 14.2(e), 15.1(f)**	I43	Processor contact details - **30.2(a)**
I15	Statutory or contractual obligation details - **13.2(e)**	I44	Processor's representative identity - **30.2(a)**
I16	Existence of automated decision making - **13.2(f), 14.2(g), 15.1(h), 22.1, 22.4**	I45	Processor's representative contact details - **30.2(a)**
I17	Categories of personal data - **9.1, 14.1(d), 15.1(b), 28.3, 30.1(c), 33.3(a)**	I46	Nature of data breach - **33.3(a), 34.2**
I18	Source of personal data - **14.2(f), 15.1(g)**	I47	Approximate number of data subjects - **33.3(a)**
I19	Grounds to not comply with information right - **13.4, 14.5**	I48	Approximate number of personal data records - **33.3(a)**
I20	Safeguards related to the transfer to a third country - **15.2, 30.1(e), 30.2(c)**	I49	Consequences of personal data breach - **33.3(c), 34.2**
I21	Copy of personal data - **15.3, 20.1**	I50	Measures to address and mitigate data breach's effects - **33.3(d), 34.2**
I22	Request to complete incomplete personal data - **16**	I51	Systematic description of processing operations - **35.7(a)**
I23	Grounds to request erasure of data - **17.1**	I52	Assessment of the necessity and proportionality of the processing operations - **35.7(b)**
I24	Technical measures taken to erase data - **17.2**	I53	Assessment of the risks to the rights and freedoms of data subjects - **35.7(c)**
I25	Recipients contact details - **17.2, 19**	I54	Responsibilities of the controller - **36.3(a)**
I26	Grounds to not comply with right of erasure - **17.3**	I55	Responsibilities of the processors - **36.3(a)**
I27	Grounds to request restriction of processing - **18.1**	I56	Means of processing - **36.3(b)**
I28	Transfer data directly between controllers - **20.2**	I57	Data protection impact assessment (DPIA) - **35, 36.3(e)**
I29	Grounds to not comply with right to object - **21**		

(RRP) the *'right to restriction of processing'* of personal data: the data subject has the right to request the ceasing of the processing when the accuracy of the data is being contested; when the processing is unlawful and the data subject does not wish to erase the data; when the purposes stated by the controller are no longer valid but the data subject needs it for any legal claims; or when the data subject objects to the processing.

(RN) the *'right to be notified'* about the rectification, erasure or restriction of processing: the data controller has the obligation of notifying the data subject and the recipients to whom the data was disclosed, as well to disclose these recipients to the data subject.

(RDP) the *'right to data portability'*: the data subject has the right to receive its data in a commonly used and machine-readable format and has the right to request that its data be transferred directly from one controller to another.

(RO) the *'right to object'* to any processing, including profiling.

(RDM) the *'right to not be subjected to automated decision-making'*, including profiling.

The informational items to be granted to the data subject in function of the established GDPR rights are represented in Table 2.

Table 2

Informational items (I*) to be provided to the data subject, according to the rights (R*) defined under Chapter III of the GDPR.

Rights (R*)	Informational items (I*)
RI1	I1 to I17, I19
RI2	I1 to I14, I16 to I19
RA	I6, I9, I11, I12, I14, I16 to I18, I20, I21
RR	I22
RE	I9, I23 to I26
RRP	I25, I27
RN	I9, I25
RDP	I21, I28
RO	I29
RDM	I30

2.3. Rights and obligations of controllers and processors

Data controllers must be ready to demonstrate that their processing activities are in accordance with the GDPR and that they have in place the appropriate security measures to ensure people's right to privacy and data protection. These measures must take into account the nature, context and risks associated with each processing activity and should be embedded by design and by default in the data controllers' services.

The following rights and obligations must be observed by the data controllers and processors so that they can comply with the regulation:

(JC) the *'joint controllers'* responsibilities: in the case where there are two or more controllers determining the purposes and means of processing, they are joint controllers. They must determine the responsibilities of each controller in relation to the duties generated by the data subject's rights and this information should be communicated to the data subjects.

(CPr) contract with *'processors'*: the controller can establish contracts with processors, that have in place the appropriate security measures, for the processing to be carried out on behalf of them. This processing must be governed by a contract between controller and processor, that establishes the subject-matter, duration, nature and purpose of processing, as well as the personal data types, categories of data subjects and the rights of obligations of both the data controller and the data processor. The processor can only hire a sub-processor with the authorization of the controllers.

(ROPA-DC) *'records of processing activities'* of data controllers: each controller and its representative should keep a record of the processing activities under their responsibility, which must be available to the supervisory authorities when requested.

(ROPA-DP) the *'records of processing activities'* of data processors: each processor and its representative should keep a record of the processing activities carried out on behalf of a controller, which must be available to the supervisory authorities when requested.

(NDB) the *'notification of a data breach'* to the supervisory authority: the data controller has 72 hours to notify the competent supervisory authority that a personal data breach has occurred. The processor should inform the controller without delay as soon as it is aware of the breach.

(CDB) the *'communication of a data breach'* to the data subject: the data subjects have the right to be informed about any data breach that results on a high risk to their rights and freedoms. This communication should contain at least the nature of the breach and the measures that are being taken to mitigate it.

(DPIA) the *'data protection impact assessment'*: in the case where the data controllers are going to perform an extensive evaluation of personal data based on automated processing, processing activities over special categories of data or criminal data or a systematic monitoring on large scale, the controller should draft an assessment of the impact of the processing activities, and respective risks to the protection of personal data, with the guidance of the data protection officer.

(PC) the *'prior consultation'* right: the controller has the right to consult the supervisory authority, prior to the processing, when the DPIA illustrates that the processing activities will result in a high risk to the privacy of the data subjects if the proper measures to mitigate risks are not implemented.

The informational items that must be represented, in function of the rights and obligations of the data controllers and processors, are represented in Table 3.

Table 3

Informational items (I*) to be modelled, according to the rights and obligations of the controllers and processors, defined under Chapter IV of the GDPR.

Rights / Oblig.	Informational items (I*)
JC	I31 to I33
CPr	I6, I12, I17, I34 to I39
ROPA-DC	I1 to I6, I9 to I11, I17, I20, I31, I32, I37, I40, I41
ROPA-DP	I1 to I5, I10, I20, I36, I40 to I45
NDB	I5, I17, I37, I40, I46 to I50
CDB	I5, I40, I46, I49, I50
DPIA	I6, I8, I41, I51 to I53
PC	I5, I6, I33, I41, I54 to I57

3. Existing and related work

Some articles review the existing privacy-related policy languages, however, for the most part, they were published before the GDPR was enacted.

Kumaraguru et al. [2] provided a literature review on available privacy policy languages with the goal of developing a framework with metrics for their analysis. This framework classified languages based on the situations in which they could be used, also considering whether the policy language was user-centered or company-centered.

Kasem-Madani and Meier [3] produced a survey focused on security and privacy policy languages. The survey's goal is to present an overview of the existing solutions as well as providing a categorization framework to facilitate the adoption of policy languages. The main categories of the framework to classify the languages are the scope, syntax, extensability, context, type (focused on issues such as security, privacy or accountability), intention of use (user-centred, enterprise-centred or both) and usability (language oriented to humans or machines).

The most recent review work on privacy languages, by Leicht and Heisel [4], intends to provide a survey on languages in the context of privacy policies that can help users to easily understand them and that are compatible with data protection legislations such as the GDPR. Therefore, this framework identifies the criteria to compare the languages through the GDPR legislation. The identified criteria are system obligations, time constraints and formalization of the language.

Other review works related to the privacy and data protection domain have been published, namely overviews of access control frameworks, rights expression languages or other semantic approaches related to the representation of consent.

Kirrane et al. [5] provide an overview of access control models, such as the Mandatory Access Control (MAC), the Discretionary Access Control (DAC) and the Role Based Access Control (RBAC) models, and other RDF-based standards and policy languages frameworks. A collection of access control requirements is proposed and are used to categorize the described frameworks accordingly.

Pellegrini et al. [6] produced a preliminary survey on Rights Expression Languages (RELs). RELs are used to define machine-readable permissions, obligations and prohibitions, and are an essential component of any Digital Rights Management (DRM) system. This work also proposed a framework to classify RELs according to their application area in the DRM domain, namely for the purpose of specifying access and trust policies, license policies and contract policies.

Pandit [7] PhD thesis describes and analyses state of the art semantic-based technologies used to support and assess GDPR compliance, including privacy policy solutions, consent-related approaches and other solutions developed in the context of data privacy and data protection projects. The solutions are compared according to a set of categories, such as the representation of GDPR concepts, consent-related information or personal data handling activities, evaluation of GDPR compliance or resource accessibility.

3.1. Methodology

The main sources of information used to compile the existing efforts were academic publications and other Web specifications about languages, vocabularies and ontologies related to the data privacy and data protection domain. The collected results were reviewed and, if relevant for this analysis, included in this article. In the cases where the identified solutions were found to be developed within the framework of a project, the main goals and research directions of said project are described through information gathered on the project's website.

3.2. Privacy-related policy languages

In this subsection, we aim to identify privacy-related policy languages, describing the structure and information provided by each privacy policy as well as identify its compatibility with the GDPR to describe not only rights, but also obligations and duties.

A few languages, cited below, are left out of the review since they are considered to be obsolete or fall out of the context of this work. IBM Research's[2] Enterprise Privacy Authorization Language (EPAL) [8], and its predecessor Platform for Enterprise Privacy Practices (E-P3P) [9], are left out of this document since EPAL policies are mapped into Platform for Privacy Preferences (P3P) promises, to match the enterprises privacy policies with the users preferences, and World Wide Web Consortium (W3C) does not recommend further adoption of this language. Similarly, Bohrer and Holland [10] developed the Customer Profile Exchange (CPExchange) language, an XML specification for the transfer of customer data among enterprise services, implements P3P privacy policies applicable to the data that is being exchanged.

[2] http://www.research.ibm.com/

3.2.1. P3P

P3P, implemented by Cranor et al. [11], emerged as a specification for websites to disclose privacy protocols in a machine readable format so that web user agents could easily interpret them and notify the users about the decisions based on these practices. However, these mechanisms, that allow the user to be informed about the websites privacy policies in relation to its respective data collection, do not mean that the sites are actually implementing these policies since P3P does not provide a way to enforce them. Thus, the P3P vocabulary was not built to comply with a specific regulation but rather to establish the practices of each website.

