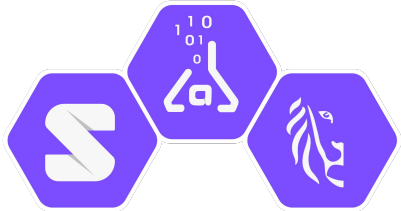




Enhancing Data Use Ontology (DUO) for Health-Data Sharing by Extending it with ODRL and DPV

Harshvardhan Pandit, ADAPT - Dublin City Uni., Ireland
Beatriz Esteves, Uni. Ghent – imec, Belgium



Agenda

1. Introductions
2. Our understanding of DUO and DUC/CCE
3. What is the Data Privacy Vocabulary (DPV)?
4. What is the Open Digital Rights Language (ODRL)?
5. Using ODRL and DPV to improve DUO + DUC/CCE
6. Next Steps / Going Forward

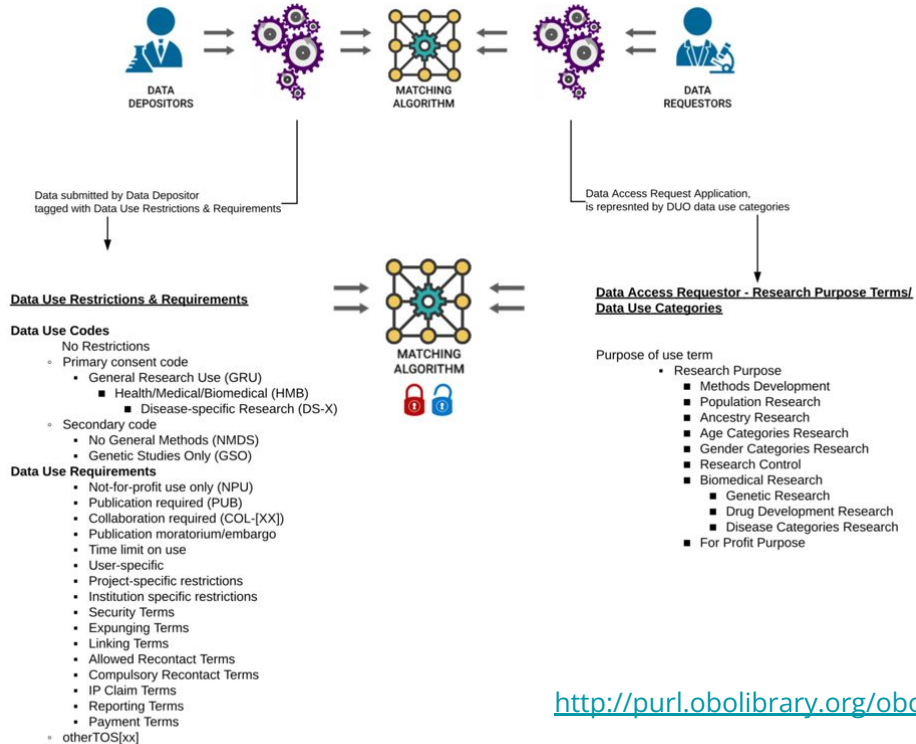
Refer to: **Enhancing Data Use Ontology (DUO) for Health-Data Sharing by Extending it with ODRL and DPV** by H. J. Pandit and B. Esteves. Semantic Web Journal (2024) <https://doi.org/10.3233/SW-243583>

DUO - Data Use Ontology

The (GA4GH) Data Use Ontology (DUO) includes terms describing data use conditions, particularly for research data in the health/clinical/biomedical domain.

Datasets are annotated with semantic information using DUO, and then this is used to 'match' available data with requests using the matching algorithm

This workflow and modelling is based on ***practical*** needs of organisations involved



Data Use Ontology (DUO)

Data Use Limitation

- General Research Use**
Motif: Microscope
ID: DUO_000042
- Health/Medical/Biomedical**
Motif: Hospital
ID: DUO_000006
- Disease Specific**
Motif: Patient and bed
ID: DUO_000007
- Population Origins or Ancestry**
Motif: Family tree
ID: DUO_000011
- No Restrictions**
Motif: Disabled prohibited mark
ID: DUO_000004

Data Use Requirements

- Not-for-Profit Use**
Motif: Dollar and prohibited mark
ID: DUO_000018
- Use Within Geographic Region**
Motif: Pinned map
ID: DUO_000022
- Research Ethics Approval Required**
Motif: Balance
ID: DUO_000021

x2.3

DUO aims to define:

- *permissions e.g. use data for X*
- *prohibitions e.g. don't do Y*
- *constraints e.g. do Z after X*

These 'rules' are defined using DUO concepts to create "*Data Use Limitations*" (DUL) for datasets

DUO uses OWL (so it is semantic-aware) and is linked to OBO

DUO and DULs have both machine- and human-readable meaning e.g. code for machine and text for humans

DUC/CCE for creating "data sharing policies"

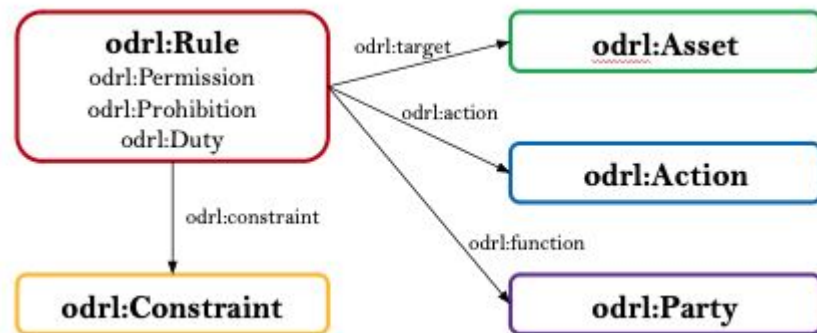
- Common Conditions of Use Elements (CCE): – a controlled vocabulary representing concepts for use in data sharing policies
 1. atomic i.e. each term should represent a single concept
 2. no directionality i.e. term should not specify if reuse or sharing is allowed, forbidden, etc.
 3. generalised i.e. term should be modular category without conditionality, or dependencies
 4. the term should be “widely applicable and relevant”.
- Digital Use Conditions (DUC) – a policy expression mechanism to specify rules regarding conditions for sharing and reuse of datasets
 1. a condition term, which is a CCE concept;
 2. a rule, which is one of Obligatory, Permitted, Forbidden, and No Requirement;
 3. a scope, which is either ‘Whole of asset’ by default or ‘Part of asset’; and
 4. optionally a condition parameter with an optional value.

