Creating A Vocabulary for Data Privacy *

The First-Year Report of Data Privacy Vocabularies and Controls Community Group (DPVCG)

Harshvardhan J. Pandit[1], Axel Polleres[2], Bert Bos[3], Rob Brennan[4], Bud Bruegger[5], Fajar J. Ekaputra[6], Javier D. Fernández[2], Roghaiyeh Gachpaz Hamed[1], Elmar Kiesling[6], Mark Lizar[7], Eva ~~Schlehan~~Schlehahn[5], Simon Steyskal[8], and Rigo Wenning[3]

[1] Trinity College Dublin, Ireland
[2] Vienna University of Economics and Business, Austria
[3] W3C/ERCIM
[4] Dublin City University, Ireland
[5] ~~Unabhängige~~ Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Germany
[6] Vienna University of Technology, Austria
[7] OpenConsent/Kantara Initiative, United Kingdom
[8] Siemens, Austria

Abstract. Managing privacy and understanding handling of personal data has turned into a fundamental right, at least within the European Union, with the General Data Protection Regulation (GDPR) being enforced since May 25th 2018. This has led to tools and services that promise compliance to GDPR in terms of consent management and keeping track of personal data being processed. The information recorded within such tools, as well as that for compliance itself, needs to be interoperable to provide sufficient transparency in its usage. Additionally, interoperability is also necessary towards addressing the right to data portability under GDPR as well as creation of user-configurable and manageable privacy policies. We argue that such interoperability can be enabled through agreement over vocabularies using linked data principles. The W3C Data Privacy Vocabulary and Controls Community Group (DPVCG) was set up to jointly develop such vocabularies towards interoperability in the context of data privacy. This paper presents the

* Corresponding authors: Harshvardhan J. Pandit `pandith@tcd.ie` and Axel Polleres `axel.polleres@wu.ac.at`. We thank all members of the W3C DPVCG for their feedback and input to this work: a preliminary outline of the goals of CG has been presented in ISWC2018's SWSG workshop [**?**] where we also gathered valuable feedback by the participants; this work is the first complete presentation of the resulting, proposed vocabulary elaborated by the DPVCG since. This work was supported by the European Union's Horizon 2020 research and innovation programme under grant 731601 (SPECIAL), by the Austrian Research Promotion Agency (FFG) under the projects "EXPEDiTE" and "CitySpin", by the ADAPT Centre for Digital Excellence funded by SFI Research Centres Programme (Grant 13/RC/2106), and co-funded by European Regional Development Fund.

resulting Data Privacy Vocabulary (DPV), along with a discussion on its potential uses, and an invitation for feedback and participation.

Keywords: Privacy · GDPR · Interoperability · Semantic Web

1 Introduction

Concerns regarding privacy and trust have been raised to a point where regulators, citizens, and companies have started to take action. Services on the Web are often very complex orchestrations of co-operation between multiple actors, and the processing of personal data in Big Data environments is becoming more complex while being less transparent.

Yet, while from a legal point of view, the adoption of the General Data Protection Regulation (GDPR) [**?**] in April 2016, as well the California Consumer Privacy Act (CCPA) [**?**] of 2018 regulate processing of personal data, their technical implementation in operative IT systems is far from being standardised. While building privacy-by-design [**?**] into systems is a much wider scope, we lack the tools, standards, and best practices for those wanting to be good citizens of the Web to provide interoperable and understandable privacy controls, or to keep records of data processing in an accountable manner, with the possible exception of work on permissions [**?**] and tracking protection [**?**], but even those only cover partial aspects.

To this end, the work presented in this paper aims to set a basis for the establishment of *interoperable standards* in this domain. In particular, it addresses the following gaps by complementing existing (W3C) standards:

- There are no standard vocabularies to describe and interchange personal data. Such vocabularies are relevant, for instance, to support data subjects' right to data portability under Article 20 of the GDPR [**?**].
- There are no agreed upon vocabularies or taxonomies for describing *purposes* of personal data handling and *categories of processing*: the GDPR requires legal bases for data processing, including consent, be tied to the specific *purposes* and *processing* of personal data to justify their lawful use. Consequently personal data processing should be logged with a standard reference to a purpose which complies with the norms set by the legal bases - such as the individual's consent. The concrete taxonomies for representing this information in the context of personal data handling are not yet standardised.
- There are no agreed upon vocabularies or ontologies that align the *terminology of privacy legislations* - such as the GDPR, to allow organisations to claim compliance with such regulations using machine-readable information.

