Data Privacy Vocabulary Version 0.1

W3C Community Group Draft 22 April 2019

Please make sure your changes are authored (not-anonymous). We have started work towards first HTML draft, which supersedes this document. See
https://lists.w3.org/Archives/Public/public-dpvcg/2019Jun/0016.html
Current (work-in-progress) draft is at https://dpvcg.github.io/extract-sheets

Editors:
Axel Polleres, Vienna University of Economics and Business
Harshvardhan J. Pandit, Trinity College Dublin

Contributors:
(In alphabetical order)
Bert Bos, W3C/ERCIM
Rob Brennan, Dublin City University
Bud Bruegger, Unabhängige Landeszentrum für Datenschutz Schleswig-Holstein
Fajar J. Ekaputra, Vienna University of Technology
Javier D. Fernández, Vienna University of Economics and Business
Ramisa Gachpaz Hamed, Trinity College Dublin
Elmar Kiesling, Vienna University of Technology
Mark Lizar, OpenConsent/Kantara Initiative
Eva Schlehan, Unabhängige Landeszentrum für Datenschutz Schleswig-Holstein
Simon Steyskal, Siemens
Rigo Wenning, W3C/ERCIM

Abstract

The Data Privacy Vocabulary provides terms (classes and properties) to annotate and categorize instances of legally compliant personal data handling according to the EU General Data Protection Regulation. This scope could be extended by later versions to other data and privacy protection regulations. The vocabulary provides terms to describe which **personal data Categories** are undergoing a specified kind of **processing** by a specific **data controller** and/or transferred to some **recipient** for a particular **purpose**, based on a specific **legal ground** (e.g. consent, or other legal grounds such as legitimate interest, etc.), with specified **technical and organisational measures and restrictions** (e.g. storage locations and storage durations) in place.

Please Send Comments

This document was published by the Data Privacy Vocabularies and Controls Community Group (DPVCG) as a first public working draft. If you wish to make comments regarding this document, please send them to public-dpvcg@w3.org (archives). All comments are welcome.

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. In order to propose/approve new terms, please include the following in your email:

- **term** - suggested URI/identifier
- **class/property** - is the suggested term a class or a property?
- **description** - a natural language description of the term as to be included in the vocabulary to describe the term in an unambiguous manner.
- **domain/range** - for properties include information about domain and range
- **superclasses/subclasses** - for classes include (if applicable) information about known sub-/ or superclasses.
- **superproperties/subproperties** - for properties include (if applicable) information about known sub-/ or superproperties.
- **related terms** - include if applicable otherwise related terms from other vocabularies and explain how they are related (e.g. skos:broader/skos:narrower, rdfs:seeAlso).
- **source** - mention where the new term originates from (e.g. from a legal regulation or from an existing ontology, if possible with a link confirming the reference.

The DPVCG also requests participation regarding open issues and welcomes suggestions on their resolution or mitigation. The list of open issues and their discussions to date can be found at https://www.w3.org/community/dpvcg/track/issues/open

Table of Contents

Introduction

The Data Privacy Vocabulary provides terms (classes and properties) to annotate and categorize instances of legally compliant personal data handling. In particular, the vocabulary provides extensible taxonomies of terms to describe these components, namely:

- Personal Data Categories
- Purposes
- Processing Categories
- Technical and Organisational Measures
- Legal Basis
- Consent
- Recipients, Data Controllers, Data Subjects

These terms are intended to annotate Legal Personal Data Handling in a machine-readable fashion, by specifying which **personal data categories** are undergoing a specific kind of **processing,** by a specific **data controller** and/or shared with some **recipient** for a particular **purpose**, based on a specific **legal ground** (e.g. consent or legitimate interest), with specific **technical and organisational measures and restrictions** (e.g. storage location and storage duration) in place.