Challenges in using DUO + DUC/CCE

1. Lacking a formal specification and basis in standards means the onus is on the community to define '*every little thing*' necessary for interoperable communication and interpretation
 - a. how to express different 'concepts' e.g. disease, publishing, organisation type = non-profit
 - b. how to express and interpret 'conditions' e.g. permission and prohibition within each other
 - c. how to 'use' this information e.g. check conditions have been satisfied in the future
2. Not linked to regulations - which vary across jurisdictions
 - a. e.g. EU GDPR has specific requirements for 'notice' and 'consent'
 - b. e.g. EU DGA enables data reuse through intermediaries
 - c. e.g. US/UK regulations may have opt-outs
3. These are the same challenges also faced in 'regtech' by organisations, so what *existing* solutions can we reuse to help GA4GH in its mission?

Open Digital Rights Language (ODRL)

- W3C Recommendation
- Maintained by the W3C ODRL Community Group
- Composed by several specifications
 - ODRL Information Model – W3C Recommendation
 - ODRL Core Vocabulary – W3C Recommendation
 - ODRL Implementation Best Practices
 - ODRL Profile Best Practices
 - ODRL Formal Semantics [Under development]
- Easily extendable through the use of ODRL profiles



Who [**can** | **cannot** | **must**] act **what**
in **which** resource **how**

Aligning DUO concepts with ODRL

Concept	Code	Rule Type	Constraint	Placeholder
DU00000001	Data Use Permission			
DU00000006	HMB	<i>Permission</i>	Purpose is :HMB and not :POA	
DU00000007	DS	<i>Permission</i>	Purpose is :DS and mondo :0000001	:TemplateDisease
DU00000017	Data Use Modifier			
DU00000020	COL	<i>Duty</i>	Action is :CollaborateWithStudyPI	
DU00000016	GSO	<i>Prohibition</i>	Purpose is :GS and not :GSG	
DU00000022	GS	<i>Permission</i>	Spatial is equal to specified :Location	:TemplateLocation
OBI00000066	Data Use Permission			
DU00000034		<i>Permission</i>	Age is specified :Age	:TemplateAgeCategory
DU00000040		<i>Permission</i>	Purpose is :DS and mondo :0000001	:TemplateDisease

Aligning CCE with ODRL

CCE Term	DUO Term	DPV/ODRL Mapping
Use As Control	Research Control	dpv:Purpose
Clinical Research Use	Biomedical Research	dpv:Purpose
Geographical Area + Permitted	Geographical restriction	dpv:Location
Research Use + Permitted	General research	dpv:Purpose
Clinical Care Use + Permitted	Clinical Care Use	dpv:Purpose
Return Of Results + Obligated	Return to database or resource	dpv:Data + dpv:Recipient + odr:Obligation
Collaboration + Obligated	Collaboration required	dpv:Purpose + odr1:Obligation

Enhancing DUO with ODRL

```
:DUO_0000011 a odrl:Set ;  
  rdfs:label "DUO_0000011" ;  
  rdfs:comment "This data use permission indicates that use of the data is limited to the  
  ↪ study of population origins or ancestry (POA - population origins or ancestry research  
  ↪ only)" ;  
  dct:source obo:DUO_0000011 ;  
  odrl:permission [  
    odrl:action odrl:use ;  
    odrl:target :TemplateDataset ;  
    odrl:constraint [  
      odrl:leftOperand odrl:purpose ;  
      odrl:operator odrl:isA ;  
      odrl:rightOperand :POA ] ] ;  
  odrl:prohibition [  
    odrl:action odrl:use ;  
    odrl:target :TemplateDataset ;  
    odrl:constraint [  
      odrl:leftOperand odrl:purpose ;  
      odrl:operator :isNotA ;  
      odrl:rightOperand :POA ] ] .
```

odrl:Set = conditions
odrl:Offer = offer to use data
odrl:Request = request to use data
odrl:Agreement = approved request with one or more offers

Permission = allowed
Prohibition = not allowed
Obligation = MUST be done
Duty = MUST be done after something

Constraints can be any triple
<subject> <relation> <object>

Proof of Concept: matching offers with requests

DUO term	Constraint	Offer	Rule	Request	Decision	Reason
GS	Location	Spain	Permission	Europe	DENY	$Europe \not\subseteq Spain$
GS	Location	Europe	Permission	Spain	GRANT	$Spain \subseteq Europe$
GS	Location	Spain	Prohibition	Europe	DENY	$Europe \cap Spain \neq \emptyset$
GS	Location	Europe	Prohibition	Spain	DENY	$Spain \cap Europe \neq \emptyset$
GS	Location	UK	Prohibition	Spain	GRANT	$Spain \cap UK = \emptyset$
GRU	Purpose	HMB	Permission	DS-Cancer	GRANT	$DS-Cancer \subseteq HMB$
GRU	Purpose	DS-Cancer	Prohibition	HMB	DENY	$HMB \cap DS-Cancer \neq \emptyset$

Fig. 1. Proof-of-concept implementation showing generation of `odrl:Offer` policies

Proof of Concept GUI interface

<https://w3id.org/duodrl/repo>

Dataset policy offer editor
Generate an `odrl:Offer` to express the access conditions to a certain dataset

