Network Working Group
Internet-Draft
Intended status: Informational
Expires: 11 December 2025

D. Smullen
B. Scriber
CableLabs

9 June 2025

Privacy Preference Declaration Taxonomy
draft-dsmullen-ppd-taxonomy-01

Abstract

This document defines a standardized taxonomy for describing data handling practices of Internet-connected devices within home networks. It complements the Privacy Preference Declaration (PPD) Protocol by providing the necessary vocabulary and semantic structure to represent and reason about data types, purposes, actions, sources, and destinations. This taxonomy supports both machine reasoning and human interpretation and can be implemented using ontological frameworks such as OWL-DL.

About This Document

This note is to be removed before publishing as an RFC.

The latest revision of this draft can be found at https://drspangle.github.io/draft-dsmullen-ppd-taxonomy/draft-dsmullen-ppd-taxonomy.html. Status information for this document may be found at https://datatracker.ietf.org/doc/draft-dsmullen-ppd-taxonomy/.

Source for this draft and an issue tracker can be found at https://github.com/drspangle/draft-dsmullen-ppd-taxonomy.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 11 December 2025.

Table of Contents

1. Introduction

The effectiveness of the Privacy Preference Declaration (PPD) architecture depends on a shared understanding of the semantics of privacy preferences. (TODO: reference the main architecture i-d here.) A well-structured taxonomy enables the clear articulation of user-defined privacy constraints and provides a common language for devices to report their data handling practices.

This document introduces such a taxonomy, allowing policy declarations to express what kind of data is being handled, why it is being handled, how it is used, where it originates and is sent, and who is involved. These taxonomic categories enable reasoning over complex privacy configurations and enforceable policies.

To support interoperability and consistency, the taxonomy defined herein is coupled with a centralized registry governed by IANA or a designated authority. This registry ensures that all terms used in privacy declarations are semantically defined, unambiguous, and maintained through a community-driven process. The registry plays a critical role in policy validation, enforcement logic, and device interoperability across ecosystems.

2. Conventions and Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Design Goals

* Semantic Clarity: Enable precise and unambiguous expression of privacy concepts.

* Machine Reasoning: Support ontology-based reasoning to detect policy violations or mismatches.

* Extensibility: Allow addition of new concepts without disrupting existing deployments.

* Alignment: Reflect terminology familiar from privacy regulations (e.g., GDPR, CCPA).

* Registry Governance: Ensure terms are publicly documented, versioned, and governed through a formal review process to maintain ecosystem coherence.

* Validation Support: Facilitate automated validation of policies and declarations using machine-readable registry definitions.

* Interoperability: Promote uniform understanding of privacy semantics across diverse vendors and devices, backed by a shared taxonomy registry.

4. Core Taxonomy Structure

The taxonomy consists of five orthogonal but interrelated categories. These categories reference the concepts defined by Contextual Integrity theory in describing data flows. (TODO: cite this properly as an informative reference)

4.1. Data Type (What)

Defines the nature of the data being collected, used, or transmitted.

Examples:

* temperatureReading

* videoCapture

* locationCoordinate

* audioTranscript

* deviceIdentifier

* healthStatus

This dimension aligns with data classification in privacy laws and informs sensitivity.

4.2. Purpose (Why)

Describes the rationale for processing the data.

Examples:

* coreFunctionality (e.g., heating control)

* analytics

* advertising

* securityMonitoring

* personalization

* diagnostics

Each purpose can be mapped to categories of lawful basis for processing under regulations like GDPR (e.g., contract, consent, legitimateInterest).

4.3. Action (How)

Specifies what is being done with the data.

Subclasses:

* collection (retrieval or ingestion)

* usage (processing or decision-making)

* transfer (sharing externally)

* retention (storage over time)

* deletion (erasure procedures)

These actions enable both auditing and fine-grained controls.

4.4. Source and Destination (Where and To Whom)

Defines the data origin and endpoint.

Source may be, in the abstract:

* userInput

* sensor

* inferred (e.g., derived from other data)

* thirdPartyImport

Destination may include, in the abstract:

* localProcessing

* cloudStorage

* manufacturerServer

* thirdPartyPartner

* dataBroker

Concrete examples of these abstract categories of source and destination may also be used, such as [RFC3986]. This classification enables constraints on data flow (e.g., local-only, no third-party sharing).