Descriptions and respective annotations of Legal Personal Data Handling could be used for example to:

1. declare personal data handling **policies**,
2. declare **consent** for a personal data handling policy (by a specific consent action),
3. transparently **log/document** specific personal data handling actions by a data controller, and finally
4. to support automated **checking of legal compliances** of data handling ex ante (before data handling is actually performed), or ex post (i.e. check compliance to a certain policy and/or consent)

The ‘Base Vocabulary’ 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, e.g. the Personal Data taxonomy. While all concepts within the vocabulary

share a single namespace, the modular approach makes it possible to use the sub-vocabularies without the PersonalDataHandling class, for example to refer only to purposes. Exceptions to this are the NACE purpose taxonomy extending the Sector concept in Purposes vocabulary, and the GDPR legal basis taxonomy extending the top-level Legal Basis class, which are provided under a separate namespace to indicate their specialisation.

ISSUE-2 should we do something like SPECIAL

Namespaces

We use the following RDF namespace for all terms defined in the DPVCG community group:

dpv: http://www.w3.org/ns/dpv#

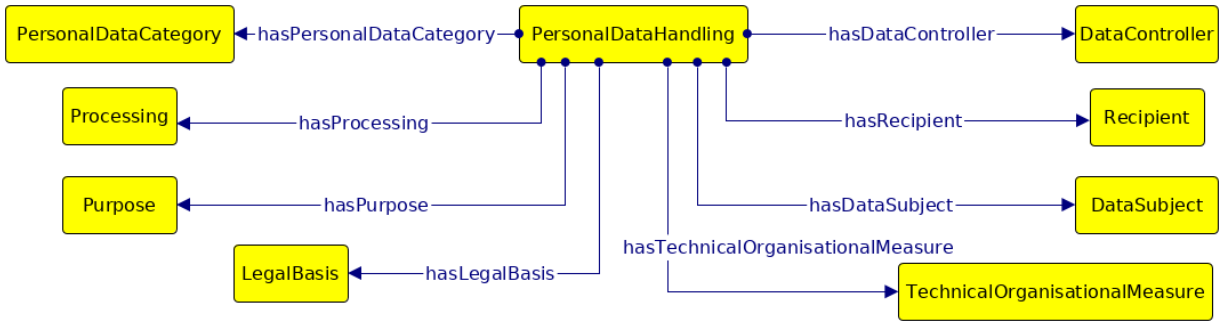
Besides that, the present specification occasionally refers to the following external namespaces:

rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	The normative W3C RDF vocabulary
rdfs	http://www.w3.org/2000/01/rdf-schema#	The normative W3C RDFS vocabulary
owl	http://www.w3.org/2002/07/owl#	The normative W3C OWL vocabulary
skos	http://www.w3.org/2004/02/skos/core#	The normative W3C SKOS vocabulary
dct	http://purl.org/dc/terms/	Dublin core terms
odrl	http://www.w3.org/ns/odrl/2/	Namespace of the Open Digital Rights Language
xsd	http://www.w3.org/2001/XMLSchema#	XML Schema
spl	http://www.specialprivacy.eu/langs/usage-policy#	SPECIAL Log vocabulary cf. SPECIAL Deliverable D2.5
svd	http://www.specialprivacy.eu/vocabs/data#	Namespace for SPECIAL data categories
svpu	http://www.specialprivacy.eu/vocabs/purposes#	Namespace for SPECIAL purposes
svpr	http://www.specialprivacy.eu/vocabs/processing#	Namespace for SPECIAL processing
svr	http://www.specialprivacy.eu/vocabs/recipients	Namespace for SPECIAL recipients
svl	http://www.specialprivacy.eu/vocabs/locations#	Namespace for SPECIAL locations
svdu	http://www.specialprivacy.eu/vocabs/duration#	Namespace for SPECIAL durations
dpv-nace	http://www.w3.org/ns/dpv-nace#	Auxiliary namespace for casting NACE codes as an RDFS hierarchy

dpv-gdpr	http://www.w3.org/ns/dpv-gdpr#	Auxiliary namespace for referring to the legal basis for personal data handling provided by the GDPR

Base Vocabulary

The base vocabulary describes Legal Personal Data Handling using core classes and properties which can be connected to terms defined in the other sub-taxonomies. The class PersonalDataHandling acts as the top class representing an instance of Personal Data Handling, and is connected to other classes using properties to describe specific information such as Personal Data Categories, Purpose, or Processing. The PersonalDataHandling class provides a way to represent a policy for processing personal data using the classes provided in the base vocabulary.