Target dataset

Data use permission

Data use modifier

```

@prefix dct: <http://purl.org/dc/terms/> .
@prefix duodrl: <https://w3id.org/duodrl/#> .
@prefix odr: <http://www.w3.org/ns/odrl/2/> .

<https://example.com/offer> a odr:Offer ;
dct:source duodrl:DUO_0000006,
duodrl:DUO_0000019 ;
odrl:permission [ odr:action odr:luse ;
odrl:duty [ odr:action odr:ldistribute ;
odrl:output duodrl:ResultsOfStudies ] ;
odrl:target <https://example.com/Dataset> ],
[ odr:action odr:luse ;
odrl:constraint [ odr:leftOperand odr:rpurpose ;
odrl:operator odr:isA ;
odrl:rightOperand duodrl:HMB ] ;
odrl:target <https://example.com/Dataset> ],
[ odr:action odr:luse ;
odrl:constraint [ odr:leftOperand odr:rpurpose ;
odrl:operator duodrl:isNotA ;
odrl:rightOperand duodrl:POA ] ;
odrl:target <https://example.com/Dataset> ] .
                    
```

(a) from DUO concepts

Dataset policy offer editor
Generate an `odrl:Offer` to express the access conditions to a certain dataset

Target dataset

Data use permission

Data use modifier

```

@prefix dct: <http://purl.org/dc/terms/> .
@prefix dpv: <https://w3id.org/dpv/#> .
@prefix duodrl: <https://w3id.org/duodrl/#> .
@prefix odr: <http://www.w3.org/ns/odrl/2/> .

<https://example.com/offer> a odr:Offer ;
dct:source duodrl:DUO_0000042 ;
odrl:permission [ odr:action dpv:Use ;
odrl:constraint [ odr:leftOperand dpv:hasLegalBasis ;
odrl:operator odr:isA ;
odrl:rightOperand dpv:Consent ] ;
odrl:target <https://example.com/Dataset> ],
[ odr:action dpv:Use ;
odrl:constraint [ odr:leftOperand dpv:hasOrganisationalMeasure ;
odrl:operator odr:isA ;
odrl:rightOperand dpv:ImpactAssessment ] ;
odrl:target <https://example.com/Dataset> ],
[ odr:action odr:luse ;
odrl:constraint [ odr:leftOperand odr:rpurpose ;
odrl:operator odr:isA ;
odrl:rightOperand duodrl:GRU ] ;
odrl:target <https://example.com/Dataset> ] .
                    
```

(b) from DUO and DPV concepts

Benefits of ODRL

1. Using ODRL solves the challenges for having a standardised representation
2. It is a W3C standard - so we have interoperability and consistency
3. All tools can use the information in the same manner
4. We have a good way to define how datasets should be represented (odrl:Offer) and requests (odrl:Request) and do the matching between them (odrl:Agreement)

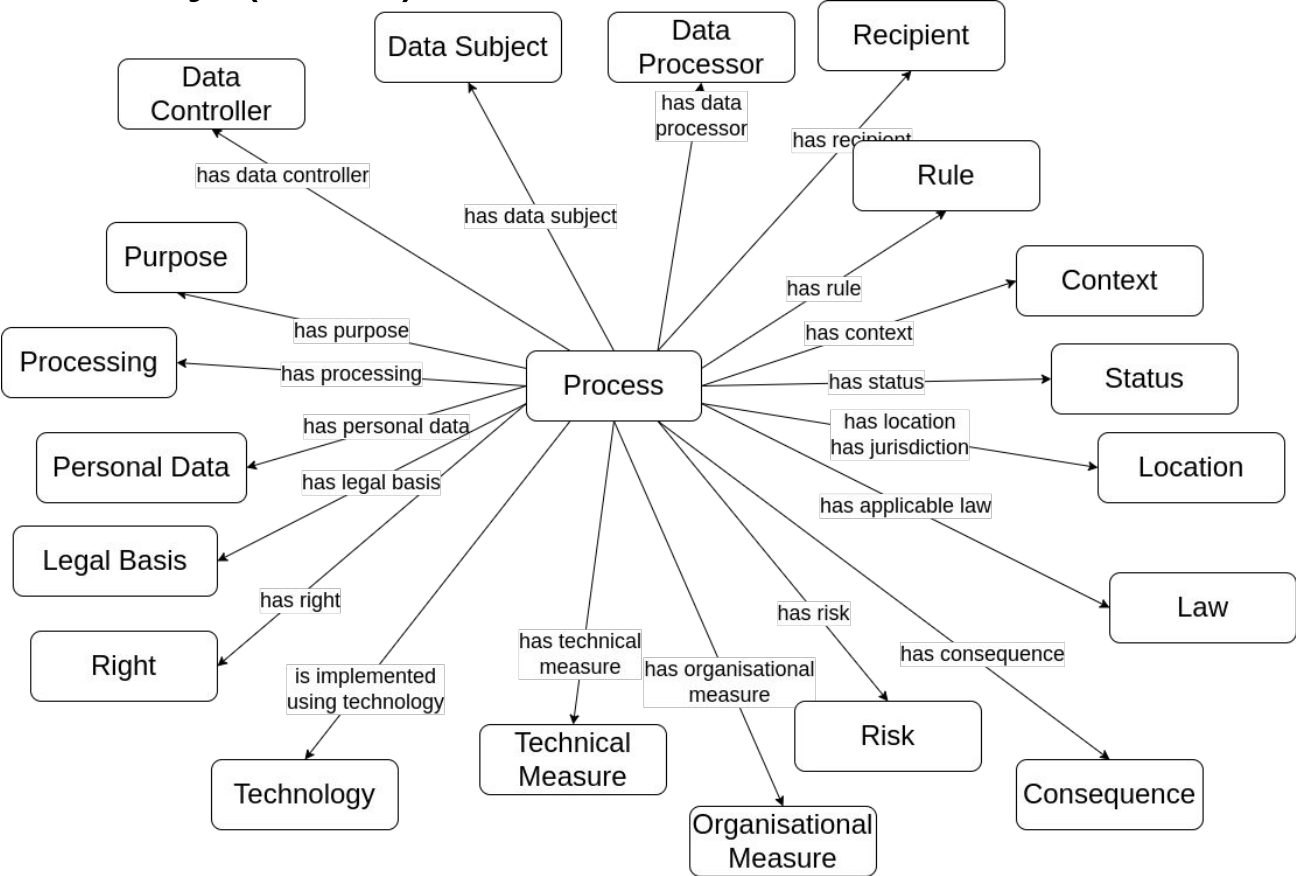
Remaining challenge: Where do we get the concepts like purpose, consent, and the legal stuff from? → DPV!

