



**Engaging Content**  
Engaging People



Trinity College Dublin  
Coláiste na Trionóide, Baile Átha Cliath  
The University of Dublin



# Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance

*Harshvardhan J. Pandit*

*PhD Candidate*

*Supervisor: Dave Lewis*

*Co-supervisor: Declan O'Sullivan*

*School of Computer Science & Statistics*

*Trinity College Dublin*



European Union  
European Regional  
Development Fund



- GDPR compliance is associated with information
  - Technological solutions for GDPR face challenges regarding information representation
- 
- Semantic Web technologies have proved useful in legal compliance and for addressing GDPR
    - Commercial solutions e.g. Top Quadrant, Thomas Reuters, Signatu
    - H2020 projects such as SPECIAL, MIREL, BPR4GDPR
    - Use is growing in legal domain for information representation, querying, reasoning, and interoperability

To what extent can

- information regarding activities
- associated with processing of personal data and consent
- be represented using Semantic Web technologies
- for GDPR compliance?

## Tasks

- 1) Representation of information
- 2) Querying
- 3) Validation
- 4) Assessment/Evaluation

## Information Requirements

- A) to conduct evaluation of compliance
  - B) document compliance
- 
- a) State of system at a given time
  - b) Stakeholders

- 1) Machine-readable representation of GDPR
- 2) Glossary of terms and concepts associated with GDPR
- 3) Representation of activities associated with processing of personal data in ex-ante/ex-post phases
- 4) Representation of consent information
- 5) Demonstration using authoritative compliance queries
- 6) Validating information for compliance evaluation

**RO1** Identify the **subset of GDPR** and **RO2 Identify information** required to represent activities associated with processing of personal data and consent in investigation of GDPR compliance → **Compliance queries and analysis of information (Ch.4)**

major -> **RO3 Create OWL2 ontologies** for representation information about:

**Ontologies presented in Chapter 5**

(a): concepts and text of GDPR → **GDPRtEXT (Sec. 5.2)**

(b): activities associated with processing of personal data and consent → **GDPRov (Sec. 5.3)**

(c): consent required to determine compliance → **GConsent (Sec 5.4)**

minor -> **RO4 Represent compliance questions using SPARQL to query information** about activities associated with processing of personal data and consent → **queries (Sec. 6.1)**

minor -> **RO5 Utilise SHACL to:** (a): **validate information** for GDPR compliance regarding activities associated with processing of personal data and consent (b): link validation results with GDPR → **(Sec 6.2)**

**Querying and Validation of information presented in Chapter 6**

# Research Contributions and Outcomes

<https://openscience.adaptcentre.ie/>

- Information modelling for GDPR compliance
- Linking information to GDPR
- Representing processing activities
- Representing Consent
- Querying information using SPARQL
- Validation framework using SHACL
- Impact on Data Privacy Vocabulary (external contribution)

Name	Type	GDPR	Result	Node
One Processing x Many Purposes	A	R32	F	Q:Consent20190415140000
Personal Data → Storage Period	A	A13-2-a	F	Q:CATQInfoStorageAccess
Personal Data → Storage Period	A	A13-2-a	F	Q:CATTIInfoStorageAccess
Personal Data → Storage Period	A	A13-2-a,R39	F	Q:Consent20190415120753
Personal Data → Storage Period	A	A13-2-a,R39	F	Q:Consent20190415140000
Right to Withdraw	A	A7-3	P	
Separation of Processing	M	R43	P	
Third Party Categories	A	A44	P	
Third Party Identities	A	A13-1-e	P	
Third Party Identities	A	A30-1-d	P	
Third Party Identities	A	A44	P	
Third Party Safeguards	A		P	
Withdraw Consent Information	M	A7-3	P	

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# Research Dissemination and Publication

<https://openscience.adaptcentre.ie/>

## 18 Publications (Sec. 1.4.7)

- 1) GDPRtEXT – ESWC 2018
- 2) GDPRov – ISWC PrivOn 2017
- 3) GConsent – ESWC 2019
- 4) SPARQL queries – SEMANTiCS 2018
- 5) SHACL validation – SEMANTiCS 2018,2019, ISWC CKG 2018
- 6) GDPR interoperability analysis – EURAS 2018, IJSR 2018, Book Chapter 2019
- 7) Investigated applications
  - 1) information management – ICSC 2019, ESWC MEPDaW 2018
  - 2) privacy policies – IPAW 2018, ISWC WOP 2018, TELERISE/ADBIS 2018
- 8) Data Privacy Vocabulary – ODBASE 2019

## Publish developed resources

- FAIR / Open Source / Open Access
- Ontologies: W3ID, Github, Gogs ; Code / Data / Documents: Github, Gogs, Datahub, Zenodo
- Publications: university archive (TARA), Zenodo

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | [pandith@tcd.ie](mailto:pandith@tcd.ie) | [@coolharsh55](https://coolharsh55) | <https://harshp.com/research/phd-thesis>

## 1 W3C Data Privacy Vocabularies & Controls Community Group (DPVCG)

- Author/Editor of Data Privacy Vocabulary (2019)
- Chair since Jan 2020

## 2 EU Publications Office interest in GDPRtEXT

- modelling of clauses, extension of ELI
- discussion of future plans, opinion as POI and research stakeholder, invitation as sem-web expert

## 3 Consent Receipt

- invitation to edit next version of consent standard addressing GDPR and global privacy laws with Kantara
- recently accepted as project at ISO/IEC 27560