5. Ontological Representation

To support semantic reasoning, the taxonomy is expressed in OWL-DL (Web Ontology Language - Description Logic). (TODO: perhaps add a normative reference to this?) This allows:

* Inference of non-compliance: e.g., if a device claims to process only temperatureReading for coreFunctionality, but attempts to collect locationCoordinate for advertising.

* Subclassing and equivalence: Allowing extension through subClassOf and equivalentClass definitions.

* Integration with existing vocabularies: Such as Data Privacy Vocabulary (DPV), and schema.org. (TODO: add an informative reference to DPV)

5.1. Example OWL Classes

```
<owl:Class rdf:ID="DataType"/>
<owl:Class rdf:ID="temperatureReading">
  <rdfs:subClassOf rdf:resource="#DataType"/>
</owl:Class>

<owl:Class rdf:ID="Purpose"/>
<owl:Class rdf:ID="coreFunctionality">
  <rdfs:subClassOf rdf:resource="#Purpose"/>
</owl:Class>

<owl:ObjectProperty rdf:ID="hasPurpose">
  <rdfs:domain rdf:resource="#DataHandlingAction"/>
  <rdfs:range rdf:resource="#Purpose"/>
</owl:ObjectProperty>
```

6. Use in Privacy Policies

Policies referencing this taxonomy can be expressed in a structured format such as JSON-LD or RDF/XML, allowing for both enforcement at runtime and static policy analysis. (TODO: add normative references here)

6.1. Sample Policy Statement (JSON-LD)

```
{
  "@context": "http://example.org/ppd-taxonomy",
  "@type": "Policy",
  "appliesTo": "device:smart-thermostat-123",
  "allows": {
    "action": "collection",
    "data": "temperatureReading",
    "purpose": "coreFunctionality",
    "destination": "localProcessing"
  },
  "prohibits": [
    {
      "action": "transfer",
      "data": "temperatureReading",
      "destination": "thirdPartyPartner"
    },
    {
      "action": "collection",
      "data": "locationCoordinate"
    }
  ]
}
```

7. Security Considerations

7.1. Tamper resistance

Devices must not forge or misrepresent declared purposes. Term identifiers MAY include cryptographic hashes for integrity. All entries MUST be tamper-resistant and digitally signed where applicable. Devices SHALL reject policies using unrecognized or invalid terms.

7.2. Immutable references

Policy enforcement relies on exact matching; hash-based identifiers may be used.

7.3. Cross-device reasoning

Shared taxonomy supports detecting conflicts or inconsistencies in multi-device settings.

8. IANA Considerations

This specification requests the creation of a new IANA registry titled:

Privacy Preference Declaration (PPD) Taxonomy Registry

This registry defines structured terms for use in Privacy Preference Declarations, organized across five core categories: DataType, Purpose, Action, Source, and Destination. It is intended to support semantic validation, enforcement, and interoperability in privacy-aware networked systems.

8.1. Purpose and Justification

The Privacy Taxonomy Registry described in this document serves as the authoritative catalog of privacy-related semantic terms used across the Privacy Preference Declaration (PPD) architecture. Managed under the Internet Assigned Numbers Authority (IANA), this registry provides a consistent, governed vocabulary to ensure interoperability, enforcement, and semantic alignment among privacy declarations, devices, and policy engines.

Unlike many traditional IANA registries that define protocol-level constructs such as status codes or media types, the Privacy Taxonomy Registry defines semantically rich terms suitable for reasoning over privacy constraints. Each term is designed to be both human-readable and machine-processable, enabling automated policy enforcement, auditing, and semantic validation in distributed environments like home networks. The registry helps prevent semantic drift, ensures privacy declarations are interpretable across vendor ecosystems, and provides a compliance anchor point for policy analysis and device certification. It supports a unified approach to policy expression that is extensible yet constrained by formal definitions, such as those expressed in OWL-DL. This registry distinguishes itself by supporting semantic reasoning and structured validation, not just name-value mappings. It is foundational to privacy-preserving automation.