The herein presented Data Privacy Vocabulary (DPV) aims at addressing these challenges by providing a comprehensive, standardized way set of terms for annotating provacy policies, consent receipts, and - in general - records of personal data handling. To this end, the rest of the paper is structured as follows: Section 2 explains the setup and governance of the DPV Community group

within the World Wide Web Consortium (W3C) whereafter Section 3 summarizes pre-existing relevant vocabularies and standards that served as inputs. Section 4 describes the methodology that we applied in reconciling these towards the DPV vocabulary. The vocabulary itself and its modules are described in Section 5 (ommitting detailed descriptions of all classes and properties, which can be found in the published W3C CG draft at `https://www.w3.org/ns/dpv`). We close with a discussion of applications and adoption (Section 6) followed by conclusions and a call for participation and feedback (Section 7).

2 DPVCG: Data Privacy Vocabularies and Controls CG

To address the gaps mentioned in Section 1, a W3C workshop was announced[9], which received 32 position statements and expressions of interest. These were used to create an agenda based on standards and solutions for interoperable privacy. The workshop took place on 17th and 18th April 2018 in Vienna and consisted of about 40 participants. Discussions and interactions were structured into sessions around the four themes of: (1) 'relevant vocabularies and initiatives', (2) 'industry perspective', (3) 'research topics', and (4) 'governmental side and initiatives'. The workshop concluded with a discussion of the next steps and priorities in terms of standardisation and interoperability. The identified goals were (from highest to lowest priority): taxonomies for regulatory privacy terms (including GDPR), personal data, purposes, disclosure and consent (as well as other legal bases), details of anonymisation (and measures taken to protect personal data), and for recording logs of personal data processing.

Following this, a W3C Community Group (CG) with the title '*Data Privacy Vocabularies and Controls CG*' (DPVCG) was formally established on 25th May 2018 - the implementation date of the GDPR. The group has a total of 55 participants to date representing academia, industry, legal experts, and other stakeholders. Its discussions are open via the public mailing list[10], along with a wiki[11] documenting meetings, resources, general information.

The CG had its first face-to-face meeting on 30th August 2018 co-located with the MyData 2018[12] conference at Helsinki, Finland. The goal of this meeting was agreement on the first steps and deliverables of the CG as well as establishment of meeting and management procedures. The outcome of this meeting was agreement on working towards the following deliverables:

- **Use cases and requirements:** Collect and align common requirements from industry and stakeholders to identify areas where interoperability is needed in the handling of personal data. The outcome of this was a prioritised list of requirements to enable interoperability in the identified use-cases.

[9] `https://www.w3.org/2018/vocabws/`
[10] `https://lists.w3.org/Archives/Public/public-dpvcg/`
[11] `https://www.w3.org/community/dpvcg/wiki/Main_Page`
[12] `https://mydata2018.org/`

- **Alignment of vocabularies and identification of overlaps:** Collect existing vocabularies and standardisation efforts, and identify their overlaps and suitability for covering the requirements prioritised in step one. The identified vocabularies are presented in Section 3.
- **Glossary of GDPR terms:** An understandable and interoperable glossary of common terms from the GDPR and an analysis of how they are covered by the agreed vocabularies.
- **Vocabularies:** Based on the heterogeneity or homogeneity of identified use-cases and requirements, create a set of (modular) vocabularies for exchanging and representing information in an interoperable form for personal data, purposes, processing, consent, anonymisation, and transparency logs. The resulting vocabulary is presented in Section 5.

A second face-to-face meeting was conducted on 3rd and 4th December 2019 at Vienna, Austria. The goal of this meeting was to analyse the collected use-cases and vocabularies, to establish agreement on the requirements for vocabularies to be delivered, and to plan ahead towards their conception and completion. A third face-to-face was organised on 4th and 5th April 2019 in Vienna and Dublin to finalise the vocabulary and reach an agreement towards the first public draft. The outcome of the meeting was agreement of terms used and its expression using RDF and OWL. The meeting also provided agreement over the namespace of the vocabulary, its hosting, and documentation. After over a year of collaborative effort, the CG published the '*Data Privacy Vocabulary*' (DPV) on 26th July 2019. The CG is currently welcoming feedback for DPV from the community and stakeholders in terms of comments, suggestions, and contributions.

3 Existing and Relevant Vocabularies

Existing relevant use cases and vocabularies were collected and documented in the wiki[13] through individual submissions by CG members. The wiki page for each vocabulary presents a summary, its relevance, covered requirements, uptake, and applicable use-cases. Relevant terms were then identified from each vocabulary and categorised as per requirements. These were used as the basis for discussions regarding terms to be included and aligned in the DPV.