## 4 Involved EU / H2020 project funding applications as domain expert

# Future Plans & Opportunities

<https://openscience.adaptcentre.ie/>

- Coverage of more GDPR topics
  - Data Protection Impact Assessments
  - Controller – Processor agreement and data governance
- Data Privacy Vocabulary
  - Expand in terms of concepts, areas, documentation
  - Incorporate other laws e.g. CCPA
- Student Supervision (ongoing)
  - Use of PhD work on topics of privacy and consent online
  - Alignment and application to AI accountability / ethics
- Research applications
  - Policy and Consent
  - ISO 27560 for work on consent receipt
  - Consent extension for Schema.org
- Collaborations with researchers

## Funding Calls – applied/in-progress

- 1) Ireland Postdoc Fellowship (individual)
- 2) H2020 ICT on IoT (WP lead)
- 3) NGI Trust 2020 (co-ordinator)
- 4) NGI DAPSI (team lead)
- 5) ICO (UK) grants (research lead)

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | [pandith@tcd.ie](mailto:pandith@tcd.ie) | [@coolharsh55](https://coolharsh55) | <https://harshp.com/research/phd-thesis>

To what extent can

- information regarding activities
- associated with processing of personal data and consent
- be represented using Semantic Web technologies
- for GDPR compliance?

**RO1** and **RO2** → Compliance queries and analysis of information (Ch.4)

**RO3** → Ontologies presented in Chapter 5

(a): concepts and text of GDPR → **GDPRtEXT** (Sec. 5.2)

(b): activities associated with processing of personal data and consent → **GDPRov** (Sec. 5.3)

(c): consent required to determine compliance → **GConsent** (Sec 5.4)

**RO4** → SPARQL queries (Sec. 6.1)

**RO5** → (Sec 6.2) SHACL for validation and link results to GDPR (Sec 6.2)

Resources: documented and open access at <https://openscience.adaptcentre.ie>

## Impact:

- Data Privacy Vocabulary: community agreement and specification
- Consent Receipt: data model and standard, project at ISO

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance

PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin

CC-BY-NC 4.0 | [pandith@tcd.ie](mailto:pandith@tcd.ie) | [@coolharsh55](https://coolharsh55) | <https://harshp.com/research/phd-thesis>

# Notes Index

<https://openscience.adaptcentre.ie/>

12. GDPR: compliance requirements
13. GDPR: sources of information
14. Ex-ante & Ex-post
15. ISO standards for GDPR
16. Technological development of legal compliance solutions
17. Approaches for legal compliance
18. Commercial solutions for GDPR
19. Timeline of research
20. RQ: definitions
21. Strengths and Weaknesses of my work
22. Publications list
23. Why not ODRL?
24. Are they really OWL2 ontologies?
25. SotA: Scope
26. SotA: Survey of legal approaches
27. Why not BPMN?
28. SotA: Legal Ontologies
29. SotA: Analysis criteria
30. SotA: Approaches using Sem-Web
31. SotA: Approaches other
32. SotA: Approaches for privacy policies
33. SotA: Approaches for consent
34. SotA: upcoming research projects
35. SotA: SPECIAL
36. SotA: MIREL + DAPRECO
37. SotA: BPR4GDPR
38. SotA: Ujcich et al.
39. SotA: RestAssured
40. SotA: Analysis
41. SotA: Analysis – representation of GDPR
42. SotA: Analysis – representation activities
43. SotA: Analysis – representation of consent
44. SotA: Analysis – querying
45. SotA: Analysis – compliance evaluation
46. Analysing GDPR compliance requirements
47. GDPR req. – stakeholder requirements
48. GDPR req. – interoperability model
49. GDPR req. – compliance questions
50. GDPR req. – consultations
51. GDPR req. - categories
52. GDPR req. – use-cases
53. GDPR req. – assumptions / constraints
54. Ontology engineering methodology
55. GDPRtEXT – overview
56. GDPRtEXT – ELI extension
57. GDPRtEXT – script for text extraction
58. GDPRtEXT – DPD mapping
59. GDPRtEXT – compliance reporting
60. GDPRtEXT – evaluation
61. GDPRtEXT – compare with SotA
62. GDPRov – overview
63. GDPRov – ex-ante and ex-post phase
64. GDPRov – Extending PROV-O and P-Plan
65. GDPRov – Data Lifecycle
66. GDPRov – Consent Lifecycle
67. GDPRov – change detection
68. GDPRov – evaluation
69. GDPRov – compare with SotA
70. GDPRov – data storage and location
71. GDPRov – use-case from SPECIAL
72. GConsent – overview
73. GConsent – relationship with GDPRov
74. GConsent – context and concepts
75. GConsent – example use-cases
76. GConsent – evaluation
77. GConsent – compare with SotA
78. GConsent – use-case from SPECIAL
79. DPV – overview
80. DPV – concepts
81. DPV – comparison with phd ontologies
82. Querying – compliance questions
83. Querying – GDPR readiness checklist
84. Querying – methodology
85. Querying – analysis
86. Querying – demo
87. Querying – evaluation
88. Querying – compare to SotA
89. Validation – model
90. Validation – link to GDPR
91. Validation – manually evaluated constraint
92. Validation – ex-ante and ex-post results
93. Validation – demo
94. Validation – Quantcast data
95. Validation – constraints
96. Validation – SPARQL queries
97. Validation – evaluation
98. Validation – compare to SotA
99. Conclusion – fulfilment of RO and RQ
100. Conclusion – extent of sem-web

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GDPRtEXT - overview

<https://openscience.adaptcentre.ie/>

- Ontology to specify GDPR at a granular level + RDF dataset
  - Extends ELI (in a compatible manner)
  - PID individual clause/section
- SKOS glossary of GDPR concepts
  - Personal Data
  - Legal Basis
  - Principles, Rights
  - Compliance obligations

article12-3 :: GDPRtEXT  
<https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3>