The main contributions of the P3P specification are a P3P-based data schema for the data that the website intends to collect, a standard group of purposes, data categories and recipients and a XML standard to define privacy policies. The P3P policies are made up of general assertions and specific ones, called statements, that are related only to certain types of data. General assertions are constituted by the legal **entity** that applies the policy and an **access**, **disputes** and **remedies** elements. The access element expresses whether the website provides access to the data it collects. The disputes element provides a procedure for disputes on privacy practices, while the remedy specifies the possible solutions in case a policy breach happens. In relation to the statements applied only to specific data types, they are composed of the specified **data group**, that could contain one or more data elements, and of the **purpose**, **recipient** and **retention** elements. P3P defines a list of web relevant purposes for data processing, such as the completion and support of the activity for which the data was provided, research and development or individual analysis, and the purpose element should contain at least one of them. The recipient element should specify the beneficiaries of the collected data according to the recipient types established by P3P and the retention element must reflect the retention policy that covers the statement data.

As P3P was designed to express web services policies, A P3P Preferences Exchange Language (APPEL) by Cranor et al. [12] was developed as an extension of P3P so that users can express their preferences. Therefore both languages should be used in order to match the user's privacy preferences with the services' privacy policies.

The P3P 1.0 Specification became a W3C recommendation on April 16, 2002. However, it has had a limited implementation, since its use needs to be adopted by both Web services and users and, in addition, no protocol has been implemented for these P3P policies to reflect the actual privacy practices of the sites. Its status has turned to W3C obsolete recommendation on August 30, 2018 and thereby future implementations are not recommended.

3.2.2. Open Digital Rights Language (ODRL)

The ODRL Vocabulary & Expression 2.2 [13] is a W3C recommendation since February 2018, published by the Permissions & Obligations Expression (POE) Working Group (WG), being that its first version was released in 2001. The aim of this vocabulary is to define a language that can translate natural language policies to machine readable formats, providing information about permissions, prohibitions and duties related to an asset. This vocabulary is based on the merge of the previous work performed by the ODRL Community Group (CG), the ODRL V2.1 Common Vocabulary, the ODRL V2.1 XML Encoding, the ODRL V2.1 Ontology and the ODRL V2.1 JSON Encoding. ODRL is now supported and maintained by the POE WG, whereby new implementations should follow the deliverables of the WG.

Two vocabularies are used to describe ODRL: the ODRL Core Vocabulary and the ODRL Common Vocabulary. ODRL's Core Vocabulary main class is the **policy**, that allows for the identification of a particular policy using its unique identifier. Each policy may contain several **rules** - a rule is an abstract class that defines the common features of permissions, prohibitions and duties. **Permission** represents the concept of allowing an action related to an asset to take place, while the **prohibition** notion is related to the inability to execute the action. The permission may also be associated with a **duty** in the case where the action is mandatory. The rules are further refined by using **constraints** to determine the conditions in which the rule is applied, for example, to establish that a certain permission is only valid until the end of 2018. The **parties** (can be a group of people, an organization or an agent) that enforce the rules can take different roles, depending on their position in relation to the asset - a party that issues the rule takes on the assigner role, while the recipient of the rule is the assignee. An **asset** is an identifiable entity, such as data, software, services or even a collection of these resources, that is subject to a rule. The ODRL Common Vocabulary further specifies the policy sub-classes, the functions that can be exercised by the parties involved, the actions to which the rules apply and the temporal, spacial, sector, ... constraint

operands that can be set. Of particular interest in relation to the GDPR is the privacy policy subclass. This sub-class is related to policies that express rules over assets incorporating personal data. Therefore, the privacy policies that implement the ODRL language must inform the parties involved in which way the policy is being used and also with whom and for what purpose the policy is being shared with other parties.

ODRL has already been used in several contexts, for instance by the working groups on Open Mobile Alliance SpecWorks[3] and by the International Press Telecommunications Council (IPTC) Rights Expressions WG for the RightsML Standard, a rights expression language for the media industry[4].

3.2.3. XPref

Agrawal et al. [14] established XPref as an alternative to APPEL, which only allows for the definition of P3P policies that are unacceptable for the user. XPref resorts to XPath (XML Path Language) 1.0 expressions to replace APPEL rules, making the preferences formulation more user-friendly and less error prone. XPath 1.0, by Clark and DeRose [15], is a W3C Recommendation since November 16th, 1999, although no further maintenance will be performed to this version since later versions exist and have achieved the Recommendation statute. XPath's main goal is to provide a way to navigate through the hierarchical elements present in a XML document. To accomplish this task, XPath treats a XML document as a tree of nodes and a XPath expression, when applied to the document, establishes the ordered sequence of the nodes to produce a compact path notation. The path is then comprised of expressions that return nodes, such as root, element, text, attribute, name-space, processing instruction or comment nodes.

XPref was designed so that its rules can not only identify combinations of P3P elements which make a policy unacceptable, according to the user's preferences, but also to verify that the presented elements are specified as acceptable. XPref manages these goals maintaining the APPEL syntax and semantics and its top classes, **ruleset** and **rule**. However, the rule bodies are replaced by XPath expressions since P3P policies are XML documents and thus can be easily matched with the XPath based rules. These expressions are specified by adding a *condition* attribute to the rule, which is responsible for triggering the rule when the XPath expression provides a non-empty outcome. Thus with XPref rules, using the *behavior* attribute, it is possible to establish a preference to block or allow services according to the P3P policy elements, e.g. purposes and recipients, specified on the *condition* attribute.

3.2.4. S4P

S4P (*SecPAL for Privacy*), developed by Becker et al. [16], is a language framework to express user's privacy preferences and web services data handling policies. This language was developed by Microsoft Research[5] and it is an extension of the company previous work, SecPAL, to define the handling of Personally Identifiable Information (PII).

SecPAL [17] is an extensible and decentralized authorization language, developed to express policies and better disclose expressiveness features such as delegation, domain-specific constraints, and negation. An authorization policy is composed of a group of assertions that have an issuer, that vouches for the assertion, the collection of conditional facts and constraints related to times, dates or addresses. Then, when requesting access to the service, this request is transformed into a series of queries, which are checked against the clauses defined to represent the system's policy, so that the decision is made. S4P extends SecPAL to treat granted rights and required obligations as assertions and queries and, based on these, a satisfaction checking algorithm is defined for the disclosure of PII between users and data collecting services. Therefore, services express data-handling policies as SecPAL queries, defining what is going to be their behaviour in relation to the users' PII, and the users express their preferences as SecPAL assertions, precising what the services are permitted to do and what their obligations are towards the users' PII. The satisfaction algorithm then checks if the services data collecting activities match the behaviours permitted by the users and if the obligations defined on the users' preferences are respected by the services' policies. If the outcome of this algorithm is positive, meaning the service's policy satisfies the preferences of the user, the service can proceed with its data collecting activities. S4P also defines a data disclosure protocol to ensure that the users' preferences are regarded when their data is provided to third parties.

[3] https://www.omaspecworks.org/

[4] https://www.iptc.org/std/RightsML/2.0/RightsML_2.0-specification.html

[5] https://www.microsoft.com/en-us/research/

In addition to having an XML schema for implementations, S4P has a human-readable and unambiguous syntax that allows it to be used in other applications.

3.2.5. Accountability in RDF (AIR)

Khandelwal et al. [18] implemented AIR, a declarative language to make assertions of facts and addition of rules, based on N3Logic [19], that supports rule nesting, rule reuse, and automated explanations of rule-based actions performed by the AIR reasoner. These explanations are customizable and, since they can be a source of sensitive information such as PII, can be used to provide privacy, for instance, to hide actions performed under certain rules.

N3Logic is an extension of the RDF data model that aims at expressing logic rules in the web, so that the same language is used for data and logic.

AIR builds on N3Logic's built-in functions, nested graphs and contextualized reasoning, allowing the AIR rules to adopt the usage of graphs as literal values, universally or existentially quantified variables in graphs and built-in functions or operators expressed as RDF properties.

Each rule has a unique Internationalized Resource Identifier (IRI), an HTTP Uniform Resource Identifier (URI), so that it is part of the linked data cloud and can be reused. These rules are defined using the following structure: **air:if** *condition*; **air:then** *then-actions*; **air:else** *else-actions*. The action instances can be annotated through the **air:description** properties. These annotations are then incorporated by the AIR reasoner in its justifications and can be used to hide PIIs present in the rule set. Also, the rules graph format allows for the nesting of rules within the same rule set, thus providing a way to segment the conditions stated by the rule in order to only expose part of them in the justifications.

3.2.6. Privacy Option Language (POL)

POL was developed by Berthold [20] in order to define privacy contracts between data controllers and data subjects, based on the concepts of financial option contracts and respective data disclosure agreements. Its framework applies the data minimization principle by automatically transforming privacy contracts into a canonical form. This canonical form allows the differences among contract compositions to be normalized and so contracts have a similar semantic structure. Also, POL can be used as a core language to be linked with other sub-languages or otherwise be used as a sub-language within a more broad-domain framework.

In POL, each privacy contract is focused on defining the rights and duties regarding data disclosure. As this language emerged in the financial context, contract formulations are mainly based on duties, unless there is no trivial formulation of them. To implement these formulations, POL resorts to several modules that can also be extended. The main components defined by the language are the **syntax** module, the data-related modules for **personal data**, **purpose**, **observable** values and **time**, and the semantics modules for **management** and **human readability**. The syntax module contains the language primitives to define the POL contracts' canonical form. The data modules can then be hooked to the contracts through data support structures as simple as an attribute-value pair, such as *(eye color, brown)*, or as complex as tree-like data organizations. Specifically, the observable module specifies comparison and Boolean operators, which are available in the contract execution environment, to evaluate data retention periods for instance. The time component is useful to module distinct time models, i.e. event/driven time, discrete time, continuous time. The semantic modules, for management and human readability, are used to manage changes in observables, i.e. when time elapses, and to translate POL contracts into natural language, respectively.

This language was developed on the PETWeb II[6] project, with the main goal of addressing the societal questions in relation to the electronic identifiers and electronic identities progresses. The online documentation provides application scenarios for the usage of POL.