3.1 Existing Standards and Standardisation Efforts

The CG considered several web-relevant standards for terms relevant towards identified requirements: PROV-O [**?**] (and its extension P-Plan [**?**]) for provenance, ODRL [**?**] for expressing policies, vCard [**?**] for describing people and organisations, Activity Streams [**?**] for describing activities on the web, and Schema.org [**?**] for metadata used in description of web pages.

[13] `https://www.w3.org/community/dpvcg/wiki/Use-Cases,_Requirements,_Vocabularies`

The CG also considered standardisation efforts undertaken by bodies relevant to the areas of privacy and interoperability. Classification of Everyday Living (COEL) [?] describes a privacy-by-design framework for the collection and processing of behavioural data with a focus on transparency and pseudo-anonymisation. It was developed by OASIS, which is a non-profit organisation dedicated to the development of open standards.

The ISA^2 is a programme by the by the European Parliament and the Council of European Union for development of interoperable framework and solutions, which includes a set of vocabularies, termed 'Core Vocabularies' [?], for person, business, location, criterion and evidence, and public organisation. IEEE P7012 [?] is a work-in-progress effort to standardise privacy terms in a machine-readable manner for use and sharing on the web.

Consent Receipt [?] is an interoperable standard developed by the Kantara Initiative for capturing the consent given by a person regarding use of their personal data. The standard enables creation of receipts in human as well as machine readable formats for expressing information using pre-defined categories for personal data collection, purposes, and its use and disclosire. However, it does not address the requirements specified by the GDPR.

The Platform for Privacy Preferences Project (P3P) [?] is a (now-abandoned) protocol for websites to declare their intended use of personal data collection and usage with an emphasis on providing users with more control of their personal information when browsing the web. P3P provided a machine-readable vocabulary for websites and users to define their policies, which were then compared to determine privacy actions.

3.2 Vocabularies addressing Privacy and GDPR

The Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance (SPECIAL) is an European H2020 project that uses semantic-web technologies in the expression and evaluation of information for GDPR compliance. SPECIAL has developed vocabularies for expressing Usage Policy [?] and Policy Log [?] in order to evaluate whether the recorded use of personal data is compliant with a given consent.

Mining and Reasoning with Legal Texts (MIREL) is another European H2020 project that uses semantic-web technologies for GDPR compliance. It has developed PrOnto (Privacy Ontology for Legal Reasoning) [?] - a legal ontology of concepts consisting of privacy agents, personal data types, processing operations, rights and obligations.

GDPRtEXT [?] provides a linked data version of the text of the GDPR that makes it possible for links to be established between information and the text of the GDPR by using RDF and OWL. It also provides a thesauri or vocabulary of concepts defined or referred to within the GDPR in a machine-readable manner using SKOS.

GDPRov [?] is an ontology to represent processes and activities associated with the lifecycle of personal data and consent as an abstract model or plan indicating what is supposed to happen, as well as the corresponding activity

logs indicating things that have happened. It extends PROV-O and P-Plan with GDPR-specific terminology. GConsent [?] is an ontology for expressing necessary information for management and evaluating compliance of consent as governed by the obligations and requirements of the GDPR.

Considered ontologies developed prior to implementation of GDPR also include an ontology to express privacy preferences [?], a data protection ontology based on the GDPR [?], and an ontology for expressing consent [?].

4 Methodology

~~Following the collection of vocabularies, relevant terms were documented in the wiki[14], and were used as the basis for further discussion for addressing the requirements. While initially working towards a taxonomy of terms, the necessity of representing relationships and logic led towards an RDF/OWL based ontology.~~

The process of ~~ontology development was~~ vocabulary development was largely shaped by discussions and interactions between CG members, and was (informally) ~~loosely~~ based on NeOn methodology scenarios [?]~~, with the CG using the SPECIAL Usage Policy Language [?] as the base ontology combined with modular ontologies representing personal data categories , purposes, processing , technical and organisational measures, legal basis, and consent~~. The CG decided to work towards creating a generic or top-level vocabulary rather than restricting it to a particular domain or use-case in order to facilitate universal application and adoption. For this, existing work and approaches were analysed to identify terms relevant for describing specific categories of information, such as purposes of processing and personal data.

The ~~aim of the ontology was stated~~ analysis of existing vocabularies revealed a lack of top-level or abstract concepts necessary to provide an extendable mechanism for representing information ~~by providing~~ in a hierarchical structure. Therefore, the CG decided to work towards creating a vocabulary that provided the necessary top-level concepts and relationships in a hierarchical structure. ~~To this end, an analysis of existing vocabularies~~ Agreement over categories of terms to be included in the vocabulary and relevance of existing terms was carried out ~~to determine their suitability, which revealed a lack of top-level concepts which could be readily incorporated. Therefore, the CG created the necessary concepts by inviting contributions and reviewing them through discussions~~through discussions, and documented in the wiki[14].