Property	Value
<code>o:isDescription</code>	<ul style="list-style-type: none"><li>• The controller shall provide information on action taken on a request under Articles 20 to 22 to the data subject without undue delay and in any event within one month of receipt of the request. That period may be extended by two further months where necessary, taking into account the complexity and number of the requests. The controller shall inform the data subject of any such extension within one month of receipt of the request, together with the reasons for the delay. Where the data subject makes the request by electronic form means, the information shall be provided by electronic means where possible, unless otherwise requested by the data subject.</li></ul>
<code>o:isPartOf</code>	<ul style="list-style-type: none"><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapter1">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapter1</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-2">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-2</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-3">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-3</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-4">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-4</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-5">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-5</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-6">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-6</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-7">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-7</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-8">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-8</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-9">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-9</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-10">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-10</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-11">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-11</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-12">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-12</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-13">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-13</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-14">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-14</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-15">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-15</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-16">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-16</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-17">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-17</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-18">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-18</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-19">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-19</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-20">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-20</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-21">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-21</a></li><li>• <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-22">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapters1-22</a></li></ul>
<code>o:isNumber</code>	<ul style="list-style-type: none"><li>• 3 (xsd:string)</li></ul>
<code>o:titleAlternative</code>	<ul style="list-style-type: none"><li>• Article12-3 (xsd:string)</li></ul>
<code>o:type</code>	<ul style="list-style-type: none"><li>• <code>o:LegalResourceTableSection</code></li><li>• <code>gdptext:Point</code></li></ul>
Metadata	
<code>Anon_3</code>	<ul style="list-style-type: none"><li>• <code>o:isDataItem</code></li><li>• <code>o:isType</code> <a href="http://www.w3.org/2004/03/rdf-schema#Graph">http://www.w3.org/2004/03/rdf-schema#Graph</a></li><li>• <code>o:isPrimaryTopic</code> <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3</a></li><li>• <code>o:hasTopic</code> Anon_0</li><li>• <code>o:realizes</code> <a href="https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3">https://openscience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article12-3</a></li><li>• <code>o:isCreatedBy</code> Anon_3 (xsd:string)</li></ul>
<a href="#">expand all</a>	
This page shows information obtained from the SPARQL endpoint at <a href="https://openscience.adaptcentre.ie/sparql">https://openscience.adaptcentre.ie/sparql</a> . <a href="#">Anon 1</a>   <a href="#">Anon 2</a>   <a href="#">Browse in Dracula</a>   <a href="#">Browse in Tabulator</a>   <a href="#">Browse in OpenLink Browser</a>	

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

- LegalResource → document
- LegalSubResource → section in document (GDPRtEXT uses this)
- has\_part property → specify relation between sections
- FRBR model → allows specification of language and format specific publications
  - E.g. PDF, XML
  - E.g. EN, FR, DE
- GDPRtEXT only models GDPR at a conceptual level.
- FRBR can be extended on top of GDPRtEXT same as ELI
- ELI+ is proposed/planned extension to ELI

- Official GDPR publication lists its contents inside a *table* element
- Does not specify metadata (e.g. clause, number, title)
- JS script parses this output to extract data
- Alternative is to use XML – much more consistent
- Issue highlighted to publication office
- Response:
  - ELI+ will have more structured metadata
  - HTML will also be better structured

- Create LOD version of DPD (assign PID to each clause)
- Create mapping between DPD clause and GDPR clause
  - Based on interpretation of obligation
  - Based on available information at time
- Annotate corresponding XACML statements based on change
  - Extended, same, reduced, completely changed, slightly changed
- Collaboration work (collaborator was leading work on XACML)

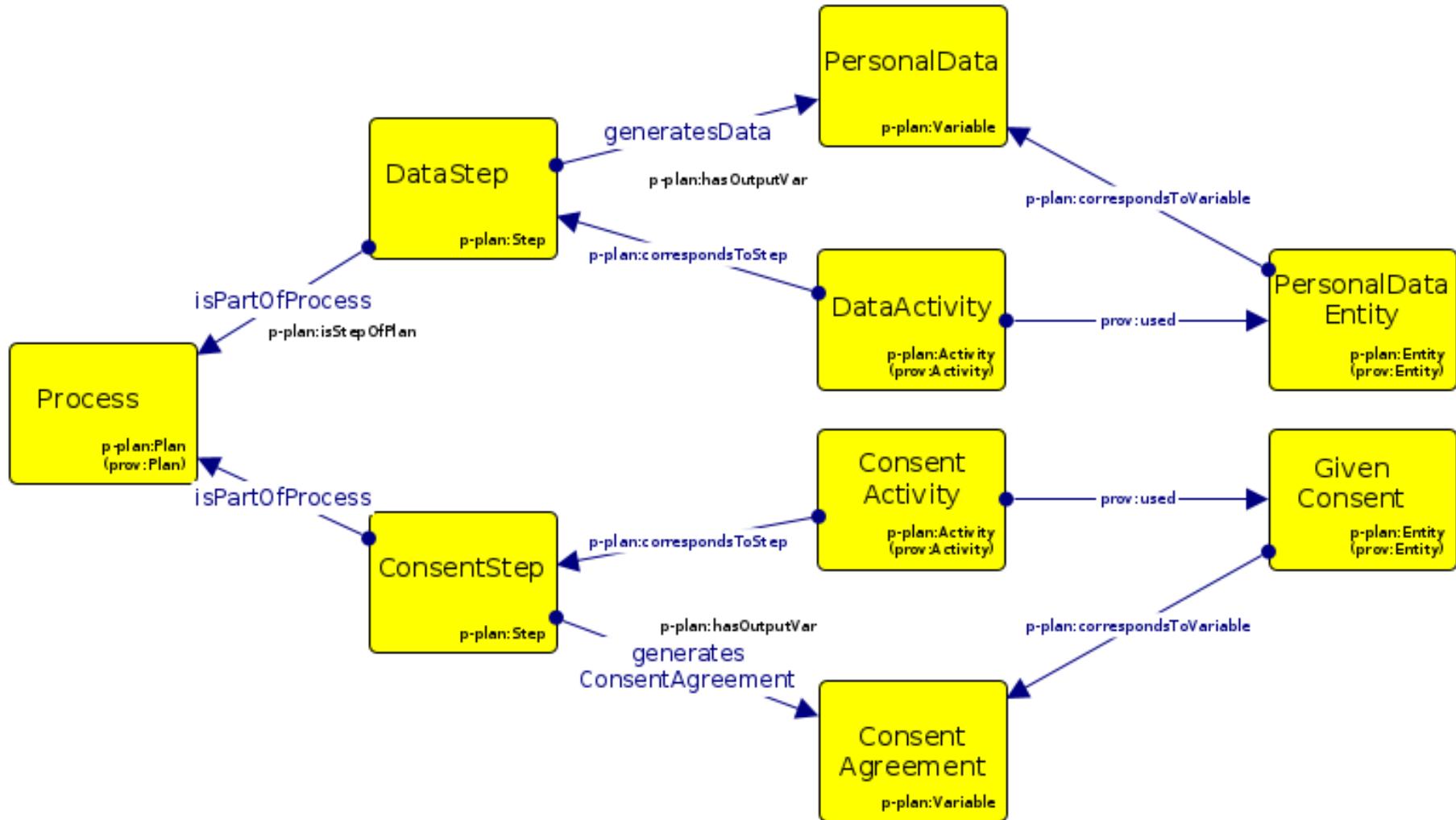
Table 5.2: Comparison of GDPRtTEXT with SotA