An instance of the PersonalDataHandling class is associated with one or more instances of the Personal Data Category, Purpose, Processing, Recipient, Technical and Organisation Measures, Legal Basis, and Data Controller classes. The vocabulary provides relevant properties for associating each top-level concept with the PersonalDataHandling class, such as hasPurpose and hasDataSubject to associate with Purpose and Data Subject respectively.

Currently, the DPVCG vocabulary does not provide any constraints on the inclusion or exclusion of concepts used to define an instance of PersonalDataHandling. Possibilities for this include use of OWL 2 semantics and SHACL to define concepts mandatory in an instance of PersonalDataHandling. For example, one could define that every instance of PersonalDataHandling *MUST* have at least one Personal Data Category, Controller, Purpose, and Legal Basis.

#import BaseOntology.html

Personal Data Categories

DPVCG 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 DPVCG vocabulary 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 personal data (e.g. “John Doe”) or as categories (e.g. name). Additionally, the class 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 DerivedPersonalData 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.

ISSUE: the property hasPersonalDataCategory has the range PersonalDataCategory, therefore, object of any triple using this property would be inferred as an instance of PersonalDataCategory. Now, how to refer to a personal data category (class) using this property, as it would infer that class to be an instance as well as a subClass of PersonalDataCategory - is this intended and okay as we suggest the following use of personal data taxonomy:

```
dpv:hasPersonalDataCategory dpv:EmailAddress .
```

Other alternative is to scope it within a blank node as:

```
dpv:hasPersonalDataCategory [ a dpv:EmailAddress] .
```

But this would conflate it with instances of email address e.g. <me@harshp.com> a dpv:EmailAddress.

#import PersonalDataCategory.html

Purposes

DPVCG at the moment defines a hierarchically organized set of ***generic*** categories of data handling purposes. Regulations such as GDPR generally require a ***specific*** purpose to be declared in an understandable manner. We therefore suggest to declare the specific context as an instance of one or several dpv:Purpose categories and to always declare the specific purpose with a human readable description (e.g. by using rdfs:label and rdfs:comment).

e.g. :NewPurpose rdfs:subClassOf dpv:DeliveryOfGoods, dpv:FraudPreventionAndDetection ;
rdfs:label “New Purpose” ;
rdfs:comment “Does some processing for delivery of goods and fraud prevention” .

dpv:Purpose rdfs:subClassOf [onProperty rdfs:comment owl:someValuesFrom …]

Purposes can be further restricted to specific contexts using the class dpv:Context and the property dpv:hasContext. DPVCG provides a way to restrict purposes to a specific business sector using the class dpv:Sector and the property dpv:hasSector. Hierarchies for defining business sectors include NACE (EU) (https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2), NAICS (US) (https://www.census.gov/eos/www/naics/), ISIC (UN) (https://unstats.un.org/unsd/classifications), and GICS (https://www.msci.com/gics). Multiple classifications can be used through mappings between sector codes such as NACE - NAICS (https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2). At the moment, we recommend to specify this sector in terms of the EU NACE codes, where the codes can be specified using the following URI scheme -

dpv-nace:*NACE-CODE*

For example, a purpose can be restricted to the domain of scientific research using its corresponding NACE code M72, which can be represented as:

:SomePurpose a dpv:Purpose ;
rdfs:label “Some Purpose” ;
dpv:hasSector dpv-nace:M72 .