Data Privacy Vocabulary (DPV)

Developed by the W3C Data Privacy Vocabularies and Controls Community Group (DPVCG)

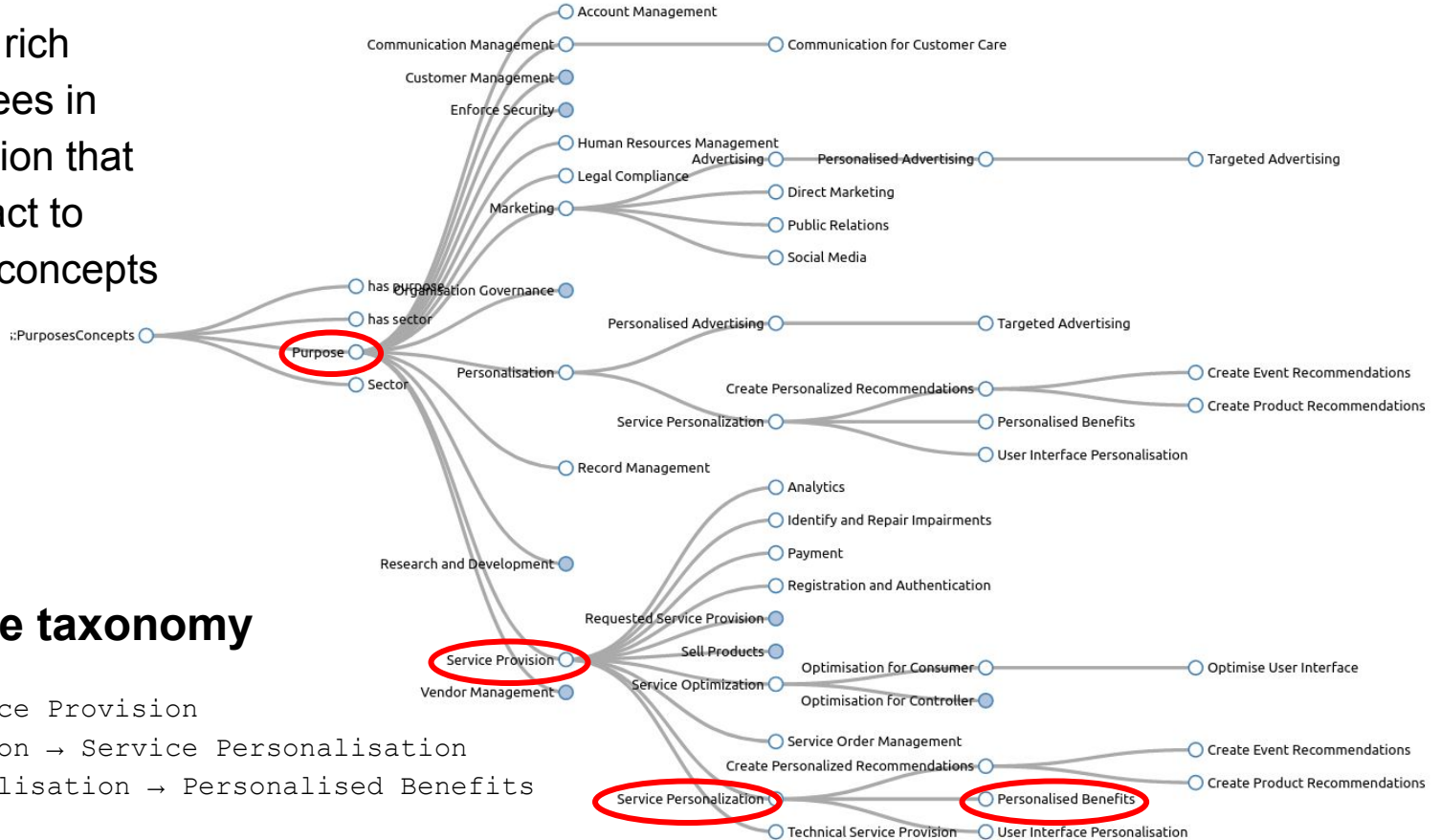
Defines an 'ontology' for defining how data and technologies are used

Provides 'taxonomies' for using the concepts in practical use-cases



DPV Taxonomies

DPV provides rich hierarchical trees in top-down fashion that go from abstract to more specific concepts