Work	GDPRtTEXT	ELI	ELI+	Agarwal et al	PrOnto
Vocabulary	ELI	OWL2	OWL2	RDFS	Akoma Ntoso
Granularity	Sub-Paragraph	Legislation	Sub-Paragraph	Paragraph	Sub-Paragraph
Glossary	✓	✗	✓	✗	✗
PID	✓	✓	✓	✗	✗
OA	✓	✓	✓	✗	✗
GDPR text	✓	✗	✓	✗	✓

- Provenance (of plans and activities) ontology for GDPR
- Extends PROV-O and P-Plan
  - PROV-O → provenance of activities (ex-post) + prov:Plan
  - P-Plan → scientific workflow (ex-ante + ex-post)
- Represents Ex-ante and Ex-post phases
- Models data and consent lifecycles
- Granular abstraction of processes and steps
- Concepts
  - Agents
  - Data
  - Consent
  - Activities (Processing, Purposes)
  - Anonymisation
  - Rights
  - Data Breach
- Composition of activities as steps and processes

# GDPRov – Extending PROV-O and P-Plan

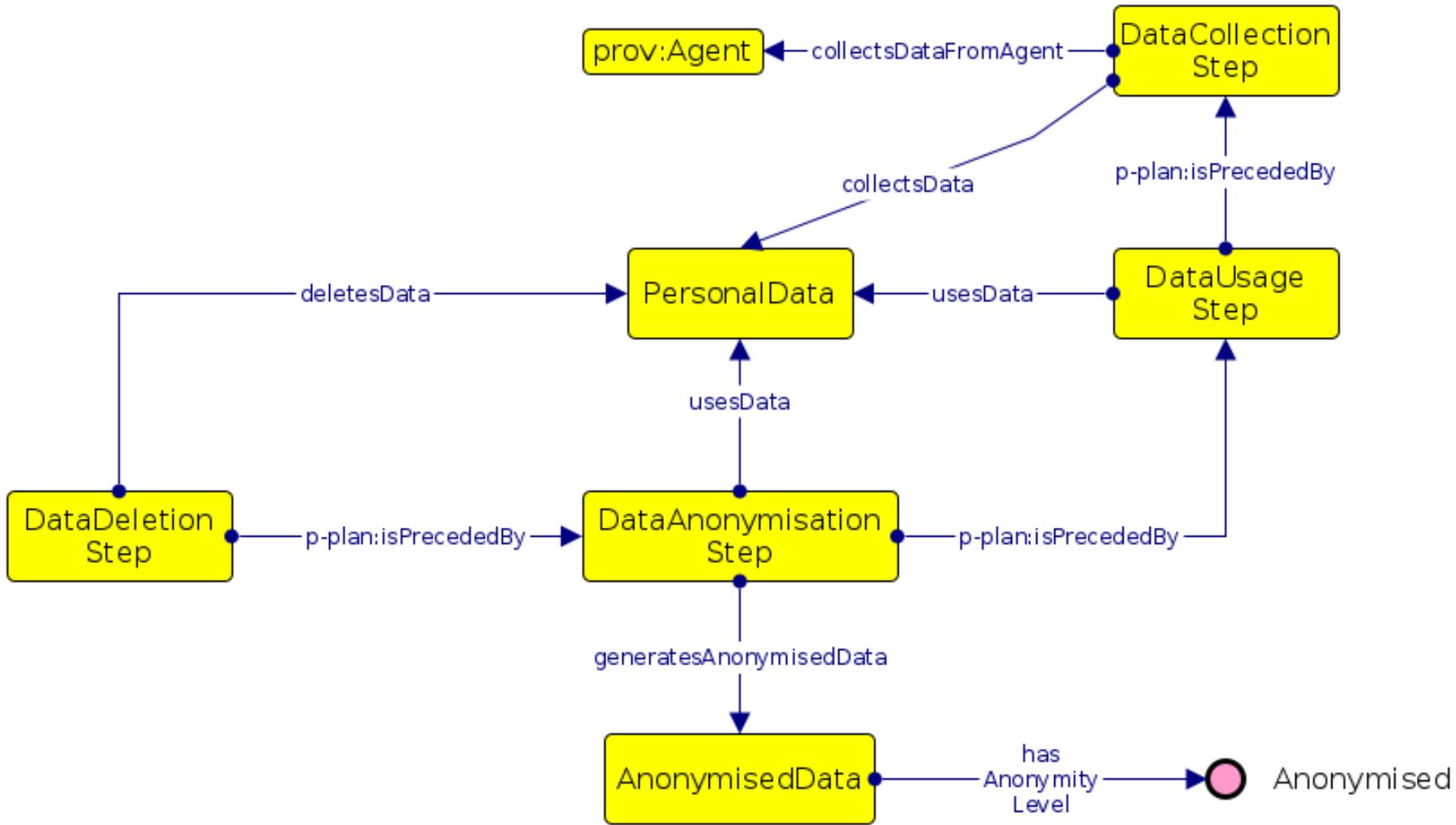
<https://openscience.adaptcentre.ie/>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GDPRov – Data Lifecycle