While initially working towards a hierarchical taxonomy, the need for representing relationships and logic between terms led towards the creation of an ontology, with RDF/OWL being used to provide standardised serialisation. A base vocabulary was created based on the SPECIAL Usage Policy Language [?] to represent a policy, and additional terms were structured to extend them as modular

14

[14] https://www.w3.org/community/dpvcg/wiki/Taxonomy

(sub-)ontologies. Terms were then added in a top-down fashion, based on existing work or its identified absence.

~~The agreement over~~ Documentation of how terms were proposed, discussed, and added was ~~documented~~ recorded through a collaborative spreadsheet hosted on the Google Sheets platform[15]. The spreadsheet contained separate tabs for each 'modular' ontology and a base ontology representing combined their combined usage to represent personal data handling. The columns in the spreadsheet were mapped to semantic web representations, as depicted in Table 1. The vocabulary was created by using the Google Drive API in a script[16] that extracted terms ~~and generated~~ to create documentation of the taxonomy. This was then modified to generate RDF serialisations using rdflib[17] as RDF/OWL was later adopted[18]. and documentation using ReSpec[19].

Plans for additions and changes to the vocabulary will follow a similar approach, where the proposal is documented and agreed upon through the public mailing list or CG meetings.

Table 1. Columns in spreadsheet for generating RDF serialisations and documentation

Column Name	**Description**	**Representation**
Class/Property	If term is Class or Property	*rdfs:Class*/*rdfs:Property*
term	The IRI of the term	as IRI
description	Description or definition	*dct:description*
domain	Domain if it is a property	*rdfs:domain*
range	Range if it is a property	*rdfs:range*
super classes/properties	Parent classes or properties	*rdfs:isSubClassOf*
sub classes/properties	Child classes or properties	N/A
related terms	Terms relevant to this	*rdfs:seeAlso*
how related?	Nature of relation	use as is
comments	Comments used for discussion	N/A
source	The source of the term	*rdfs:isDefinedBy*
date	Date of creation	*dct:created*
status	Status e.g. accepted,proposed	*sw:term_status*
comments	Comments to be recorded	*rdfs:comment*
contributor	dc:creator	*dct:creator*
date-accepted	Date of acceptance	*dct:date*-accepted
resolution	Record e.g. minutes of meeting	as IRI

[15] `https://www.google.com/sheets/about/`
[16] `https://github.com/dpvcg/extract-sheets/`
[17] `https://github.com/RDFLib/rdflib`
[18] In hindsight, a better alternative was mapping languages such as R2RML https://www.w3.org/TR/r2rml/ for creating RDF data from spreadsheets.
[19] `https://github.com/w3c/respec`

5 Data Privacy Vocabulary

As a result of the process above, the '*Data Privacy Vocabulary*' (DPV) has been published on 26th July 2019 at the namespace `http://w3.org/ns/dpv` (for which we will use the prefix `dpv:`) as a public draft for feedback. The current vocabulary provides terms (classes and properties) to annotate and categorise instances of *legally compliant personal data handling*. In particular, DPV provides extensible concepts and relationships to describe the following components (which are elaborated in further sections):

1. Personal Data Categories
2. Purposes
3. Processing Categories
4. Technical and Organisational Measures
5. Legal Basis
6. Consent
7. Recipients, Data Controllers, Data Subjects

These terms are intended to express *Personal Data Handling* in a machine-readable form by specifying the *personal data categories* undergoing some *processing*, for some *purpose*, by *data controller*, justified by *legal basis*, with specific *technical and organisational measures*, which may result in data being shared with some *recipient*.

The vocabulary is built up in a modular fashion, where each 'module' covers one of the above listed aspects, and which is linked together using a core Base Vocabulary.

5.1 Base Ontology

The 'Base Ontology' describes the top-level classes defining a policy for legal personal data handling. Classes and properties for each top-level class are further elaborated using sub-vocabularies, which are available as separate modules and are outlined in subsequent sections. While all concepts in DPV share a single `dpv:` namespace, the modular approach of providing the base ontology as a separate module makes it possible to use sub-vocabularies without the `dpv:PersonalDataHandling` class, for example to refer only to purposes. Exceptions to this are the NACE purpose taxonomy (cf. details Section 5.3) extending the `dpv:Sector` concept in the *Purposes* vocabulary, and the GDPR legal bases taxonomy (cf. details in Section 5.6) extending the top-level `dpv:LegalBasis` class - which are provided under a separate namespaces to indicate their specialisation. The core concepts of the Base Ontology module and their relationships are depicted in Figure 1.