3.2.7. Privacy Preference Ontology (PPO)

As privacy is one of the challenges of the open data era, it is of the utmost importance to define whom has access to what, specially in the context of the web. In this light, the PPO [21] proposes to represent users' privacy preferences for the restriction or permission of access to specific RDF data within a RDF document. This ontology extends the Web Access Control (WAC) vocabulary [22], a taxonomy for detailing access control privileges that uses Access Control Lists (ACL) to determine which data users have access to. Its fundamental concepts are the **Read** and **Write** terms, as well as the **Control** privilege to specify and modify the ACL, although this control can only be exercised to de-

[6] http://petweb2.projects.nislab.no/index.php/Main_Page

fine whom can access the full RDF document and not to specify access restrictions over specific data within the document. Therefore, PPO's main goal is to offer highly granular mechanisms to regulate users access to specific data represented as Linked Data, building on the work previously carried out by the WAC.

PPO's restriction abilities should apply to particular statements, to groups of statements (such as RDF graphs) and to resources, that can be particular subjects or objects within statements. The type of restriction must also be defined, as the user can either have read, write or both privileges to the data. Through the defined *hasCondition* property, certain conditions can be set to define privacy preferences in relation to specific resources, instances of particular classes or properties or even to specific values of properties. The access space should also be defined so that the requirements are met by the users to access certain resources. These requirements can be verified through a SPARQL ASK query that contains all attributes and properties that must be met by the users.

Particularly, the same authors focused in developing a specific tool for the semantic web domain, a privacy preference manager [23] based on PPO with the target of providing users with a way to specify their particular privacy choices and regulate the access to their data depending on profile characteristics such as relationships, interests or other common features. This ontology can be used to cover any social data that is modeled on RDF format or through RDF wrappers that can be applied to any major website through their API.

3.2.8. Purpose-To-Use (P2U)

P2U, by Iyilade and Vassileva [24], has taken inspiration from P3P to build a policy language for the sharing of user information across different services and data consumers, resting on the principle of purpose of use. Its main focus is to provide a language for the secondary sharing and usage of data, making sure that the user's privacy is maintained. It is designed to combine information about the data sharing purpose, its retention time and, in the case the user wants to sell it, the selling price and simultaneously allows the data consumers to negotiate prices and retention periods.

This policy framework involves the interaction of the *users* (the owners of the data), the *data consumers* (services that need the data), *the data providers* (services that collect and share the data) and the *data brokers* (services that monitor the consumers and providers activities and execute the negotiations, among other tasks). The main elements of P2U are the **policies**, the **data provider**, the **user**, the **purposes**, the **data consumers**, the **retention**, the **data groups** and respective **data** elements. Policies are the root element of P2U, and each one needs to have an associated provider, an user and at least one purpose of use. Each policy should have a name, and optionally an attribute with the path to the human-readable policy, and the name and identifier of the data provider and user to which the policy refers to. A P2U policy can specify more than one purpose for the sharing of data, along with information on how long it can be retained, with whom and the relevant data it applies to. The data consumer element has the particularity of containing an attribute, *name*, that can be set to 'public' if the data can be shared with any third party service. Also, the retention period of the purpose should be defined in days and a *negotiable* attribute, set to false by default, can also be detailed. The same attribute is available for the data group element. This component is composed by one or more data elements and each one can have an expiry period, which overrides the retention period, and the possibility of setting an initial price for the data in cases where the user is willing to sell it.

An application scenario where a user allows the data sharing between several mobile applications is further specified in an additional publication by the same authors [25]. However this implementation does not enforce compliance of the data consumers with the policies defined by the users and does not specify any special treatment for cases dealing with sensitive data.

3.2.9. Accountable Policy Language (A-PPL)

The A-PPL language, implemented by Azraoui et al. [26], has its origin on the A4Cloud[7] project, with the objective of applying accountability requirements to the representation of privacy policies. To accomplish this goal, the A-PPL expands the PrimeLife Policy Language (PPL) to take into guidelines on notification, data location and retention, and auditability. PPL by Ardagna et al. [27] is an extensible privacy policy language designed on the context of the PrimeLife[8] project, based on the eXtensible Access Control Markup Language (XACML) [28], an OASIS[9] standard for access control policies. PPL's core classes to express an obligation are **triggers** and **actions**. Triggers are events that can be filtered using certain con-

[7] http://www.a4cloud.eu/

[8] http://primelife.ercim.eu/

[9] Non profit organization focused on open standards for cloud, security and other areas, https://www.oasis-open.org/

ditions and are connected to an obligation. These triggers are responsible to fire the data controller's actions, that are executed according to the data subject's authorizations. However, PPL does not cover requirements such as data location and retention rules or auditability to be in line with data handling regulations such as the GDPR.

A-PPL introduced a role attribute identifier and added the data protection authority role to the ones already modeled by PPL, the data subject, data controller and data processor. Also, two new triggers to allow or prohibit access to personal data were included. Duration and region attributes related with a particular data processing purpose are used to enforce data retention and location rules. A-PPL further extends the PPL notification system to define the recipient and the type of notification to be sent in relation to a particular action. For auditing purposes, A-PPL added a trigger to monitor the data controller and collect evidence of data-related events which are logged with parameters such as the purpose of the action, the time-stamp or the executed action on the data.

3.2.10. SPECIAL vocabularies

The EU H2020 SPECIAL (Scalable Policy-awarE linked data arChitecture For prIvacy, trAnsparency and compLiance) project aimed to develop technology that supports today's on-going struggle between privacy and Big Data innovation, providing tools, for data subjects, controllers and processors, that facilitate the management and transparent usage of such data. Two vocabularies were produced as outcomes of this project: the SPECIAL Usage Policy Language (SPL) and the SPECIAL Policy Log Vocabulary (SPLog) [29].

An usage policy represents a set of lawful activities that can be performed in accordance with the data subject's consent. To specify these in formal terms in compliance with the GDPR, the SPL establishes five core elements: the **data** that is going to be processed, the **purpose** of such processing, a description of the **processing** itself, the **storage** information and the **recipients** of the processing results. The data storage element needs two attributes to be instantiated, as both the location and the duration of the storage need to be defined. So, in mathematical terms, the usage policy is a five-element tuple, composed of instantiations of the five core classes, that specifies an authorized operation. A general usage policy can then be defined with an union of authorized operations. The vocabularies designed to specify each of the elements on the SPL are based on previous privacy-related ontologies, such as **ODRL**, for the processing terms, and the **P3P**, for the data categories, recipients, purposes and storage duration. The vocabularies can be further extended by introducing new sub-classes to its terms [30].

SPLog was designed to provide a record of the processing events related to the consent actions given by the data owners. This vocabulary builds upon **PROV-O** to have information on the provenance of the log and is in line with the terms developed for the SPL vocabulary. The main concepts defined by SPLog are the **log** itself and the actual **log entries**. Each log has meta-data associated to it, such as the software agent it belongs to, and log entries that contain information about each event. The log entries can be from one of two types: policy entries - related to a consent form and related policy terms - or data events such as data processing or sharing. These entries should also contain information about the data subject involved in the event, a description, the event's content itself, timestamps, related data-set and so on. Therefore these logs can be used to track the provenance of an event. SPLog uses the SPL vocabulary to instantiate a log entry content. This vocabulary is easily extendable and allows the grouping of events to promote scalability [31].

SPECIAL ontologies were implemented in various use-cases in distinct sectors: to build personalized touristic recommendations in collaboration with *Proximus*[10]; for traffic alert notifications with *Deutsche Telekom*[11]; with *Thomson Reuters Limited*[12] to support anti-money laundering requirements.

3.2.11. Declarative Policy Framework (DPF)

DPF [32, 33] is being developed by an established team under the Defense Advanced Research Projects Agency (DARPA) Brandeis programme[13] with the main goal of providing a privacy policy framework based on ontology engineering and a formal shareability theory. DPF's policy engine builds on the ontology to define policy objects which are used in the development of User Interfaces (UIs). These UIs allow non-technical users to create, validate and manage privacy policies without the need to burden them with technical formalisms of a policy language. DPF's engine can also be integrated into systems supporting the management of data requests and other Privacy Enhancing Technologies (PETs).

[10] https://www.proximus.be/
[11] https://www.telekom.com/en
[12] https://www.thomsonreuters.com
[13] https://www.darpa.mil/program/brandeis

Therefore, DPF uses a defined ontology as a common data model to specify a particular domain in order to support the definition of permissive and restrictive privacy policies. Each policy rule corresponds to an allow or disallow statement that should have an identifier and description, a Policy Authority (PA), the data requesters to whom the policy applies to, and also the affected data and effectiveness time imposed by the policy. Optionally, in the case of a permissive statement, there is the possibility to define a set of constraints to establish the conditions in which the data can be shared. The PA evaluates whether a certain data request is in accordance with the defined policies. Hence, each data request must include, in addition to the data being requested, the PA that will be consulted to grant or refuse access, and the time of the request. Then the request follows the policy engine pipeline and if there is a matching rule the engine returns the decision, the identifier and description of the analogous rule and, in the case the request is authorized, the valid conditions in which it is allowed. Since a single request can trigger multiple policy rules, the engine must be equipped to deal with conflicting decisions. To achieve this, DPF implements baselines policies and then exceptions are created to define policy rules with higher priority in relation to the data that is being shared. With this mechanism in place, this privacy framework can override decisions based on detailed constraints.

The ontologies and consequently the policy rules, can be defined using OWL. To illustrate this framework, the authors provide a pandemic use-case where nation and community PAs implement data sharing policies about their residents and respective health status to monitor the disease's outbreak.

3.2.12. LegalRuleML

LegalRuleML is a rule interchange language applied to the legal domain, defined by the OASIS LegalRuleML Technical Committee, with the status of Committee Specification since April 2020 [34]. It is a XML-schema specification that reuses and extends RuleML concepts and syntax - RuleML is an XML language for rule representation [35] - with formal features to represent and reason over legal norms, guidelines and policies. LegalRuleML's main features include the use of multiple semantic annotations to represent different legal interpretations, the modeling of deontic operators, the temporal management of rules, the authorial tracking of rules and a mapping to RDF triples.

Thus, the core elements of a LegalRuleML document are the **metadata**, the **context** and the **statements**. The metadata section contains information about the **legal source** of the norms, to ensure that they are connected with the legal text statements that specify them, and also about the **actors** and the **roles** they execute in relation to the established rules, about the **jurisdiction** and the **authorities** that create, endorse and enforce the rules and information about the **temporal parameters** that define the period of validity of the rules. The context element allows to express alternative interpretations of the source of the rule, which can change over time or according to jurisdiction, and also enables the representation of the **association** element, which connects the legal sources with the rules. The statements section encompasses the formalization of the norms, including the expression of constitutive and prescriptive statements, overrides statements or violation-reparation statements. The **constitutive** rules represent the definitions present of the legal documents, while the **prescriptive** rules encode the deontic specifications. **Override** statements can be used to deal with incompatible rules and **violation and reparation** statements formalize the penalties applied to norm' breaches.