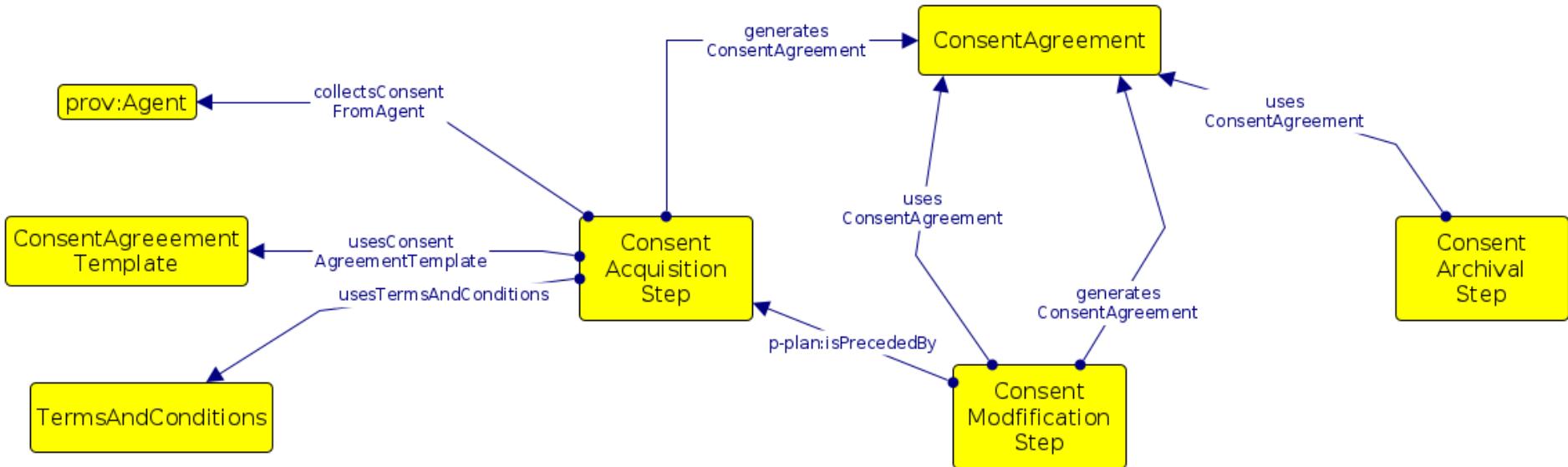
<https://opensource.adaptcentre.ie/>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GDPRov – Consent Lifecycle

<https://opensource.adaptcentre.ie/>

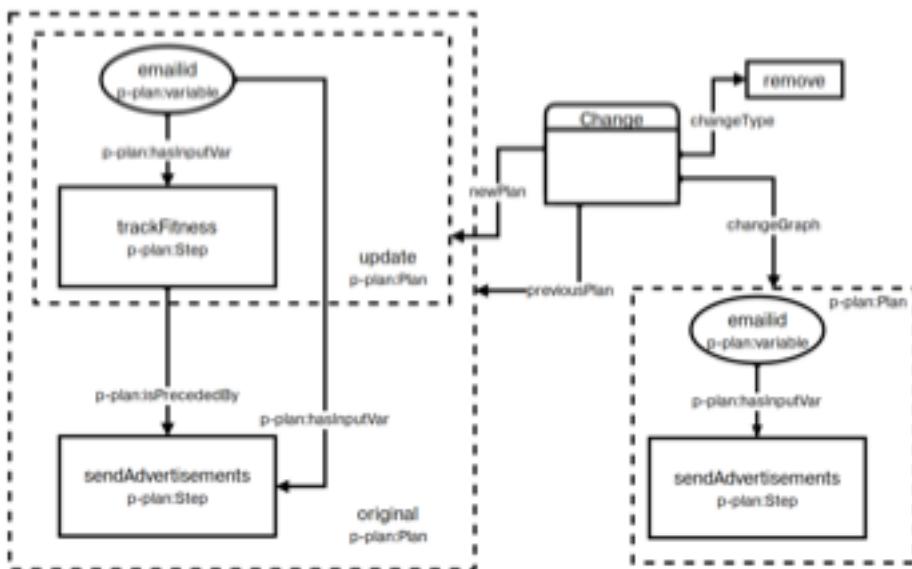


Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GDPRov – Change Detection Use-Case