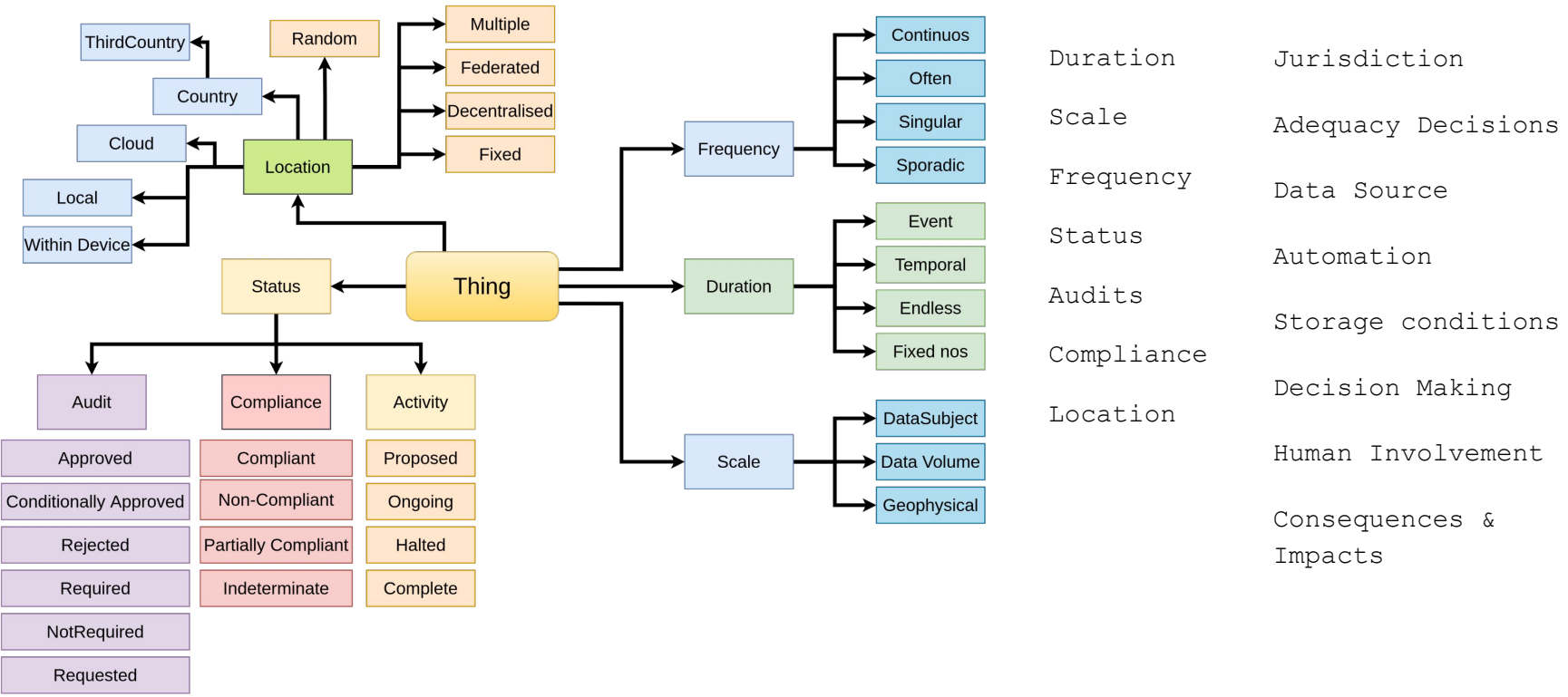
E.g. Purpose taxonomy

Purpose → Service Provision

Service Provision → Service Personalisation

Service Personalisation → Personalised Benefits

Contextual Information

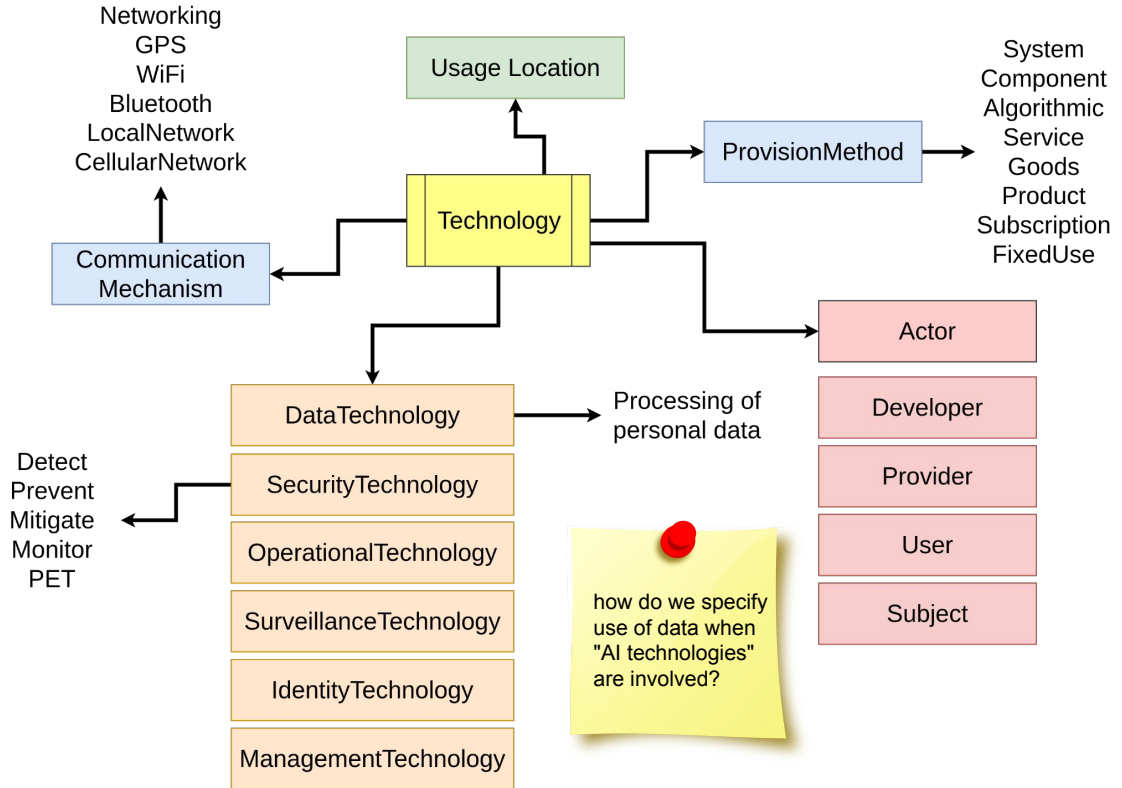


Technologies and Implementations

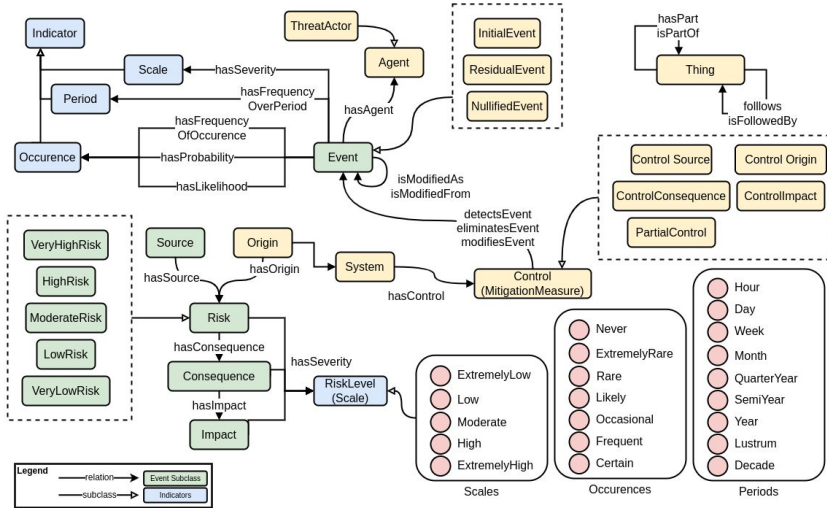
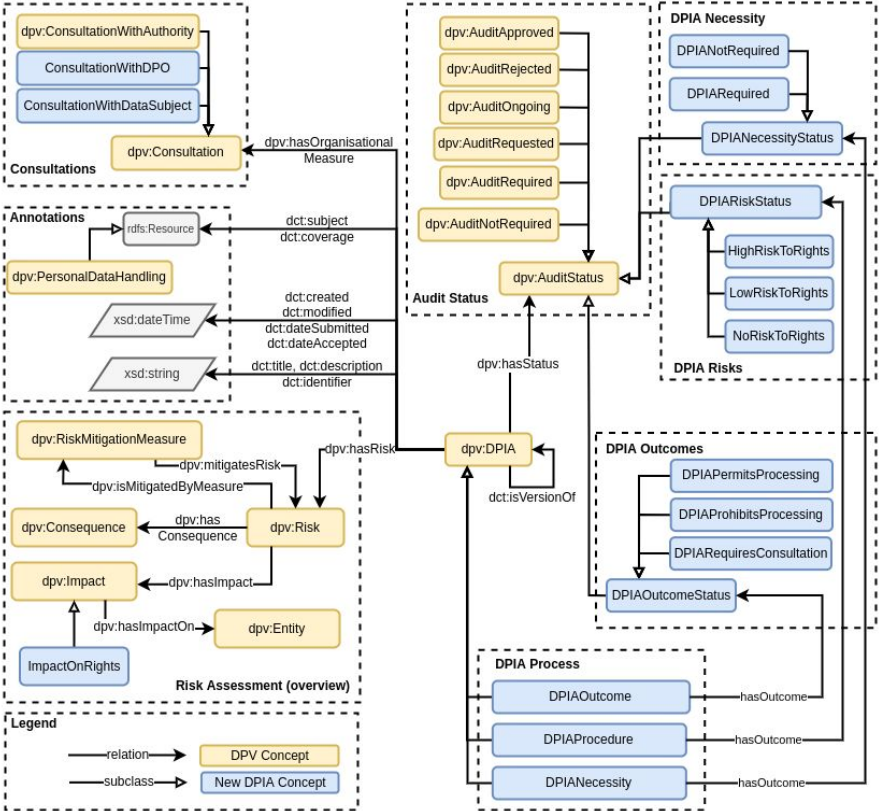
"Technology" is how you implement a process.

It can be a service, tool, or a system you develop, reuse, or purchase from a vendor.

Representing them is important to provide *practical knowledge* about personal data and its processing.

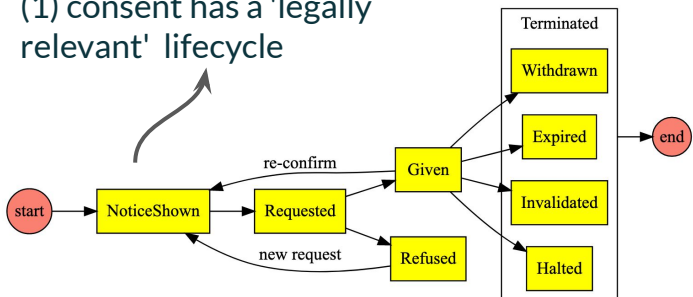


Using DPV for DPIA and Risk Assessments



Maintaining consent records using DPV based on ISO/IEC TS 27560:2023

(1) consent has a 'legally relevant' lifecycle



(2) records are maintained for legal obligation and also organisation's needs, we map ISO with GDPR

Mappings between Clauses

Note: EU GDPR (2016) clauses do not have titles. The table only provides informative descriptions
The notations used as: R42 – Recital 42, 4-11 – Article 4 Point 11.

ISO/IEC TS 27560 (CIB)	ISO/IEC 29184:2020	EU GDPR (2016)
3.1 consent		4-11 definition of consent
3.2 consent receipt	Annex B	R42, 7-1 demonstrating consent
3.3 consent record		R42, 7-1, 13, 14, 30 recording information related to consent