3.3. Data protection vocabularies and ontologies

In this subsection, we describe the found data protection vocabularies and ontologies, the core classes they implement and, when available, information about use cases where their resources are applied. Pre-GDPR ontologies are mentioned since they can be useful to identify missing concepts and relations between terms.

3.3.1. NEURONA ontologies

Developed by S21SEC[14] and IDT-UAB[15], the main focus of the NEURONA project [36] is the correctness of files containing personal data information and the measures of protection applied to them. Its legal basis was the Spanish protection of personal data regulation that was in effect prior to the GDPR enforcement in all Europe.

The core classes implemented are the **personal data**, **consent**, **purpose** and the **data security measures**. In relation to the data class, categories such as the data regarding religion or racial origin are well de-

[14] https://www.s21sec.com/
[15] http://idt.uab.cat/

fined and fall under special protection security measures. The consent should be given by the data subject in an unambiguous way and for a specific purpose and how it is given depends on the type of data it is related to. Technical and organizational measures for data security should also be in place to regulate the activity of data controllers and processors. These measures should be intrinsically related to the nature of the data and should also reflect the risk associated with its unfulfillment. For this, the concept of **level of security** is introduced by the NEURONA project, a variable that can have three states: low, medium or high. For instance, a file obtained by the police without the consent of the data subjects or a file with data related to the health status of a patient should have high level measures, such as access control policies and backup procedures, associated with it.

These concepts constitute the core ontology of the project, the Data Protection Knowledge Ontology, from which the Data Protection Reasoning Ontology derives with the goal of classifying files based on its compliance with the legislation. Therefore, the NEURONA ontologies could prove useful in the context of companies that deal with great amounts of data stored in files, however, they are not publicly available for usage.

3.3.2. Data Protection Ontology

Bartolini and Muthuri [37] and Bartolini et al. [38] developed an ontology to deal with the new personal data rights and obligations stated by the GDPR, prior to its implementation in May 2018, using an early version of the regulation. The ontology was built focusing on the obligations of the data controller and corresponding rights of the data subject. Therefore the foundations of the ontology are the data protection principles defined in the GDPR, such as the purpose limitation, data quality or data minimization principles.

The ontology was created following the established METHONTOLOGY guide, by Fernández et al. [39], and it is based on the concepts collected from the GDPR, Data Protection Directive (DPD) and the *Handbook on European data protection law* [40], reusing concepts defined on the Legal Knowledge Interchange Format (LKIF) Core [41] and Simple Knowledge Organization System (SKOS) [42] ontologies. The core classes are the **data protection principles**, the **rules of data processing**, that constitute most of the data controller's duties, and the **data subject's rights**, and the ontology is designed so that each data processing rule and data subject's right is connected to at least one data protection principle. For instance, data subjects have the right to access their own data, so the data controller must provide the means for their access to such data. Furthermore, this data protection ontology defines consent as a legal justification connected with the principle of trust and also specifies the special case where parents give consent in the name of the child, although the concept of consent given by delegation is left out. The several entities involved in the data usage, such as the controller, the supervisory authorities or the processor, are also modeled under the **Person** class.

The ontology has been used to extend the Business Process Model and Notation (BPMN), a language to model business processes [43], with the objective of applying data protection concepts that a data controller must follow so that its activity is GDPR compliant.

3.3.3. GDPR Provenance Ontology (GDPRov)

Based on the **PROV-O** and **P-Plan** ontologies, developed by Lebo et al. [44] and by Garijo and Gil [45], respectively, Pandit and Lewis [46] published an ontology with the objective to conceptualize the provenance of data and how the consent and processing of such data are managed in the domain of the GDPR. PROV-O is a provenance ontology, designed to define entities and the relations and activities between them in a generic and domain independent format. It is a W3C recommendation since 2013 and has already been validated in several domains, as demonstrated in the works of Belhajjame et al. [47, 48]. P-Plan (Ontology for Provenance and Plans) is a necessary extension of the PROV-O ontology as the latter does not expand the concept of plan nor does it give detail on the plan execution. With P-Plan extensions of the activities and corresponding steps to execute them, as well as the entities involved, it is possible to track provenance of the interaction between entities and also to monitor how their activities changed over time, for instance, if there have been changes on the consent or on the data being processed.

For queries to be GDPR compliant, provenance information on consent, third party sharing, data collection, usage and storage, anonymisation of personal data and additional rights must be available. Under the GDPR, consent must be given in an explicit and unambiguous way, so that the user knows the purpose to which its data is being processed for and which entities are involved in the data life cycle work-flow. GDPRov implements this through the *ConsentAgreementTemplate* class, a common template regarding consent per-

missions presented to the users that models how the consent is obtained. Therefore, to ensure compliance, a record must be maintained on how the consent was obtained, which processing activities were approved and in the cases where the state of the consent changes, for instance in the case of consent withdrawal, the previous consents should be recorded. Also, data collected for a specific purpose must not be used in other contexts unless the user explicitly consents to it and should only be stored as long as it is necessary. Furthermore, references to third parties with which the data is shared must be detailed to the users, along with specifications on the nature of the data that is being shared, its purpose and information about the entity and its role in the work-flow. For this, provenance meta-data on the origin, use, storage and sharing of the data must be recorded. In the cases where the data was transformed or archived, a version control system must be in place so that the provenance of the data can be tracked. As GDPR authorizes the processing of personal data without consent in the cases where the data cannot be de-anonymised, GDPRov also provides the degree of anonymisation, based on Schwartz and Solove [49]'s work, a property that can have four states: completely anonymous, pseudo-anonymous that cannot be de-anonymised by the organization with which the data was shared, pseudo-anonymous but can be de-anonymised by the organization, and not anonymous. Provenance data on the exercising of rights and obligations from users and data handlers is also kept, so that the records can be checked as proof of compliance. Therefore, for each right or obligation, a plan is defined to reflect the steps involving data or consent that need to be executed when the user wants to exercise a particular right.

3.3.4. Cloud GDPR ontology

Elluri and Joshi [50] developed a GDPR compliant ontology focused on the cloud services to express the obligations of both the cloud data consumers and the cloud data providers, also taking into account the respective Cloud Security Alliance (CSA) controls defined on the *Code of Conduct for GDPR Compliance* [51].

The **stakeholders**, the **CSA controls** and the **obligations** are the core modules of this ontology. The cloud-related obligations are extracted from the GDPR and are connected to the respective articles and also to the associated CSA requirements using the implemented *hasCSAcontrol* property. These GDPR obligations are further specified taking into account which stakeholders they apply to, so there are specific obligations to be followed by cloud consumers and cloud providers and also a few that must be met by both. For instance, maintaining records of the processing activities and notifying data breaches are common obligations, while providing European Union (EU) representatives for non-EU consumers or providers is a responsibility of the consumer and hiring a Data Protection Officer (DPO) falls on the authority of the provider.

This work was extended by Elluri et al. [52] to automate the implementation of both the GDPR and the Payment Card Industry Data Security Standard (PCI DSS) guidelines [53] to compliance. The PCI DSS legislation deals with financial data, such as the credit card number or card-holder's name. Therefore, building and maintaining a secure network, protecting card-holder's data and implementing access control measures are a few of the main requirements of the PCI DSS. As it covers a narrower scope in comparison with the GDPR, a data breach in PCI DSS automatically results in one in GDPR. Thus, the cloud-related PCI DSS requirements were used to enrich this compliance ontology and its validation was done using privacy policies from five major companies that deal with card-holder's data and PII. The ontology was also further developed to include the rights of consumers, providers and end users.

3.3.5. Privacy Ontology for legal reasoning - PrOnto

Palmirani et al. [54] presented in 2018 the first draft of PrOnto, a privacy ontology with the purpose to model the relationships between agents, processing activities, data categories and deontic specifications present on the GDPR. With the goal to support legal reasoning and compliance with the GDPR and other future regulations, PrOnto takes advantage of various other ontologies previously developed. The **LKIF Core** ontology, developed by Hoekstra et al. [41], was used to model the different classes of agents (controller, processor, ...) described in the GDPR as well as the several roles that can be assigned to them.

The **Functional Requirements for Bibliographic Records (FRBR)** ontology by Byrum et al. [55] is used to model legal documents as sources of information, that regulate the different relationships between the agents documented in the text, and to register changes in their representation over time. The FRBR model together with the **A Light Legal Ontology On Top level classes (ALLOT)** ontology, developed by Barabucci et al. [56], are used to model the relationship between the document and the data within,

according to the Akoma Ntoso[16]guidelines. Other ontologies, such as the **Time-indexed Value in Context (TVC)** [57] and the **Time Interval (TI)** [58], are used to connect time-dependant events with specific roles that emerge in certain contexts.

PrOnto was built upon five core modules: **documents and data**, **agents and roles**, **processing and workflow**, **legal rules and deontic formula**, and **purposes and legal bases**. The GDPR document is used as the source of information, from which the main data categories are defined: judicial and sensitive data (personal data) and anonymous and legal person data (non-personal data). The agent and role classes are clearly distinguished as the agent refers to the entity (person, organization, software, ...) while the role class intends to characterize the activity of the agent (data processor, data controller, supervisory authority, ...). Furthermore, an agent can be involved in different roles depending on the context. The processing activity is modulated through a work-flow of actions that should be well placed in terms of the context and time in which each event occurs. This work-flow has associated several properties that are defined in the text, transparency, fairness, lawfulness, and is prepared to deal with eventual data breaches and consequent counter measures. Each processing activity should be performed with a purpose and be committed to a legal rule, which is composed of deontic specifications (prohibitions, rights, permissions and obligations) to check if the activity being executed is in compliance or violation of the GDPR.

This ontology was tested on several use-cases: eGovernment services in the cloud[17], school services and also in the MIREL project[18] and DAPRECO[19] projects.

3.3.6. GConsent

In the Article 6 of the GDPR, the legal basis for the lawful processing of personal data are settled, consent being one of them that should be freely given in a specific, informed and unambiguous way. Information about the consent must be collected and stored, as well as maintaining a log of any changes that may be requested over time, and should be available for all parties involved - data subject, data controller and processor and the authorities.