<https://openscience.adaptcentre.ie/>

- Temporal state of Plan / System
- Compare new and old states
- Identify differences in activities
- Evaluate change in GDPR obligations (e.g. new purposes added)
- Evaluate change required for compliance (e.g. renew consent)



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GDPRov – Compare with SotA

<https://openscience.adaptcentre.ie/>

Table 5.4: Comparison of GDPRov with SotA

Work	Repr	EA	EP	Pu	Pr	DS	Rp	St	Rg	LB
GDPRov	PROV-O,P-Plan	✓	✓	✓	✓	✓	✓	✓	✓	✓
SPECIAL	PROV-O	✓	✓	✓	✓	✓	✓	✓		
SPL+CitySPIN	PROV-O	✓	✓	✓	✓	✓	✓	✓		
MIREL	PWO	✓		✓	✓			✓	✓	
MRL+DAPRECO	PWO	✓		✓	✓			✓	✓	
BPR4GDPR		✓	✓	✓	✓	✓	✓			
Ujcich et al.	PROV-O		✓	✓	✓	✓	✓	✓	✓	✓
Lodge et al		✓		✓						
Tom et al	BPMN	✓			✓	✓	✓	✓	✓	
LUCE		✓	✓			✓	✓			
Sion et al		✓		✓	✓	✓	✓	✓		✓
privacyTracker		✓	✓			✓	✓			
Basin et al		✓		✓						
RestAssured				✓	✓	✓	✓	✓		

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | [pandith@tcd.ie](mailto:pandith@tcd.ie) | [@coolharsh55](https://coolharsh55) | <https://harshp.com/research/phd-thesis>

# GDPRov – Data Storage and Location

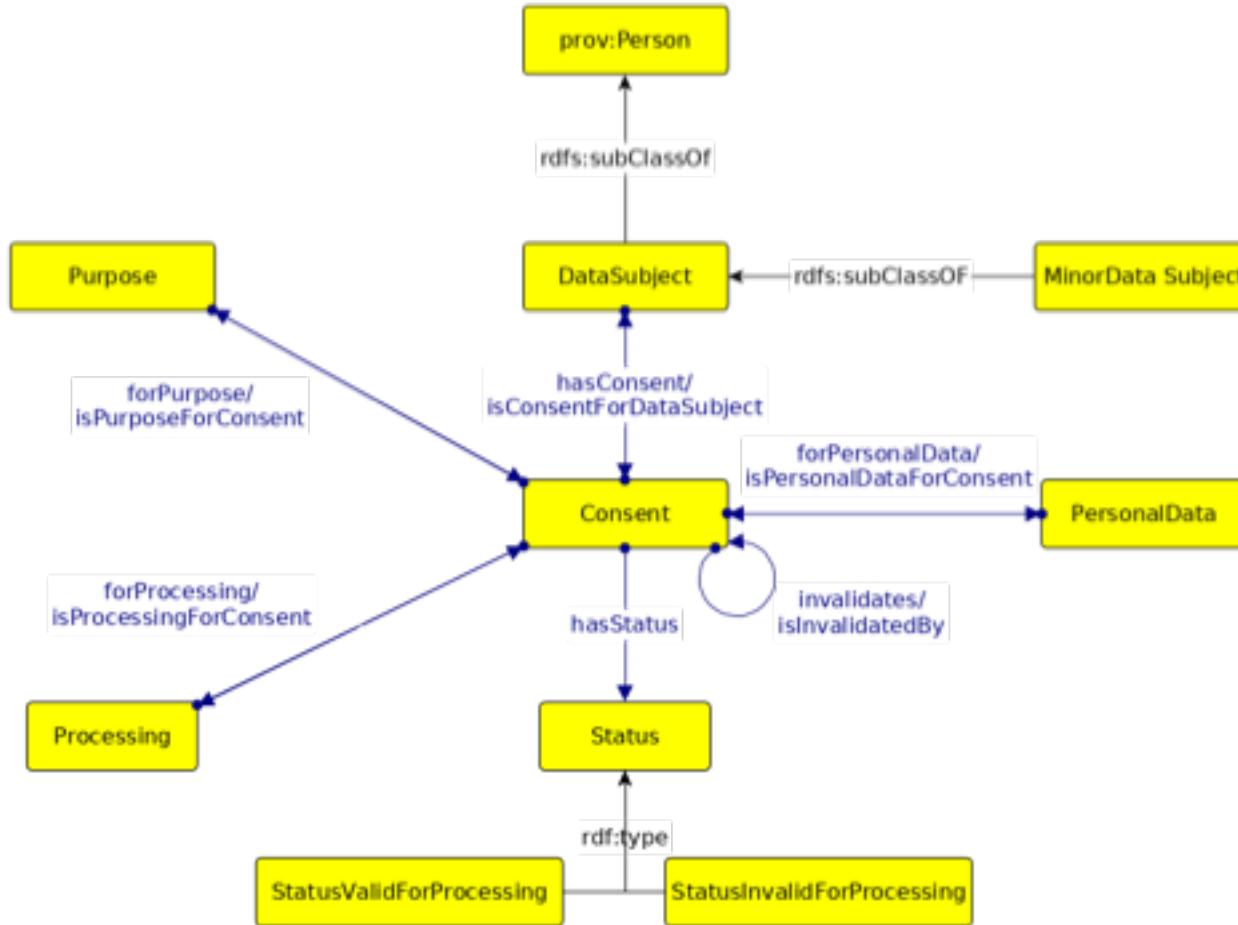
<https://openscience.adaptcentre.ie/>

- Data Storage duration (temporal properties)
  - Fixed timestamp e.g. 27-APR-2020
  - Duration e.g. 2 years
  - Indefinite e.g. “as long as required”
  - Dependent on event/condition e.g. “until you are logged in”
  - Proposal: hasStorageDuration (union of properties)
- Data Location (geospatial and political properties)
  - Country
  - Region (e.g. North America)
  - Political Unions (e.g. EU → Brexit changes memberships)
  - Multiple locations with dynamic changes e.g. servers
  - Proposal: hasDataLocation (union of properties)