We suggest selecting the most appropriate or applicable purpose over more abstract ones by selecting or extending the relevant classes in the purpose taxonomy. For example, the purpose Optimisation can be further clarified to be optimisation for the consumer, or the controller.

#import Purpose.html

Processing Categories

DPVCG provides a hierarchy of classes to specify the operations associated with the processing of personal data. Declaring the processing or processing categories associated with personal data is required by regulations such as the GDPR. Common processing operations such as collect, share, and use have certain constraints or obligations in GDPR, which makes it necessary to accurately represent 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 suggest to select and use appropriate terms to accurately reflect the nature of processing where applicable.

We define top-level classes to represent the following broad categories of processing - Disclose, Copy, Obtain, Remove, Store, Transfer, and Transform. Each of these are then further expanded using subclasses in the taxonomy, which includes terms defined in the definition of processing in GDPR (Article 4-2).

The DPVCG taxonomy 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 - which are relevant towards interpretation of legal requirements such as GDPR compliance.

#import Processing.html

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:

(a) the pseudonymisation and encryption of personal data;
(b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
(c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
(d) 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, the DPVCG defines a hierarchical vocabulary for declaring technical and organisational measures. For any of the declared measures defined below, 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.

In the future, we plan to provide a collection of URIs for identifying recommended standards and best practices. Future versions of this document may extend the vocabulary by an approved taxonomy of identifiers for such standards and best practices.

#import TechnicalOrganisationalMeasure.html

Legal Basis

Regulations such as the GDPR specify certain legal basis for carrying out the processing of personal data, which makes it mandatory for every processing to have one (or more) legal basis that justifies their compliance. DPVCG provides a list of legal bases as per the GDPR under the separate namespace of dpv-gdpr. Additional legal bases can be declared by subclassing dpv:LegalBasis.

The taxonomy lists a total of 17 legal bases as provided by Article 6 and Article 9 of the GDPR. The legal basis of ‘consent’ as defined in Article 6(1)(a) has been declared using the terms ‘explicit’ and ‘non-explicit’ to differentiate the requirements of the two in accordance of their requirements of compliance. Furthermore, legal basis provided by Article 6 apply to processing involving personal data whereas those in Article 9 apply specifically to processing involving special categories of personal data.

#import LegalBasis.html

Consent

Consent is one of the legal bases for the processing of personal data which has several requirements and obligations that determine its validity and use as a legal basis. DPVCG provides the necessary terms and relationships to describe consent and its attributes from a compliance perspective. While the focus of the taxonomy is on the use of consent as a legal basis within the GDPR, the vocabulary can be adapted or extended to define other forms of consent.

The DPVCG consent vocabulary uses the set of properties from the base vocabulary to associate instances of dpv:Consent with Personal Data Categories, Purposes, Processing, Data Controllers, Recipients, Technical and Organisational Measures, and Data Subject. Additionally, the vocabulary also specifies properties for provision (how consent was given) and withdrawal (how consent was withdrawn) to define the provenance of consent. The expiry of consent can be specified using the property dpv:expiry, with sub-properties provided to define expiry as a temporal entity using dpv:expiryTime or as a condition/event using dpv:expiryCondition.

Certain requirements for valid consent require some information to be presented before the consent is obtained in order for it to be considered an ‘informed consent’. This information and the notice it was provided in can be specified using the property dpv:consentNotice. Additionally, the property dpv:isExplicit can be used to specify if the consent is to be interpreted as ‘explicit consent’ using a boolean value (true/false).

To specify consent provided by delegation, such as in the case of a parent or guardian providing consent for a child, the properties dpv:provisionByJustification and dpv:withdrawalByJustification can be used to capture the nature of such ‘delegation’, with the fields dpv:provisionBy and dpv:withdrawalBy capturing the legal entity who provided the consent. By default, the consent is assumed to have been provided by the associated Data Subject.

#import Consent.html

Recipients, Data Controllers, Data Subjects

TODO: add rationale

#import RecipientsDataControllersDataSubjects.html