In this context, Pandit et al. [59] created the GConsent ontology based on the guidelines defined by Noy and McGuinness [60]. The GDPR was the main source adopted to collect information about consent, though other legal authorities' guidelines and reports were used, such as the guidelines on consent published by the European Data Protection Board [61]. However, this ontology only conceptualizes consent in the domain within Article 4.11 of the GDPR, so special cases where other forms of consent are allowed, such as children's personal data or scientific research, are not covered by this model. As GConsent aims at not only capturing the concept of consent, but also to represent its state, context and provenance, existing vocabularies on this subject, such as PROV-O [44], GDPRov [46] and GDPRtEXT [62], are reused.

The core classes are the **data subject**, **personal data**, **purpose** and **processing**, as well as the **consent** and the **status**. GConsent represents a step further in relation to other ontologies that conceptualized consent since it not only defines the 'given consent' concept, but also classifies other states of consent as valid or invalid for processing. Consent status can be one of the following: expired, invalidated, not given, refused, requested, unknown and withdrawn, and in these cases will be invalid for processing, or explicitly given, given by delegation and implicitly given and will be valid. To represent the context in which the context was obtained, information about the location, the time of creation and the medium is recorded, as well as about the expiry of the consent and the entity that granted it. Also, the authors plan to extend the ontology to deal with the spacial and temporal representation of processing activities, such as data storage or sharing, and continue to provide new use-cases to motivate the community's adoption of this model.

3.3.7. BPR4GDPR - Compliance Ontology

The BPR4GDPR (Business Process Re-engineering and functional toolkit for GDPR compliance) project started at May of 2018 and will be running until April 2021. It is a European Union's H2020 innovation programme with the main goal of providing a framework to reinforce the implementation of GDPR-compliant measures inside organizations at diverse scales and in several domains [63].

The Compliance Ontology, described on BPR4GDPR's deliverable D3.1 by Lioudakis and Cascone [64], is based on the BPR4GDPR's Information model, that

[16]XML vocabulary with the primary objective of providing information about the top level classes (person, event, locations, ...) in legal or legislative documents

[17]https://www.agid.gov.it/it/infrastrutture/cloud-pa/cloud-europe

[18]http://www.mirelproject.eu/

[19]https://www.fnr.lu/projects/data-protection-regulation-compliance/

aims to define the entities and respective roles that are involved in the organization processes' life-cycles. Its core classes are the **data types**, the **roles** assigned to users inside the organizations, the **operations** and **operation containers**, the **machine types** that host the operations, the **organization types**, the **events** and **contexts** in which they happened and the **purposes** for which operations are executed. The roles class is related to the responsibilities that are assigned to the user in the context of the organization and its instances can be implemented hierarchically according to the detail level of the data and connected through the *isA* property. This hierarchical structure is valid for most classes in the ontology. The data processing activities are implemented through the operations class, which have associated the *hasInput* and *hasOutput* properties that allow to connect the operations with the data that is processed and the one that is generated and respective states (i.e. plain or anonymised). These operations can be grouped in an operation container - a class that groups processing activities in contexts where they usually work together, for instance, in the management of a database, in which functions such as create, read, update or delete are commonly used. The roles and operations classes should always be connected with an instance of the purpose class. The events class, that aims at capturing all processing activities, a data breach or the revoke of consent, has associated the context class to instantiate specific cases and provide temporal and spatial details, among others.

Using this ontology, BPR4GDPR defines a policy instantiating its purpose, context, action, pre-action and pos-action. The action reflects the activity permitted, prohibited or obliged by the policy, while the pre-action and pos-action indicate the actions that must take place before and after the main action. In turn, each action is specified by the user's role, data, operation and the organization where it takes place.

BPR4GDPR is implementing services in three use-cases: for governmental services in the social security and healthcare domains with IDIKA S.A.[20]; for automotive management with CAS Software AG[21]; and for cloud-supported real state agencies with Innovazioni Tecnologiche[22].

[20] http://www.idika.gr/

[21] https://www.cas.de/en/homepage.html

[22] https://www.innovazioni-tecnologiche.com/en/index.aspx#about

3.3.8. Data Privacy Vocabularies

The Data Privacy Vocabulary (DPV) was introduced by the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG)[23] in 2018 when the GDPR came into force. This W3C CG was one of the first outputs from a W3C workshop on data privacy controls, that took place in Vienna in April 2018, with the objective of defining priorities for the standardization of this domain [65]. Initially, the group searched for relevant vocabularies that attempted to address data privacy and, in particular, the GDPR. From this state of the art review, a few conclusions emerged: there is a need for vocabularies to describe personal data and the purposes for the processing of said data, as well as vocabularies to coordinate privacy legislations. The methodology used to develop the vocabulary was based on the **NeOn** methodology by Suárez-Figueroa et al. [66] and the **SPECIAL Usage Policy Language** [67] was the core ontology used to module the processing, purpose, recipient and personal data category classes. New concepts were added to the vocabulary after being discussed and agreed upon by the CG. As a result of this process, a first version of the base vocabulary was published with the following main classes: **personal data categories**, **processing**, **purposes**, **legal basis**, **technical and organizational measures** and **legal entities**, including **data subject** and **child**, **recipients**, **data controller**, **data processor** and **third party** [68]. A second version of the base vocabulary was released in January 2021; the **risk**, **right** and **data subject right** classes were added to the base vocabulary and the previously existing classes were extended with new terms. Moreover, new legal entities, including **authority** and **data protection authority**, **vulnerable data subject**, **data sub-processor**, **data protection officer** and **representative**, were added to the vocabulary. DPV's classes are further developed as sub-vocabularies, making it possible for them to be used independently[24].

The personal data categories are split into top level classes such as financial or social data, which are further specified, and classes for sensitive and derived data are also present as required by the GDPR. The top level categories are adapted from the **EnterPrivacy** taxonomy by Cronk [69]. The purpose vocabulary is composed of 42 suggested purpose sub-classes, which are topped by classes such as R&D or Com-

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

[24] https://github.com/dpvcg/dpv/tree/master/rdf

mercial Interest, that can be extended to specify other GDPR purposes not yet conceptualized. The purpose category can be further constrained to specific contexts or business sectors. In relation to the processing categories, DPV covers the terms defined in the Article 4-2 of the GDPR, providing 40 processing categories. Properties related to the origin of the data being processed or the logic used in automated decision making algorithms are available to check compliance with the GDPR. Technical and organizational measures, such as the pseudo-anonymisation and encryption of the data, must be in place so that the processing of personal data is in line with the GDPR. These categories of measures are usually accompanied by a comment to describe the measure or the standardized practices to follow. The consent legal basis is further specified in the DPV with the withdrawal, provision and expiry concepts, based on **GConsent** [59] and **Consent Receipt** [70].

The CG also developed a GDPR extension for DPV, the DVP-GDPR vocabulary[25]. DVP-GDPR covers all the legal bases specified on the GDPR Articles 6 and 9 for the processing of personal data and also the legal bases for the transfer of personal data to third countries defined on Articles 45, 46 and 49. This vocabulary also models 12 GDPR rights of the data subjects.

The work to improve and extend the DPV vocabularies, as well as to provide more examples of application scenarios, is ongoing to date.

3.4. GDPR as a linked open data resource

Pandit et al. [62] developed **GDPRtEXT**, a linked open data resource that provides a way to connect GDPR concepts with the specific sections, chapters, articles or points of the GDPR text. GDPRtEXT is an extension of the **European Legislation Identifier (ELI)**, an ontology developed for the identification of European, national and regional legislation through URI templates [71]. Extending the properties defined by ELI, GDPRtEXT provides a way to link the correlated chapters, sections, articles or points. The ontology was developed using the *"Ontology Development 101"* guide by Noy and McGuinness [60] and the SKOS vocabulary was used to describe the GDPR terms.

The main terms represented in this ontology are the specific **entities** mentioned in the regulation's text, the **rights** and **obligations** of the entities, the **principles** and the **activities** which specify processes and actions defined in the GDPR, such as reporting a data breach, exercising rights or demonstrating consent. These terms are connected to the relevant points in the GDPR text using the *rdfs:isDefinedBy* property.

GDPRtEXT's documentation also contains two example use-cases where it was used for GDPR compliance reports and also to link obligation concepts with the previous data protection regulation, the DPD.

4. Discussion

4.1. Analysis of existing resources

Using Table 4 as a reference, it is possible to compare the policy languages described in the previous section in relation to their capacity of assisting with the representation of the GDPR data subject's rights.

Most of the analysed languages can be used to partially fulfill the representation needs identified in Section 2, related to the *'right to be informed'* (RI1 and RI2), as well as the *'right of access'* (RA) and *'right of rectification'* (RR), apart from PPO. However, only three languages, ODRL, AIR and LegalRuleML, have the resources to partially support the representation of most of the data subject's rights, excluding the *'right to data portability'* (RDP) and the *'right to not be subjected to automated decision-making'* (RDM).

Using Table 5, it is possible to conclude that most of the languages can partially cover the representation needs of the obligation to maintain *records of processing activities* (ROPA-DC and ROPA-DP), excluding PPO. P3P, APPEL, ODRL, XPref, POL, P2U, DPF, SPL and LegalRuleML can also be used to partially model the *contract with processors* (CPr) and the *notification of a data breach* (NDB) duty. In particular, ODRL stands out from other languages by having the resources to represent, at least partially, six of the rights and obligations described in Section 2.3.

Although these languages do not specifically mention the rights and obligations discussed in Section 2, they can be used to represent a few of the items of information mentioned by them, which is why they are classified as capable of partially representing each right or obligation. It should also be noted that no language seems to have the necessary resources to represent the *'right to data portability'* (RDP), the *'right to not be subjected to automated decision-making'* (RDM), the right to *prior consultation* (PC) with the

[25] https://github.com/dpvcg/dpv-gdpr/

supervisory authority and the *data protection impact assessment* (DPIA) duty.

From the described languages, solely ODRL, LegalRuleML and DPF continue to be actively maintained and developed, and only P3P, APPEL, ODRL, AIR, LegalRuleML and SPL have the resources available for reuse on the Web. Since the majority of the policy languages were developed before the GDPR came into full effect, they do not model concepts such as the legal basis for processing or the rights of the data subject.

In this context, the ontologies and vocabularies in the domain of privacy and data protection as well as the GDPRtEXT ontology, described in the previous sections and compared in Tables 6 and 7, are of particular interest to cover these gaps on the representation of informational items. When available, the name of the class that can be used to model the respective informational item is detailed, as well as the number of sub-classes which can be used to more specifically define the term. The cases in which there is still no specific concept to represent the informational item, yet there are terms that can be extended to accomplish it, are marked with an asterisk. Informational items I15, I19, I24, I26, I28 to I30, I33, I34, I46 to I48, I51, I52 and I54 to I56 are not represented in either Table 6 or 7 since they are not modeled by any of the analyzed ontologies.