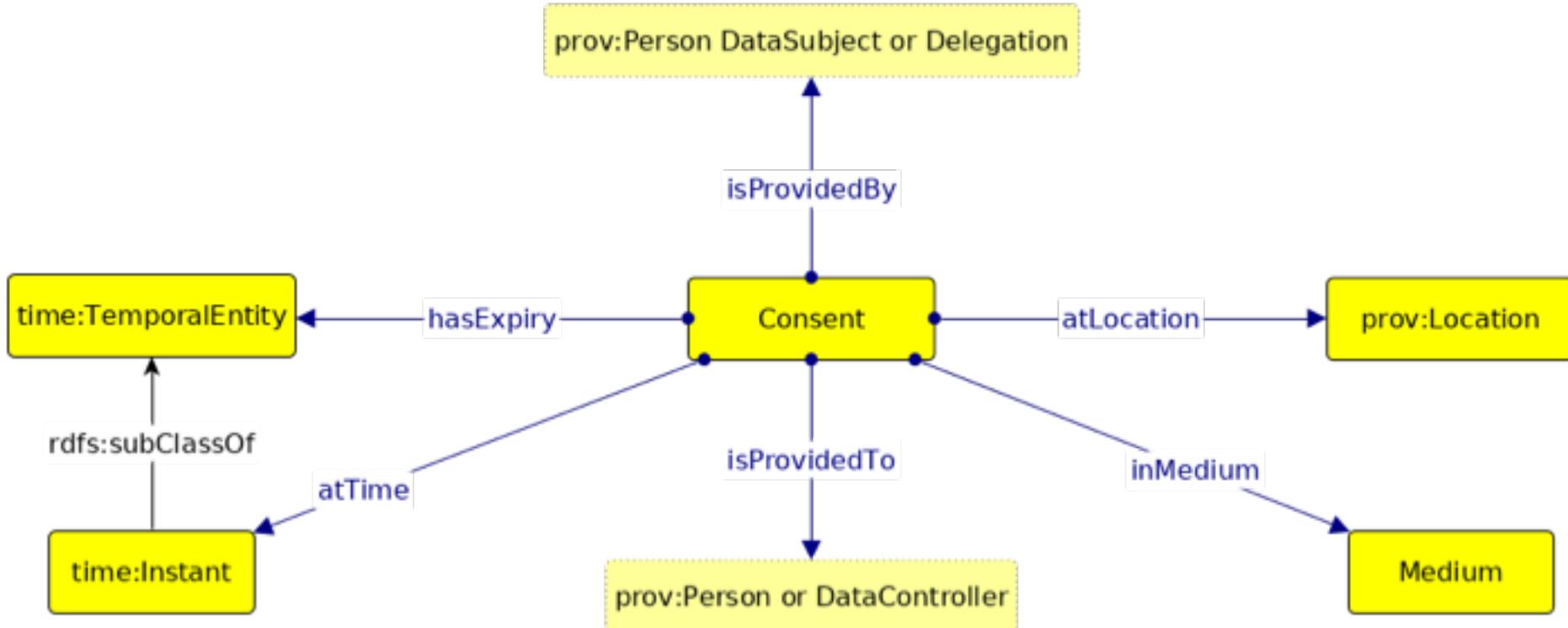
- Consent-specific ontology for GDPR
- Models purpose, processing, etc. focused on consent instance
- Companion to GDPRov
- Models context
  - Required to analyse GDPR consent requirements
  - E.g. Location, medium
- Models consent states
  - Based on consent management in real-world
  - Organisations need to track consent
  - E.g. Given, Requested, Invalidated
- Models chain of consent instances