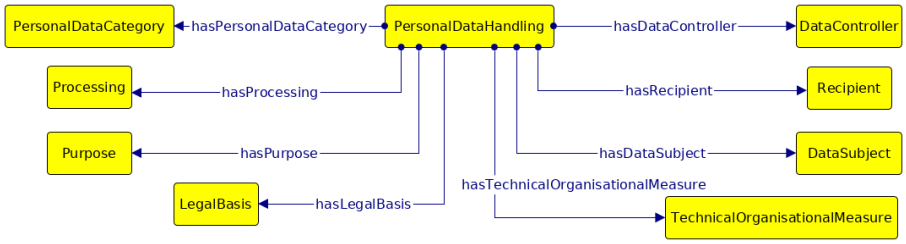

Fig. 1. DPV Base Ontology classes and properties

5.2 Personal Data Categories

DPV provides broad top-level personal data categories adapted from the taxonomy provided by EnterPrivacy [**?**]. The top-level concepts in this taxonomy refer to the nature of information (financial, social, tracking) and to its inherent source (internal, external). Each top-level concept is represented in the DPV as a class, and is further elaborated by subclasses for referring to specific categories of information - such as preferences or demographics.

Regulations such as the GDPR allow information about personal data used in processing to be provided either as specific instances of persona data (e.g., "John Doe") or as categories (e.g., name). Additionally, the class `dpv:SpecialCategoryOfPersonalData` represents categories that are 'special' or 'sensitive' and require additional conditions as per GDPR's Article 9.

The categories defined in the personal data taxonomy can be used directly or further extended to refer to the scope of personal data used in processing. The taxonomy can be extended by subclassing the respective classes to depict specialised concepts, such as "likes regarding movies" or combined with classes to indicate specific contexts. The class `dpv:DerivedPersonalData` is one such context where information has been derived from existing information, e.g., inference of opinions from social media. Additional classes can be defined to specify contexts such as use of machine learning, accuracy, and source.

While the taxonomy is by no means exhaustive, the aim is to provide a sufficient coverage of abstract categories of personal data which can be extended using the subclass mechanism to represent concepts used in the real-world. For instance, Figure 2. shows the hierarchy of concepts for classifying depictions of individuals in pictures.

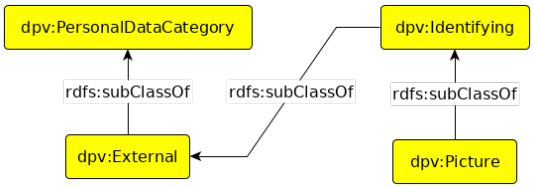

Fig. 2. Hierarchy of concepts for classifying depictions of individuals in pictures (inspired by EnterPrivacy [**?**])

5.3 Purposes

DPV at present defines a hierarchically (by subclassing) organized set of generic categories of data handling *purposes*, as depicted in Figure 3. Overall, DPV provides a list of 31 suggested purposes as subclasses of these generic purposes which may be extended as shown in Listing 1 by further subclassing to create more specific ones. As regulations such as the GDPR generally require a specific purpose to be declared in an understandable manner, we suggest to such declare specific purposes as subclasses of one or several `dpv:Purpose` categories to make them as specific as possible, and to always annotate them with a human readable description (e.g., by using `rdfs:label` and `rdfs:comment`).

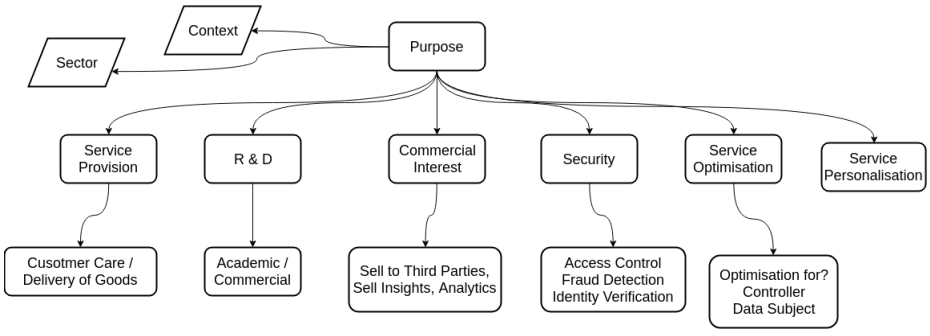

Fig. 3. Categories of Purposes for Data Processing in DPV

```
:NewPurpose
    rdfs:subClassOf dpv:DeliveryOfGoods, dpv:FraudPreventionAndDetection ;
    rdfs:label "New Purpose" ;
    rdfs:comment "Intended delivery of goods with fraud prevention" .
```