DPO, GDPRov, PrOnto, DPV and GDPRtEXT can be used to partially populate a great deal of the informational items required by the *'right to be informed'* (RI1 and RI2) and the other GDPR rights and obligations. However, we must highlight DPV and GDPRtEXT since they represent, at least partially, 31 and 25 informational items, respectively, out of the 57 described in Table 1. Furthermore, these vocabularies are the ones that have the largest number of sub-classes to specifically define the respective informational items.

Most of the ontologies and vocabularies presented are obsolete or without new developments in recent years, with BPR4GDPR's IMO, GDPRov, GConsent, DPV and GDPRtEXT being the only ones that continue to be improved. Moreover, only DPKO, IMO and PrOnto do not have open and accessible resources.

Taking into account the performed analysis, it can be concluded that ODRL, DPV and GDPRtEXT are resources that can be easily extended to support the discussed representation needs of GDPR rights and obligations. As an example, Listing 1 combines these vocabularies with a few new terms to describe a *communication of a data breach* (CDB) obligation. This example describes the need of a certain controller to inform a specific data subject in the case of a personal data breach event. A data controller keeping these obligations in this structured form can more easily fulfill them if the event actually happens.

Listing 1: Communication of a personal data breach to a data subject.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix dpv: <http://www.w3.org/ns/dpv#> .
@prefix gdprtext: <https://w3id.org/GDPRtEXT#> .
@prefix gdprif: <https://protect.oeg.fi.upm.es/def/gdprif#> .

gdprif:communicateDataBreach
    rdfs:comment "Data controller A informs Beatriz Esteves about the
                  existence of a personal data breach";
    rdfs:seeAlso gdprtext:NotifyDataSubjectOfBreach ;
    odrl:obligation [
        odrl:informedParty [
            a dpv:DataSubject, odrl:Party ;
            dpv:hasName "Beatriz Esteves" .
        ] ;
        odrl:informingParty _:ControllerA ;
        odrl:assignee _:ControllerA ;
        odrl:action odrl:inform ;
        odrl:target gdprif:I5, gdprif:I40, gdprif:I46, gdprif:I49, gdprif:I50 ;
        odrl:constraint [
            a odrl:Constraint ;
            odrl:leftOperand odrl:event ;
            odrl:operator odrl:eq ;
            odrl:rightOperand gdprif:PersonalDataBreach ;
        ] .
    ] .