3.4 consent type	5.4.3 Informed and freely given consent. 3.1 explicit consent	R32, R43, 6-1a, 9-2a condition for consent. R42 demonstrating consent. 8 child's consent. 9-2 explicit consent
6.2 recordkeeping for privacy notices and consent		R42, 7-1, 13, 14, 30 recording demonstrating consent
6.2.2.1 presentation of notice	5.2.2 providing notice, 5.2.3 appropriate expression, 5.2.7 appropriate forms, 5.2.9 accessibility	R32, R42, R43, R58, 7-2 notice for consent
6.2.2.2 timeliness of notice	5.2.5 appropriate timing	R32, R42, R43, R60, 7-2, 13, notice for providing information and requesting consent. R61, 1

```

29 "hasDataController": [ { "@id": "ex:C01" }, { "@id": "ex:C02" } ],
30 "hasDataSource": [
31   { "hasProcessing": "dpv:Collect",
32     "hasDataSource": "dpv:DataSubject" },
33   { "hasProcessing": "dpv:Infer",
34     "hasPersonalData": "ServiceUsage" }
35 ],
36 "hasProcessing": {
37   "@type": [ "dpv:Obtain", "dpv:Store" ],
38   "hasLocation": "IE" },
39 "hasStorageCondition": {
40   "hasLocation": "IE",
41   "hasDuration": "63115200"
42 },
43 "hasJurisdiction": "EU",
44 "dct:subject": [ "News Website XYZ Subscription" ],
45 "hasRecipient": [ { "@id": "ex:T01" }, { "@id": "ex:T02" } ],
    
```

(3) tools and automation requires information in interoperable and machine-readable form

Pan-Jurisdictional KG

DPV seems to incorporate several concepts from GDPR, but it is not *exclusive* to it.