Listing 1: Extending pre-defined purposes with human-readable descriptions

Moreover, purposes can be further restricted to specific *contexts* using the class `dpv:Context` and the property `dpv:hasContext`. Similarly, DPVCG provides a way to restrict purposes to a specific *business sector*, i.e., allowing/restricting data handling to purposes related to particular business activities, using the class `dpv:Sector` and the property `dpv:hasSector`. Potential hierarchies for defining such business sectors include NACE[20] (EU), NAICS[21] (USA), ISIC[22] (UN), and GICS[23]. At the moment, we recommend to use NACE (EU) codes using `dpv-nace:NACE-CODE` as shown in Listing 2, where the prefix `dpv-nace:` represents the DPV defined namespace `http://www.w3.org/ns/dpv-nace#`.

```
:SomePurpose a dpv:Purpose ;
    rdfs:label "Some Purpose" ;
    dpv:hasSector dpv-nace:M72 .
```

Listing 2: Creating a new purpose and restricting it to Scientific Research using the NACE sector code (M.72)

5.4 Processing Categories

In this module, DPV provides a hierarchy of classes to specify operations associated with processing of personal data, which are required by regulations such as the GDPR. As common processing operations such as collect, share, and use have certain constraints or obligations in GDPR, it is necessary to accurately represent and define them for personal data handling. While the term 'use' is liberally used to refer to a broad range of processing categories in privacy notices, we recommend to select the most appropriate and specific terms to accurately reflect the nature of processing as applicable.

[20] `https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2`
[21] `https://www.census.gov/eos/www/naics/`
[22] `https://unstats.un.org/unsd/classifications`
[23] `https://en.wikipedia.org/wiki/Global_Industry_Classification_Standard#cite_note-mapbook-1`

DPV defines top-level classes to represent the following broad categories of processing - Disclose, Copy, Obtain, Remove, Store, Transfer, Transform, and Use, as shown in Figure 4. Each of these are then again further expanded using subclasses to provide 33 processing categories, which includes terms defined in the definition of processing in GDPR (Article 4-2).

The DPVCG taxonomy further provides properties with a boolean range to indicate the nature of processing regarding *Systematic Monitoring*, *Evaluation or Scoring*, *Automated Decision-Making*, *Matching or Combining*, *Large Scale processing*, and *Innovative use of new solutions*, as these are relevant towards assessment of processing for GDPR compliance.

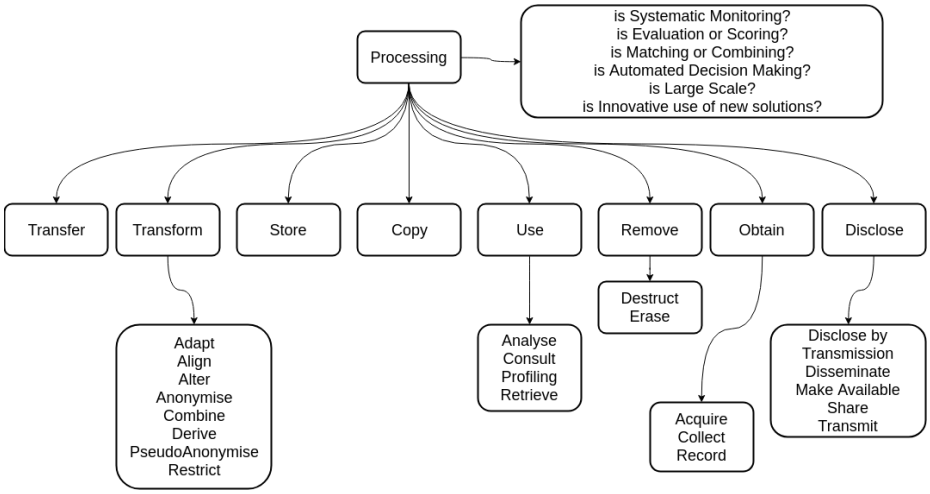

Fig. 4. Categories of Data Processing in DPV

5.5 Technical and Organisational Measures

Regulations require certain technical and organisational measures to be in place depending on the context of processing involving personal data. For example, GDPR (Article 32) states implementing appropriate measures by taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing, as well as risks, rights and freedoms. Examples of measures stated in the article states include:

- the pseudonymisation and encryption of personal data
- the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services
- the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident

– a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing

To address these requirements, DPV defines a module comprising of a hierarchical vocabulary for declaring such technical and organisational measures, as shown in Figure 5.