_:ControllerA
    a dvp:DataController, odrl:Party ;
    dpv:hasName "Controller A" .
```

4.2. Supplementary material

In order to complement the description of privacy languages, ontologies and vocabularies presented on Section 3 of this paper, an online portal[26] has been published with additional resources. For each solution, there is a brief description of the language or ontology and also links to additional documentation and available RDF serializations. There is more information about the authors of the solutions, when it was first created and last updated, about the projects or the research groups where it was developed and, when available, examples of implementations that are using it.

This webpage also includes a REST API service to find references to specific concepts in the collection of ontologies and languages that have been identified in the context of this paper. The main objective of this service is to give users a platform where they can search for ontologies that model processing activities such as 'derive' or 'disclose' or a language that can be used to represent the *'right to erasure'*.

Furthermore, we specify a lightweight ontology, the GDPR Information Flows (GDPRIF)[27], in order to model the relationships between GDPR stakeholders, informational items, GDPR rights and obligations and

[26] https://protect.oeg.fi.upm.es/sota/

[27] https://protect.oeg.fi.upm.es/def/gdprif

Table 4

Representation of GDPR's data subject rights (R*) in the identified privacy policy language solutions. The languages that can be used to partially assist with a particular right are marked with an asterisk.

	RI1	RI2	RA	RR	RE	RRP	RN	RDP	RO	RDM
P3P	*	*	*	*						
APPEL	*	*	*	*						
ODRL	*	*	*	*	*	*	*		*	
XPref	*	*	*	*						
S4P	*	*	*	*						
AIR	*	*	*	*	*	*	*		*	
POL	*	*	*	*						
PPO			*							
P2U	*	*	*	*						
A-PPL	*	*	*	*						
DPF	*	*	*	*						
SPL	*	*	*	*						
LegalRuleML	*	*	*	*	*	*	*		*	

Table 5

Representation of the rights and obligations of data controllers and processors in the identified privacy policy language solutions. The languages that can be used to partially assist with the needs of a particular right or obligation are marked with an asterisk.

	JC	CPr	ROPA-DC	ROPA-DP	NDB	CDB	DPIA	PC
P3P		*	*	*	*			
APPEL		*	*	*	*			
ODRL	*	*	*	*	*	*		
XPref		*	*	*	*			
S4P			*	*				
AIR			*	*		*		
POL		*	*	*	*			
PPO								
P2U		*	*	*	*			
A-PPL			*	*	*			
DPF		*	*	*	*			
SPL	*	*	*	*	*			
LegalRuleML		*	*	*	*	*		

also to specify information about the flows of information and about the events that trigger the rights and obligations.

5. Conclusions

There is a strong need to develop technologies to support individuals to manage their personal information and at the same time there is a need to support companies to better manage compliance. Having common vocabulary elements and common data models to refer to these rights and to denote specific GDPR concepts would favor data subjects and data controllers to speak in the same terms, and would ease the interoperability between different types of tools. Not only companies may have information systems to manage the individuals' consent and abide the law: other software systems can also help individuals to manage the consent they are constantly giving. In particular: data subjects can control the access to their personal data in distributed stores; as recommended by the Opinion 9/2016 of the European Data Protection Supervisor on Personal Information Management Systems (PIMS). Conversely, data controllers can make sure they have complied with their obligations about (i) informing the data subjects and (ii) responding to the data subjects' requests. For example, having a categorization of the

Table 6

Representation of the informational items I1 to I57 in the DPKO, DPO, GDPRov, Cloud and PrOnto ontologies. The names of the classes which can be used to specify a particular item are depicted in the table, as well as their respective number of sub-classes. The informational items which can not be fully represented by the current ontology terms are illustrated with an asterisk.

	DPKO	DPO	GDPRov	Cloud	PrOnto
I1		Controller	Controller		*
I3			ControllerRepresentative		
I6	*	Purpose			Purpose (10)
I7	*	LegalJustification (6)			
I8		LegitimateInterest			
I9		Recipient (2)			*
I10			*		
I11					*
I12		DataSubjectRight (7)	Process (10)		Right (8)
I16		AutomatedProcessing	*		
I17	*	PersonalData	PersonalData (3)		PersonalData (7)
I20		*			
I21			ProvideCopyOfPersonalData		
I22			RectifyData		
I31			JointController		
I36					Action (13)
I37		DataSubject (1)			
I38		*	*	*	*
I39		*		*	*
I40		DataProtectionOfficer	DPO		*
I41	*	Measures (2)			
I42		Processor	Processor		*
I44			ProcessorRepresentative		
I57		*	*		

types of information that an individual should receive would enable automatic labeling tools analyzing existing text communications. Aligning ontologies and vocabularies with the GDPR (or other equivalent norms in other territories) would greatly favor interoperability of the privacy-related tools both in the side of the individuals and in the side of the companies.

This paper has analyzed the value of existing policy languages, vocabularies and ontologies to support these interoperability needs, and has concluded that ODRL, DPV and GDPRtEXT are mature resources, ready to be used for representing privacy-related rights and obligations, with an explicit link to the current version of the GDPR text. Points in favor of these solutions are the fact that they are open access, have good documentation and, in the case of ODRL, it is already a W3C recommendation for digital rights management. An example of using these resources to specify the obligation to report a data breach is given to support this conclusion. In terms of future work, we intend to create ODRL-DPV-GDPRtEXT rules for each of the rights and obligations found in GDPR, as this exceeds the ambitions of this paper, but would favor its quick adoption.

Acknowledgements

This research has been supported by European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497 (PROTECT).

References

[1] A.F. Westin, Special Report: Legal safeguards to insure privacy in a computer society, *Communications of the ACM* **10**(9) (1967), 533–537.

[2] P. Kumaraguru, J. Lobo, L. Cranor and S.B. Calo, A Survey of Privacy Policy Languages, *World Wide Web Internet And Web Information Systems* (2007).

Table 7

Representation of the informational items I1 to I57 in the GConsent, IMO, DPV and GDPRtEXT ontologies. The names of the classes which can be used to specify a particular item are depicted in the table, as well as their respective number of sub-classes. The informational items which can not be fully represented by the current ontology terms are illustrated with an asterisk.

	GConsent	IMO	DPV	GDPRtEXT
I1	DataController	DataController	DataController	Controller
I2			hasContact	
I3			Representative	ControllerRepresentative
I4			hasContact	
I5			hasContact	
I6	Purpose	Purposes	Purpose (42)	
I7	*		LegalBasis (34)	LawfulBasisForProcessing (14)
I8			A6-1-f	LegitimateInterest
I9	*		Recipient (3)	*
I10			*	CrossBorderTransfer
I11	*	*	*	RecordDataRetentionPeriod
I12			DataSubjectRight (12)	Rights (10)
I13	*		A7-3	
I14			A77	
I16			AutomatedDecisionMaking	AutomatedProcessing
I17		DataTypes (52)	PersonalDataCategory (170)	PersonalData (5)
I18			DataSource	InfoAboutSourceOfData
I20			*	
I21				ProvideCopyOfPersonalData
I23				RightOfErasure (2)
I25			hasContact	
I27				RightToRestrictProcessing (3)
I31				JointController
I32			hasContact	
I35			*	
I36	Processing (18)	Operations (40)	Processing (40)	DataActivity (9)
I37	DataSubject (1)	DataSubject	DataSubject (2)	DataSubject
I38				ControllerObligation (11)
I39				ProcessorObligation (14)
I40		DataProtectionOfficer	DataProtectionOfficer	DPO
I41			TechnicalOrganisationalMeasure (37)	
I42		DataProcessor	DataProcessor	Processor
I43			hasContact	
I44			Representative	ProcessorRepresentative
I45			hasContact	
I49				*
I50				*
I53			Risk	
I57			DPIA	*

[3] S. Kasem-Madani and M. Meier, Security and Privacy Policy Languages: A Survey, Categorization and Gap Identification (2015). http://arxiv.org/abs/1512.00201.

[4] J. Leicht and M. Heisel, A Survey on Privacy Policy Languages: Expressiveness Concerning Data Protection Regulations, in: *2019 12th CMI Conference on Cybersecurity and Privacy (CMI)*, IEEE, 2019, pp. 1–6. ISBN 978-1-72812-856-6. doi:10.1109/CMI48017.2019.8962144. https://ieeexplore.ieee.org/document/8962144/.

[5] S. Kirrane, A. Mileo and S. Decker, Access control and the Resource Description Framework: A Survey, *Semantic Web* **8**(2) (2016), 311–352. doi:10.3233/SW-160236.

[6] T. Pellegrini, A. Schönhofer, S. Kirrane, A. Fensel, O. Panasiuk, V. Mireles-Chavez, T. Thurner, M. Dörfler and A. Polleres, A Genealogy and Classification of Rights Expression Languages - Preliminary Results, in: *Proceedings of the 21st International Legal Informatics Symposium*, 2018, pp. 243–250.

[7] H.J. Pandit, Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance, 2020.

[8] P. Ashley, S. Hada, G. Karjoth, C. Powers and M. Schunter, Enterprise Privacy Authorization Language (EPAL 1.2), 2003. https://www.w3.org/Submission/2003/SUBM-EPAL-20031110/.

[9] P. Ashley, S. Hada, G. Karjoth and M. Schunter, E-P3P privacy policies and privacy authorization, in: *Proceeding of the ACM workshop on Privacy in the Electronic Society - WPES '02*, ACM Press, 2002, pp. 103–109. ISBN 978-1-58113-633-3. doi:10.1145/644527.644538. http://portal.acm.org/citation.cfm?doid=644527.644538.

[10] K. Bohrer and B. Holland, Customer Profile Exchange (CPExchange) Specification, Technical Specification, 2000.

[11] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall and J. Reagle, The Platform for Privacy Preferences 1.0 (P3P1.0) Specification, 2002, Publication Title: W3C Recommendation 16 April 2002 obsoleted 30 August 2018. https://www.w3.org/TR/P3P/.

[12] L. Cranor, M. Langheinrich and M. Marchiori, A P3P Preference Exchange Language 1.0 (APPEL1.0), 2002. https://www.w3.org/TR/2002/WD-P3P-preferences-20020415/.

[13] R. Iannella, M. Steidl, S. Myles and V. Rodríguez-Doncel, ODRL Vocabulary & Expression 2.2, 2018, Publication Title: W3C Rec. https://www.w3.org/TR/odrl-vocab/.

[14] R. Agrawal, J. Kiernan, R. Srikant and Y. Xu, XPref: a preference language for P3P, *Computer Networks* **48**(5) (2005-08), 809–827. doi:10.1016/j.comnet.2005.01.004. https://linkinghub.elsevier.com/retrieve/pii/S1389128605000095.

[15] J. Clark and S. DeRose, XML Path Language (XPath) Version 1.0, 1999. https://www.w3.org/TR/1999/REC-xpath-19991116/.

[16] M.Y. Becker, A. Malkis and L. Bussard, S4P: A Generic Language for Specifying Privacy Preferences and Policies, Technical Report, Microsoft Research, 2010. https://www.microsoft.com/en-us/research/wp-content/uploads/2010/04/main-1.pdf.

[17] M. Becker, C. Fournet and A. Gordon, Design and Semantics of a Decentralized Authorization Language, in: *20th IEEE Computer Security Foundations Symposium (CSF'07)*, IEEE, 2007-07, pp. 3–15, ISSN: 1063-6900. ISBN 978-0-7695-2819-9. doi:10.1109/CSF.2007.18. http://ieeexplore.ieee.org/document/4271637/.

[18] A. Khandelwal, J. Bao, L. Kagal, I. Jacobi, L. Ding and J. Hendler, Analyzing the AIR Language: A Semantic Web (Production) Rule Language, in: *Web Reasoning and Rule Systems*, P. Hitzler and T. Lukasiewicz, eds, Lecture Notes in Computer Science, Vol. 6333, Springer Berlin Heidelberg, 2010, pp. 58–72, Series Title: Lecture Notes in Computer Science. ISBN 978-3-642-15917-6 978-3-642-15918-3. doi:10.1007/978-3-642-15918-3_6. http://link.springer.com/10.1007/978-3-642-15918-3{_}6.

[19] T. Berners-Lee, D. Connolly, L. Kagal, Y. Scharf and J. Hendler, N3Logic: A logical framework for the World Wide Web, in: *Theory and Practice of Logic Programming*, Vol. 8, 2008-05, pp. 249–269. doi:10.1017/S1471068407003213. https://www.cambridge.org/core/product/identifier/S1471068407003213/type/journal{_}article.

[20] S. Berthold, The Privacy Option Language - Specification & Implementation, Research Report, Faculty of Health, Science and Technology, Karlstad University, 2013. http://kau.diva-portal.org/smash/get/diva2:623452/FULLTEXT01.pdf.

[21] O. Sacco and A. Passant, A Privacy Preference Ontology (PPO) for Linked Data, 2011. http://ceur-ws.org/Vol-813/ldow2011-paper01.pdf.

[22] R.W.W.C. Group, WebAccessControl, 2019, Publication Title: W3C Wiki. https://www.w3.org/wiki/WebAccessControl.

[23] O. Sacco and A. Passant, A Privacy Preference Manager for the Social Semantic Web, in: *Proceedings of the 2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation, SPIM2011*, 2011, pp. 42–53. ISBN 16130073.