8.2. Registry and Extension Mechanism

The success of the Privacy Preference Declaration framework depends on a shared, extensible, and authoritative vocabulary of privacy-related concepts. A unified taxonomy ensures that:

* Users can write meaningful, enforceable policies with well-understood terms.
* Device manufacturers can interpret and comply with these policies using standard semantics.
* Policy processing engines can reason over device declarations and user constraints for compatibility, conflicts, or enforcement.

Without a centralized governance model, the ecosystem risks semantic drift — where different devices interpret similar terms differently, or invent new, incompatible terms — undermining both interoperability and policy clarity.

8.2.1. Taxonomy Registry

A central Privacy Taxonomy Registry SHALL be established and governed by a standards organization (e.g., IETF/IANA, or an independent Privacy Policy Consortium). This registry SHALL:

* Host the canonical definitions for core taxonomy terms.
* Publish OWL-DL and JSON-LD serializations for tooling.
* Allow versioning and deprecation of terms.
* Accept vetted community-submitted extensions via a structured process.
* Provide a human-readable portal and machine-readable API for lookup and validation.

8.3. Registry Structure

Each entry in the registry MUST include the following fields:

* term_id: Globally unique identifier (e.g., ppd:purpose.analytics, ppd:dataType.temperature)
* category: One of: DataType, Purpose, Action, Source, Destination
* definition: Human-readable description of the term

* owl_definition: OWL-DL class or property definition

* examples: At least one real-world usage scenario

* status: Enum: active, deprecated, or experimental

* submitted_by: Name of contributing entity (organization or individual)

* date_registered: ISO timestamp of official inclusion

* version: Semantic versioning identifier (e.g., 1.0.0)

* references: Optional legal or technical citations (e.g., GDPR, RFCs)

All entries MUST conform to this structure and be encoded in both machine-readable (JSON-LD, RDF/XML) and human-readable formats.

8.4. Initial Registry Contents

IANA SHALL initialize the registry with the baseline terms defined in this document’s core taxonomy. These include:

* Core Data Types: temperatureReading, locationCoordinate, audioRecording, videoStream, deviceIdentifier, userPreference, biometricData, healthData, presenceIndicator

* Core Purposes: coreFunctionality, security, personalization, analytics, advertising, diagnostics, regulatoryCompliance

* Core Actions: collection, usage, transfer, retention, deletion

* Core Sources: sensor, userInput, thirdPartyImport, derivedData

* Core Destinations: localProcessing, cloudStorage, manufacturerServer, thirdPartyPartner, dataBroker

Each of these SHALL be registered with complete metadata as described above.

8.5. Registry Management

The registry SHALL be maintained under the “Expert Review” policy defined in [RFC8126]. The designated expert(s) will evaluate submissions for:

* Conformance with the OWL-DL ontology

* Semantic clarity and non-ambiguity

* Necessity and non-duplication

* Privacy and security impact (if applicable)

The review process MUST include a public comment period and the ability to appeal decisions.

8.5.1. Extension Mechanism

To support innovation and domain-specific specialization, the taxonomy MUST allow third parties to register custom terms via a controlled submission process.

8.5.1.1. Extension Requirements

An extension submission MUST:

* Declare a unique namespace (e.g., vendorX:, industryGroupY:).

* Clearly define its relationship to existing concepts (e.g., subClassOf, equivalentClass).

* Include all required registry fields (as above).

* Demonstrate necessity: why existing terms are insufficient.

* Include privacy risk assessment if the term introduces sensitive or novel data practices.

8.5.1.2. Extension and Deprecation Policy

Extensions: Entities MAY submit new terms with a custom namespace (e.g., vendorX:) as long as relationships to core terms (e.g., subClassOf) are clearly declared.

Deprecation: Deprecated terms remain in the registry with status marked as deprecated. These terms MUST NOT be used in future policy declarations but MAY be preserved for historical validation.

8.5.2. Access Methods

The registry SHALL be made publicly available via:

* A web-accessible HTML directory with search and browse capabilities

* A machine-readable API for tool and device integration

* Regular snapshots for offline validation

9. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, DOI 10.17487/RFC3986, January 2005, <https://www.rfc-editor.org/rfc/rfc3986>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/rfc/rfc8126>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

Acknowledgments

TODO acknowledge.

Authors' Addresses

Daniel Smullen
CableLabs
Email: d.smullen@cablelabs.com

Brian Scriber
CableLabs
Email: brian.scriber@computer.org