For any of the DPV declared measures, we provide a generic ObjectProperty (`dpv:measureImplementedBy`), and for the values of this attribute, we either allow a blank node with a single `rdfs:comment` to describe the measure, or a URI to a standard or best practice followed, i.e. a well-known identifier for that standard or a URL where the respective document describes the standard. The class *StorageRestriction* represents the measures used for storage of data with two specific properties provided for storage location and duration restrictions. While at the moment, we do not yet refer to specific certifications or security standards, in the future, we plan to provide a collection of URIs for identifying recommended standards and best practices, as they further develop. Feedback on adding specific ones to future versions of the DPV specification is particularly welcome.

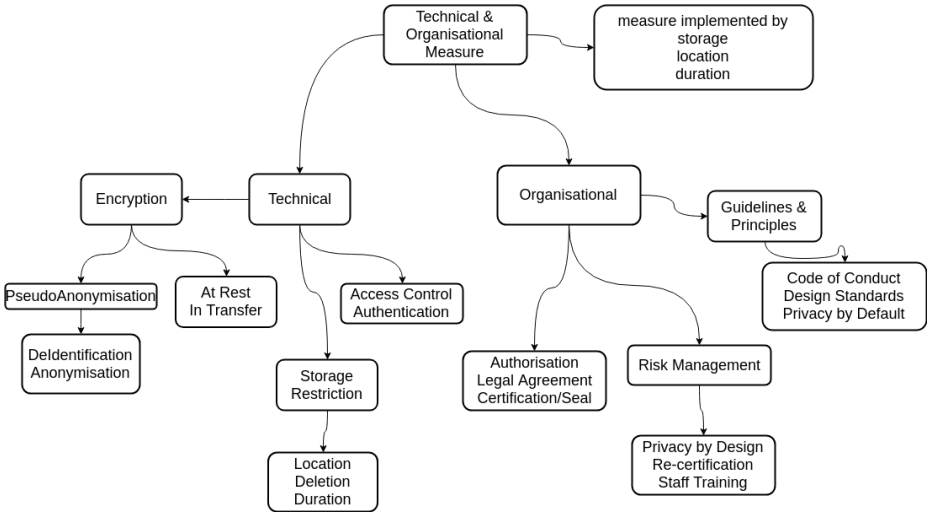

Fig. 5. Technical and Organisational Measures in DPV

5.6 Consent and other Legal Bases

While the vocabulary provides `dpv:LegalBasis` as a top-level concept representing the various legal bases that can be used for justifying processing of personal data, such legal bases may be defined differently in different legislations within the scope of legal jurisdictions. For the particular case of GDPR, we therefore

provide the legal bases specific to GDPR as a separate aligned vocabulary, under the `https://www.w3.org/ns/dpv-gdpr` namespace (prefix: `dpv-gdpr:`).

This vocabulary defines the legal bases defined by Articles 6 and 9 of the GDPR, including consent, along with their description and source within. For example, `dpv-gdpr:A6-1-b` denotes the legal basis provided by *fulfillement/performance of a contract*.

In addition to the legal bases, Consent is addressed with additional properties and classes within the core DPV vocabulary as it is a common form of legal justification across jurisdictions. The module describing consent, illustrated in Figure 6, provides the necessary terms to describe consent provision, withdrawal, and expiry. This is based on an analysis of existing work in the form of Consent Receipt [?] and GConsent [?].