DPV concepts are jurisdiction agnostic, with GDPR specific concepts declared in an extension: **EU-GDPR**

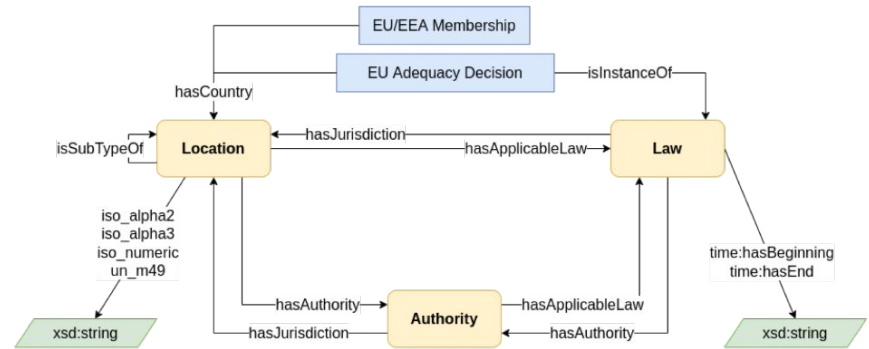
Similarly, other extensions can model different jurisdictions and domains by *extending* from DPV as a base vocabulary.

For specifying instances for different jurisdictions within the same graph, location can be utilised, e.g.

```
<Product> hasJurisdiction B .  
<Controller> hasJurisdiction C .
```

Examples of what DPV can be extended with:

us-ccpa: CCPA/US concepts and requirements
iso: Aligning DPV concepts with ISO terminology
eu-dga: Modelling Data Governance Act (DGA) as an extension of GDPR (via DPV)



The **LEGAL** extension helps with pan-jurisdictional information by providing a graph of laws, authorities, etc;

Enhancing DUO with ODRL and DPV

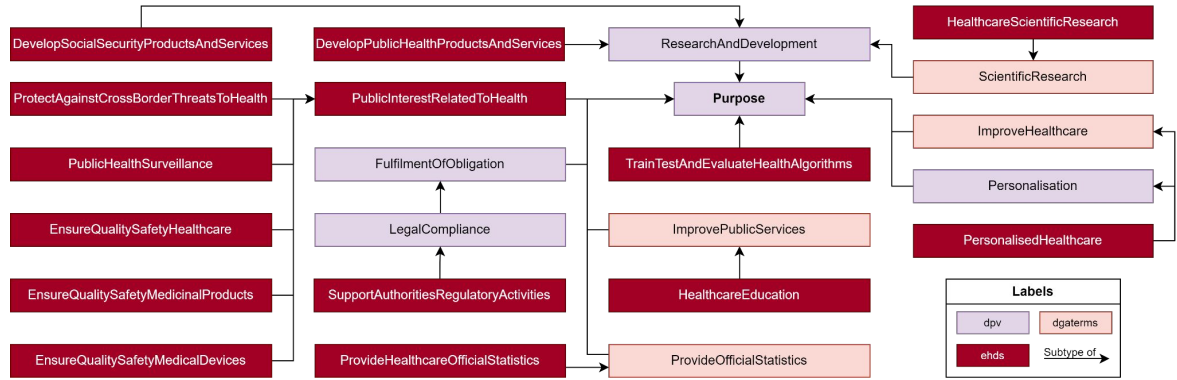
```
PREFIX dpv: <https://w3id.org/dpv#>
PREFIX dpv-legal: <https://www.w3id.org/dpv/dpv-legal#>
PREFIX dpv-gdpr: <https://w3id.org/dpv/dpv-gdpr#>
:Offer1 a odrl:Offer ;
  rdfs:label "Offer to use dataset using Consent, and requiring an Impact Assessment" ;
  odrl:target <https://example.com/Dataset> ;
  odrl:action dpv:Use ;
  odrl:permission [
    odrl:constraint [
      odrl:leftOperand dpv:hasLegalBasis ;
      odrl:operator odrl:isA ;
      odrl:rightOperand dpv:Consent ] ] ;
    odrl:permission [
      odrl:constraint [
        odrl:leftOperand dpv:hasOrganisationalMeasure ;
        odrl:operator odrl:isA ;
        odrl:rightOperand dpv:ImpactAssessment ] ] ;
  ] ;
:Offer2 a odrl:Offer ;
  rdfs:label "Offer to use dataset using GDPR's Explicit Consent, and requiring a DPIA" ;
  odrl:target <https://example.com/Dataset> ;
  odrl:action dpv:Use ;
  dpv:hasApplicableLaw dpv-legal:EU-GDPR ;
  odrl:permission [
    odrl:constraint [
      odrl:leftOperand dpv:hasLegalBasis ;
      odrl:operator odrl:isA ;
      odrl:rightOperand dpv-gdpr:A6-1-a-explicit-consent ] ] ;
    odrl:permission [
      odrl:constraint [
        odrl:leftOperand dpv:hasOrganisationalMeasure ;
        odrl:operator odrl:isA ;
        odrl:rightOperand dpv:DPIA ] ] ;
  ] ;
```