[24] J. Iyilade and J. Vassileva, P2U: A Privacy Policy Specification Language for Secondary Data Sharing and Usage, in: *2014 IEEE Security and Privacy Workshops*, IEEE, 2014-05, pp. 18–22. ISBN 978-1-4799-5103-1. doi:10.1109/SPW.2014.12. http://ieeexplore.ieee.org/document/6957279/.

[25] J. Iyilade and J. Vassileva, A Framework for Privacy-Aware User Data Trading, in: *User Modeling, Adaptation, and Personalization*, Vol. 7899, S. Carberry, S. Weibelzahl, A. Micarelli and G. Semeraro, eds, Springer Berlin Heidelberg, 2013, pp. 310–317, Series Title: Lecture Notes in Computer Science. ISBN 978-3-642-38843-9 978-3-642-38844-6. doi:10.1007/978-3-642-38844-6_28. http://link.springer.com/10.1007/978-3-642-38844-6{_}28.

[26] M. Azraoui, K. Elkhiyaoui, M. Önen, K. Bernsmed, A.S. De Oliveira and J. Sendor, A-PPL: An Accountability Policy Language, Research Report, 2014. http://www.eurecom.fr/en/publication/4372/download/rs-publi-4372.pdf.

[27] C.A. Ardagna, L. Bussard, S.D.C.d. Vimercati, G. Neven, S. Paraboschi, E. Pedrini, F.-S. Preiss, D. Raggett, P. Samarati, S. Trabelsi and M. Verdicchio, PrimeLife Policy Language, Technical Report, 2009.

[28] B. Parducci, H. Lockhart and E. Rissanen, eXtensible Access Control Markup Language (XACML) Version 3.0, 2013. http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html.

[29] S. Kirrane, J.D. Fernández, W. Dullaert, U. Milosevic, A. Polleres, P.A. Bonatti, R. Wenning, O. Drozd and P. Raschke, A Scalable Consent, Transparency and Compliance Architecture, in: *The Semantic Web: ESWC 2018 Satellite Events*, A. Gangemi, A.L. Gentile,

A.G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J.Z. Pan and M. Alam, eds, Lecture Notes in Computer Science, Vol. 11155, Springer International Publishing, 2018, pp. 131–136, Series Title: Lecture Notes in Computer Science. ISBN 978-3-319-98191-8 978-3-319-98192-5. doi:10.1007/978-3-319-98192-5_25. http://link.springer.com/10.1007/978-3-319-98192-5{_}25.

[30] P.A. Bonatti, S. Kirrane, I. Petrova, L. Sauro and E. Schlehahn, Policy Language V2 - Deliverable D2.5, Project deliverable, 2018. https://www.specialprivacy.eu/images/documents/SPECIAL{_}D25{_}M21{_}V10.pdf.

[31] S. Kirrane, U. Milosevic, J.D. Fernández, A. Polleres and J. Langens, Transparency Framework V2 - Deliverable D2.7, Project deliverable, 2018. https://www.specialprivacy.eu/images/documents/SPECIAL{_}D27{_}M23{_}V10.pdf.

[32] K. Martiny, D. Elenius and G. Denker, Protecting Privacy with a Declarative Policy Framework, in: *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, IEEE, 2018-01, pp. 227–234. ISBN 978-1-5386-4408-9. doi:10.1109/ICSC.2018.00039. http://ieeexplore.ieee.org/document/8334462/.

[33] K. Martiny and G. Denker, Partial Decision Overrides in a Declarative Policy Framework, in: *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, IEEE, 2020-02, pp. 271–278. ISBN 978-1-72816-332-1. doi:10.1109/ICSC.2020.00056. https://ieeexplore.ieee.org/document/9031488/.

[34] M. Palmirani, G. Governatori, T. Athan, H. Boley, A. Paschke and A. Wyner, LegalRuleML Core Specification Version 1.0, 2020. https://docs.oasis-open.org/legalruleml/legalruleml-core-spec/v1.0/legalruleml-core-spec-v1.0.html.

[35] H. Boley, A. Paschke, T. Athan, A. Giurca, N. Bassiliades, G. Governatori, M. Palmirani, A. Wyner, A. Kozlenkov and G. Zou, Specification of RuleML 1.02, 2017. http://wiki.ruleml.org/index.php/Specification_of_RuleML_1.02.

[36] N. Casellas, J.-E. Nieto, A. Meroño, A. Roig, S. Torralba, M. Reyes and P. Casanovas, Ontological Semantics for Data Privacy Compliance: The NEURONA Project, in: *2010 AAAI Spring Symposium*, Intelligent Information Privacy Management, AAAI, 2010, pp. 34–38. https://ddd.uab.cat/pub/artpub/2010/137891/aaaisprsymser{_}a2010n1iENG.pdf.

[37] C. Bartolini and R. Muthuri, Reconciling Data Protection Rights and Obligations: An Ontology of the Forthcoming EU Regulation, in: *Workshop on Language and Semantic Technology for Legal Domain*, 2015.

[38] C. Bartolini, R. Muthuri and C. Santos, Using Ontologies to Model Data Protection Requirements in Workflows, in: *New Frontiers in Artificial Intelligence*, M. Otake, S. Kurahashi, Y. Ota, K. Satoh and D. Bekki, eds, Lecture Notes in Computer Science, Vol. 10091, Springer International Publishing, 2017, pp. 233–248, Series Title: Lecture Notes in Computer Science. ISBN 978-3-319-50952-5 978-3-319-50953-2. doi:10.1007/978-3-319-50953-2_17. http://link.springer.com/10.1007/978-3-319-50953-2{_}17.

[39] M. Fernández, A. Gómez-Pérez and N. Juristo, Methontology: From Ontological Art Towards Ontological Engineering, *Proceedings of the Ontological Engineering AAAI-1997 Spring Symposium Series* (1997), 33–40.

[40] E.U.A. for Fundamental Rights, *Handbook on European data protection law*, Re-ed. edn, Handbook / FRA, European Union Agency for Fundamental Rights, Publ. Office of the Europ. Union [u.a.], 2014, OCLC: 931804500. ISBN 978-92-871-9934-8 978-92-9239-461-5.

[41] R. Hoekstra, J. Breuker, M. Di Bello and A. Boer, The LKIF Core Ontology of Basic Legal Concepts, *Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007)* (2007), 43–63.

[42] A. Miles and S. Bechhofer, SKOS Simple Knowledge Organization System Reference, 2009. https://www.w3.org/TR/skos-reference/.

[43] O.M.G. (OMG), Business Process Model and Notation (BPMN) Version 2.0, Specification, 2011. http://www.omg.org/spec/BPMN/2.0.

[44] T. Lebo, S. Sahoo and D. McGuinness, PROV-O: The PROV Ontology, 2013. https://www.w3.org/TR/prov-o/.

[45] D. Garijo and Y. Gil, Augmenting PROV with Plans in P-PLAN: Scientific Processes as Linked Data, in: *CEUR Workshop Proceedings*, 2012.

[46] H.J. Pandit and D. Lewis, Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies, in: *Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2017), co-located with ISWC 2017*, Vol. 1951, 2017. http://ceur-ws.org/Vol-1951/PrivOn2017{_}paper{_}6.pdf.

[47] K. Belhajjame, J. Zhao, D. Garijo, A. Garrido, S. Soiland-Reyes, P. Alper and O. Corcho, A workflow PROV-corpus based on taverna and wings, in: *Proceedings of the Joint EDBT/ICDT 2013 Workshops on - EDBT '13*, ACM Press, 2013, p. 331. ISBN 978-1-4503-1599-9. doi:10.1145/2457317.2457376. http://dl.acm.org/citation.cfm?doid=2457317.2457376.

[48] K. Belhajjame, J. Zhao, D. Garijo, M. Gamble, K. Hettne, R. Palma, E. Mina, O. Corcho, J.M. Gómez-Pérez, S. Bechhofer, G. Klyne and C. Goble, Using a suite of ontologies for preserving workflow-centric research objects, *Journal of Web Semantics* **32** (2015-05), 16–42. doi:10.1016/j.websem.2015.01.003. https://linkinghub.elsevier.com/retrieve/pii/S1570826815000049.

[49] P.M. Schwartz and D.J. Solove, PII 2.0: Privacy and a New Approach to Personal Information, Technical Report, 2012.

[50] L. Elluri and K.P. Joshi, A Knowledge Representation of Cloud Data Controls for EU GDPR Compliance, in: *2018 IEEE World Congress on Services (SERVICES)*, IEEE, 2018, pp. 45–46. ISBN 978-1-5386-7374-4. doi:10.1109/SERVICES.2018.00036. https://ieeexplore.ieee.org/document/8495788/.

[51] C.S.A.-P.L.A.W. Group, Code of Conduct for GDPR Compliance, 2017. https://downloads.cloudsecurityalliance.org/assets/research/gdpr/CSA{_}Code{_}of{_}Conduct{_}for{_}GDPR{_}Compliance.pdf.

[52] L. Elluri, A. Nagar and K.P. Joshi, An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance, in: *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, 2018-12, pp. 1266–1271. ISBN 978-1-5386-5035-6. doi:10.1109/BigData.2018.8622236. https://ieeexplore.ieee.org/document/8622236/.

[53] P.S.S. Council, Payment Card Industry (PCI) Data Security Standard - Version 3.2.1, 2018. https://www.pcisecuritystandards.org/document{_}library.

[54] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini and L. Robaldo, PrOnto: Privacy Ontology for Legal Reasoning, in: *Electronic Government and the Information Systems Perspective (EGOVIS 2018)*, A. Kő and E. Francesconi, eds, Lec-

ture Notes in Computer Science, Vol. 11032, Springer International Publishing, 2018, pp. 139–152, Series Title: Lecture Notes in Computer Science. ISBN 978-3-319-98348-6 978-3-319-98349-3. doi:10.1007/978-3-319-98349-3_11. http://link.springer.com/10.1007/978-3-319-98349-3{_}11.

[55] J. Byrum, S. Jouguelet, D. McGarry, N. Williamson, M. Witt, T. Delsey, E. Dulabahn, E. Svenonius and B. Tillett, Functional Requirements for Bibliographic Records, Technical Report, 2009. https://www.ifla.org/publications/functional-requirements-for-bibliographic-records.

[56] G. Barabucci, L. Cervone, A. Di Iorio, M. Palmirani, S. Peroni and F. Vitali, Managing semantics in XML vocabularies: an experience in the legal and legislative domain, 2010. ISBN 978-1-935958-01-7. doi:10.4242/BalisageVol5.Barabucci01. http://www.balisage.net/Proceedings/vol5/html/Barabucci01/BalisageVol5-Barabucci01.html.

[57] S. Peroni, The Semantic Publishing and Referencing Ontologies, in: *Semantic Web Technologies and Legal Scholarly Publishing*, Law, Governance and Technology Series, Vol. 15, Springer, Cham, 2014, pp. 121–193. ISBN 978-3-319-04776-8.

[58] A. for Ontology Design \& Patterns (ODPA), Ontology Design Patterns.org (ODP) - Time interval ontology. http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl.

[59] H.J. Pandit, C. Debruyne, D. O'Sullivan and D. Lewis, GConsent - A Consent Ontology Based on the GDPR, in: *The Semantic Web*, P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A.J.G. Gray, V. Lopez, A. Haller and K. Hammar, eds, Lecture Notes in Computer Science, Vol. 11503, Springer International Publishing, 2019, pp. 270–282, Series Title: Lecture Notes in Computer Science. ISBN 978-3-030-21347-3 978-3-030-21348-0. doi:10.1007/978-3-030-21348-0_18. http://link.springer.com/10.1007/978-3-030-21348-0{_}18.

[60] N.F. Noy and D.L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology (2001).

[61] E.D.P. Board, Guidelines 05/2020 on consent under Regulation 2016/679 Version 1.1, 2020. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb{_}guidelines{_}202005{_}consent{_}en.pdf.

[62] H.J. Pandit, K. Fatema, D. O'Sullivan and D. Lewis, GDPRtEXT - GDPR as a Linked Data Resource, in: *The Semantic Web*, A. Gangemi, R. Navigli, M.-E. Vidal, P. Hitzler, R. Troncy, L. Hollink, A. Tordai and M. Alam, eds, Lecture Notes in Computer Science, Vol. 10843, Springer International Publishing, 2018, pp. 481–495, Series Title: Lecture Notes in Computer Science. ISBN 978-3-319-93416-7 978-3-319-93417-4. doi:10.1007/978-3-319-93417-4_31. http://link.springer.com/10.1007/978-3-319-93417-4{_}31.

[63] BPR4GDPR, Business Process Re-engineering and functional toolkit for GDPR compliance, 2018. https://www.bpr4gdpr.eu/.

[64] G. Lioudakis and D. Cascone, Compliance Ontology - Deliverable D3.1, Project deliverable, 2019. https://www.bpr4gdpr.eu/wp-content/uploads/2019/06/D3.1-Compliance-Ontology-1.0.pdf.

[65] P.A. Bonatti, B. Bos, S. Decker, J.D. Fernandez, S. Kirrane, V. Peristeras, A. Polleres and R. Wenning, Data Privacy Vocabularies and Controls: Semantic Web for Transparency and Privacy, in: *Semantic Web for Social Good (SWSG2018) @ ISWC2018*, CEUR Workshop Proceedings, 2018.

[66] M.C. Suárez-Figueroa, A. Gómez-Pérez and M. Fernández-López, The NeOn Methodology for Ontology Engineering, in: *Ontology Engineering in a Networked World*, M.C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta and A. Gangemi, eds, Springer Berlin Heidelberg, 2012, pp. 9–34. ISBN 978-3-642-24794-1. doi:10.1007/978-3-642-24794-1_2. https://doi.org/10.1007/978-3-642-24794-1{_}2.

[67] SPECIAL, Home - SPECIAL, 2019. https://www.specialprivacy.eu/.

[68] H.J. Pandit, A. Polleres, B. Bos, R. Brennan, B. Bruegger, F.J. Ekaputra, J.D. Fernández, R.G. Hamed, E. Kiesling, M. Lizar, E. Schlehahn, S. Steyskal and R. Wenning, Creating a Vocabulary for Data Privacy: The First-Year Report of Data Privacy Vocabularies and Controls Community Group (DPVCG), in: *On the Move to Meaningful Internet Systems: OTM 2019 Conferences*, Vol. 11877, H. Panetto, C. Debruyne, M. Hepp, D. Lewis, C.A. Ardagna and R. Meersman, eds, Springer International Publishing, 2019, pp. 714–730, Series Title: Lecture Notes in Computer Science. ISBN 978-3-030-33245-7 978-3-030-33246-4. doi:10.1007/978-3-030-33246-4_44. http://link.springer.com/10.1007/978-3-030-33246-4{_}44.

[69] R.J. Cronk, Categories of personal information, 2017, Publication Title: Enterprivacy Consulting Group. https://enterprivacy.com/2017/03/01/categories-of-personal-information/.

[70] M. Lizar and D. Turner, Consent Receipt Specification v1.1.0, Technical Report, 2017. https://kantarainitiative.org/confluence/display/infosharing/Consent+Receipt+Specification.

[71] O.o. Publications on Eur-Lex, EU Vocabularies - European Legislation Identifier (ELI), 2017. https://op.europa.eu/en/web/eu-vocabularies/eli.