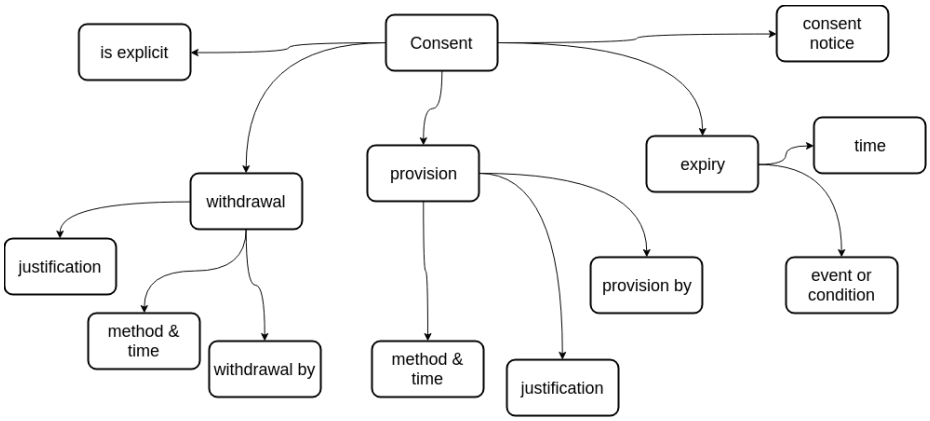

Fig. 6. Consent in DPV

5.7 Recipients, Data Controllers, and Data Subjects

Last but not least, this module of the ontology is meant for defining a taxonomy of stakeholders involved in Personal DataHandling, extending the top level classes `dpv:DataController`, `dpv:DataSubject`, and `dpv:Recipient` from the Base vocabulary module. We consider defining recipients is important in the context of data privacy as it allows tracking the entities personal data is shared/transferred with. Similarly, a categorisation of Data Controllers and Data Subjects has bearing on the privacy of personal data handling, especially when considering situations such as where data subjects are children. The vocabulary currently provides only a few top-level classes to describe such recipients and data subjects, with an invitation to suggest/provide more terms for future releases:

- `dpv:Child` as a subclass of `dpv:DataSubject` in order to capture policies and restrictions of data Handling related to children;

- `dpv:Processor` as a subclass of `dpv:Recipient` to denote natural or legal persons, public authorities, agencies or other bodies which *processes personal data on behalf of the controller*;
- `dpv:ThirdParty` as a subclass of `dpv:Recipient` to provide a generic class for third party recipients, i.e. natural or legal persons, public authorities, agencies or bodies *other than the data subject, controller, processor* and persons who, under the direct authority of the controller or processor, are authorised to process personal data.

6 Potential Adoption and Usage

The primary aim of DPV is to assist in the representation of information concerning privacy in the context of personal data processing. To this end, it models concepts at an abstract or top-level to cover a broad range of concepts. This shall enable the DPV to be used as an domain-independent vocabulary which can be extended or specialised for specific domains or use-cases. Though the DPV does not define or restrict how such extension should be created, this section highlights some suggested methods for its adoption and usage.

Firstly, the modular nature of DPV enables adoption of a selected subset of the vocabulary only to address a specific use-case. For example, an adopter may only wish to utilise the concepts under *Purpose* and *PersonalDataCategory* without using/describing all aspects of a particular *PersonalDataHandling* from the base vocabulary.

In addition, the use of RDFS and OWL enables extending the DPV in a compatible manner to define domain-specific use-cases. For example, an extension targeting the finance domain can define additional concepts by using RDFS' subclass mechanism. Such an extension, when represented as an ontology, will be compatible with the DPV, and will enable semantic interoperability of information, and ideally applications such as automated compliance checking for privacy policies and data handling records annotated with DPV and its extensions.

The DPV is intended to be used as an interoperable vocabulary where terms are structured in a hierarchy and have unambiguous definition to enable common agreement over their semantics. Such usage involves limiting the concepts to other pre-defined vocabulary, as seen in the case of Consent Receipts [**?**] and the SPECIAL vocabularies [**?**].

The SPECIAL project[24] actually has demonstrated how the above-mentioned use case of automated compliance checking can be implemented based modeling privacy policies and log records of personal data handling in a manner compatible with DPV ~~, cf.~~ [**?**]. The SPECIAL project ~~-~~[25] with its industry use case partners may also be viewed as a set of early adopters of the DPV, where currently further tools and a scalable architecture for transparent and accountable personal data processing in accordance with GDPR is being developed.

[24] http://www.specialpricacy.eu
[25] ~~http://www.specialpricacy.eu~~

7 Conclusion

The Data Privacy Vocabulary is the outcome of cumulative effort of over a year in W3C's Data Privacy Vocabulary and Controls Community Group (DPVCG). It represents the first step towards an effort to provide a standardised vocabulary to represent instances of legally compliant personal data handling. To this end, it provides a modular vocabulary representing concepts of personal data categories, purposes of processing, categories of processing, technical and organisational measures, legal bases, recipients, and consent.

With the onset of regulations in the privacy domain, the DPV fills an important gap by providing the necessary terms in an interoperable and extendable format. It is, to the best of our knowledge, currently the most comprehensive vocabulary regarding definition of privacy-related terms in addition to being aligned with regulations such as the GDPR, and attempting to comprehensively cover the relevant aspects of personal data handling. For continued development of this work, the DPVCG is currently inviting participation in the form of comments, feedback, and suggestions. Specifically, the DPVCG kindly requests proposals to extend its initial taxonomies by additional terms, where these are missing or need refinements in order to describe specific use cases of personal data handling.

Future plans also include producing documented examples of how the DPV could be adopted for further specific use-cases. Examples include annotating privacy policies, documenting information for specific laws such as GDPR, and producing transparent, machine-readable processing logs (for instance by mapping the DPV to existing database schemas and thereby generating/aggregating machine-readable transparency records directly out of their logging.