CCE Term	DUO Term	DPV/ODRL Mapping
Use As Control	Research Control	dpv:Purpose
Clinical Research Use	Biomedical Research	dpv:Purpose
Disease Specific Use	Disease Category Research	dpv:Purpose
Geographical Area + Permitted	Geographical restriction	dpv:Location
Research Use + Permitted	General research	dpv:Purpose
Clinical Care Use + Permitted	Clinical Care Use	dpv:Purpose
Return Of Results + Obligated	Return to database or resource	dpv:Data + dpv:Recipient + odrl:Obligation
Collaboration + Obligated	Collaboration required	dpv:Purpose + odrl:Obligation
Time Period + Obligated	Time limit on use	dpv:Duration + odrl:Obligation
Publication Moratorium + Obligated	Publication moratorium	dpv:Purpose + dpv:Duration + odrl:Obligation
Publication + Obligated	Publication required	dpv:Purpose + odrl:Obligation
User Authentication + Obligated	User specific restriction	dpv:TechnicalMeasure + odrl:Obligation
Ethics Approval + Obligated	Ethics approval required	dpv:OrganisationalMeasure + Obligation
(Commercial Entity + Permitted) AND (Profit Motivated Use + Forbidden)	Non-commercial use only	dpv:Purpose + odrl:Rule
Fees	None	odrl:compensate
Regulatory Jurisdiction	None	dpv:Jurisdiction
Return Of Incidental Findings	None	dpv:Data + dpv:Recipient
(Re-)Identification Of Individuals Without Involvement Of The Resource Provider	None	dpv:Processing + dpv:isImplementedBy + dpv:Entity + odrl:Constraint
(Re-)Identification Of Individuals Mediated By The Resource Provider	None	dpv:Processing + dpv:isImplementedBy + dpv:Entity + odrl:Constraint

What about Health Data Spaces?

6. Extensions

- 6.1 Personal Data (PD)
- 6.2 Locations (LOC)
- 6.3 Risk Management (RISK)
- 6.4 Technologies (TECH)
- 6.5 Artificial Intelligence (AI)
- 6.6 Justifications
- 6.7 Legal Concepts (LEGAL)
 - 6.7.1 EU GDPR
 - 6.7.2 EU DGA
 - 6.7.3 EU AI Act
 - 6.7.4 EU NIS2



Legal concepts for Germany (DE)

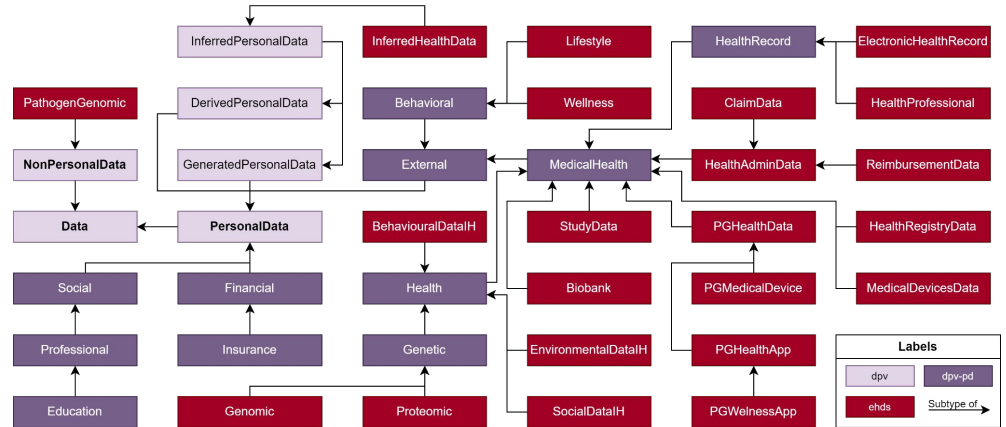
Legal concepts for European Union (EU)

Legal concepts for United Kingdom of Great Britain and Northern Ireland (UK)

Legal concepts for Ireland (IE)

Legal concepts for India (IN)

Legal concepts for United States of America (USA)



In conclusion... does everything need to change?

- Not necessary. DUO and CCE/DUC can be 'mapped' to ODRL + DPV as we have shown to keep existing uses consistent
- Future developments can utilise ODRL+DPV in various ways
- Providing a way for legal compliance with ODRL+DPV is a net benefit for the 'disruption' in change - and also has long term advantages
- DUO and CCE/DUC can be continued as a 'wrapper' around ODRL+DPV defined policies

Read our paper for more info!

Enhancing Data Use Ontology (DUO) for Health-Data Sharing by Extending it with ODRL and DPV

by H. J. Pandit and B. Esteves.

Semantic Web Journal (2024)

<https://doi.org/10.3233/SW-243583>

Enhancing Data Use Ontology (DUO) for health-data sharing by extending it with ODRL and DPV

Cite

Article type: Research Article

Authors: Pandit, Harshvardhan J.^{a*} | Esteves, Beatriz^{b*}

Affiliations: [a] ADAPT Centre, Dublin City University, Ireland | [b] Ontology Engineering Group, Universidad Politécnica de Madrid, Spain


Correspondence: [*] Corresponding author(s). E-mails: me@harshp.com, beatriz.gesteves@upm.es.

Keywords: Health data, biomedical ontologies, policy, regulatory compliance, GDPR

DOI: 10.3233/SW-243583

Journal: *Semantic Web*, vol. Pre-press, no. Pre-press, pp. 1-26, 2024

Published: 14 February 2024

 Get PDF



Abstract

The Global Alliance for Genomics and Health is an international consortium that is developing the Data Use Ontology (DUO) as a standard providing machine-readable codes for automation in data discovery and responsible sharing of genomics data. DUO concepts, which are encoded using OWL, only contain the textual descriptions of the conditions for data use they represent, and do not specify the intended permissions, prohibitions, and obligations explicitly – which limits their usefulness. We present an exploration of how the Open Digital Rights Language (ODRL) can be used to explicitly represent the information inherent in DUO concepts to create policies that are then used to represent conditions under which datasets are available for use, conditions in requests to use them, and to generate agreements based on a compatibility matching between the two. We also address a current limitation of DUO regarding specifying information relevant to privacy and data protection law by using the Data Privacy Vocabulary (DPV) which supports expressing legal concepts in a jurisdiction-agnostic manner as well as for specific laws like the GDPR. Our work supports the existing socio-technical governance processes involving use of DUO by providing a complementary rather than replacement approach. To support this and improve DUO, we provide a description of how our system can be deployed with a proof of concept demonstration that uses ODRL rules for all DUO concepts, and uses them to generate agreements through matching of requests to data offers. All resources described in this article are available at: <https://w3id.org/duodrl/repo>.