# GConsent – Core Concepts

<https://openscience.adaptcentre.ie/>



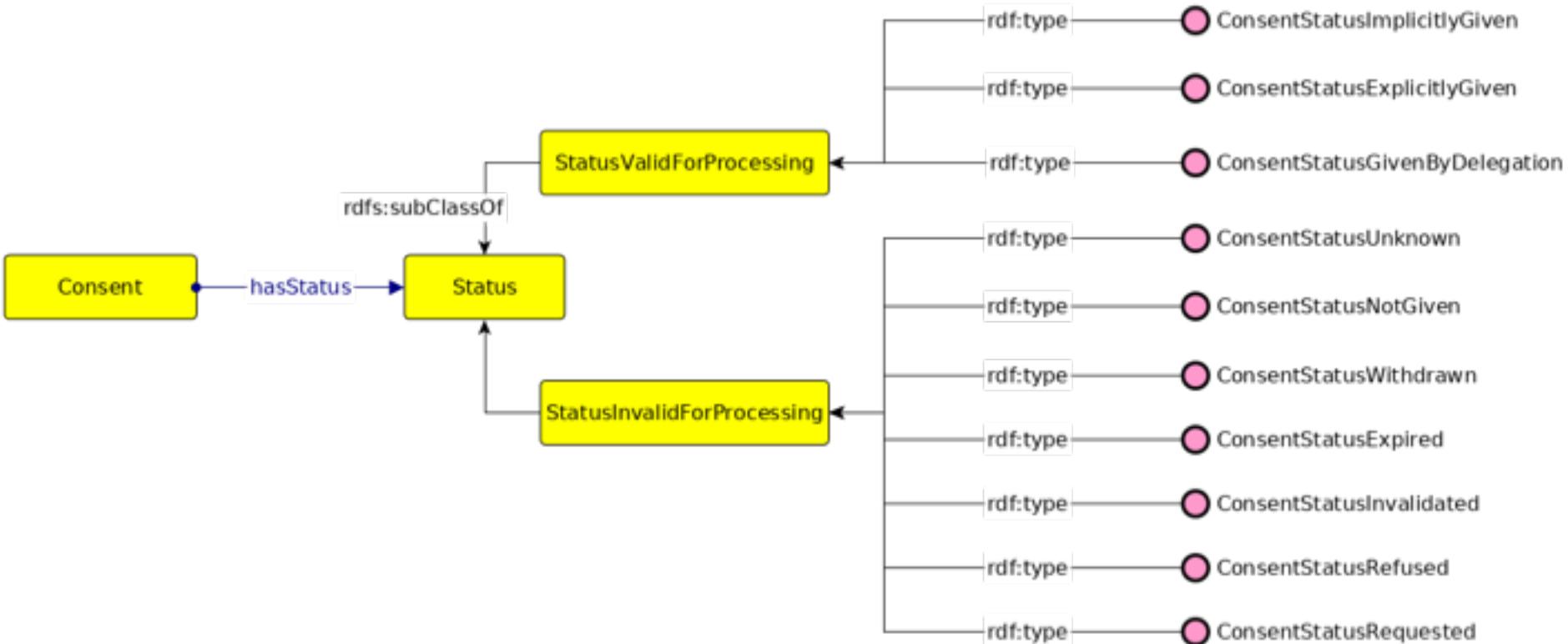
Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | [pandith@tcd.ie](mailto:pandith@tcd.ie) | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GConsent – Consent States

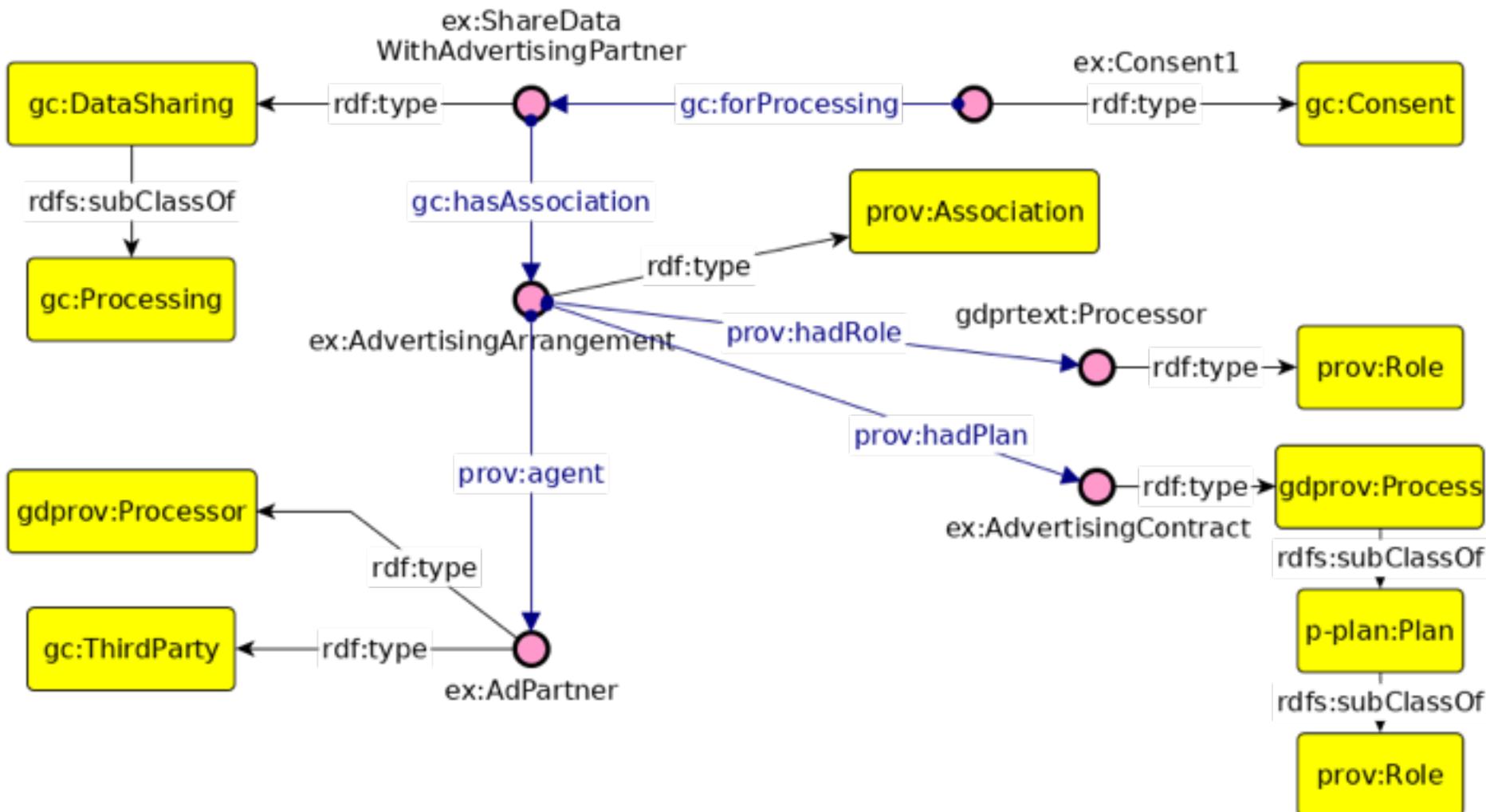
<https://openscience.adaptcentre.ie/>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GConsent – Example #3

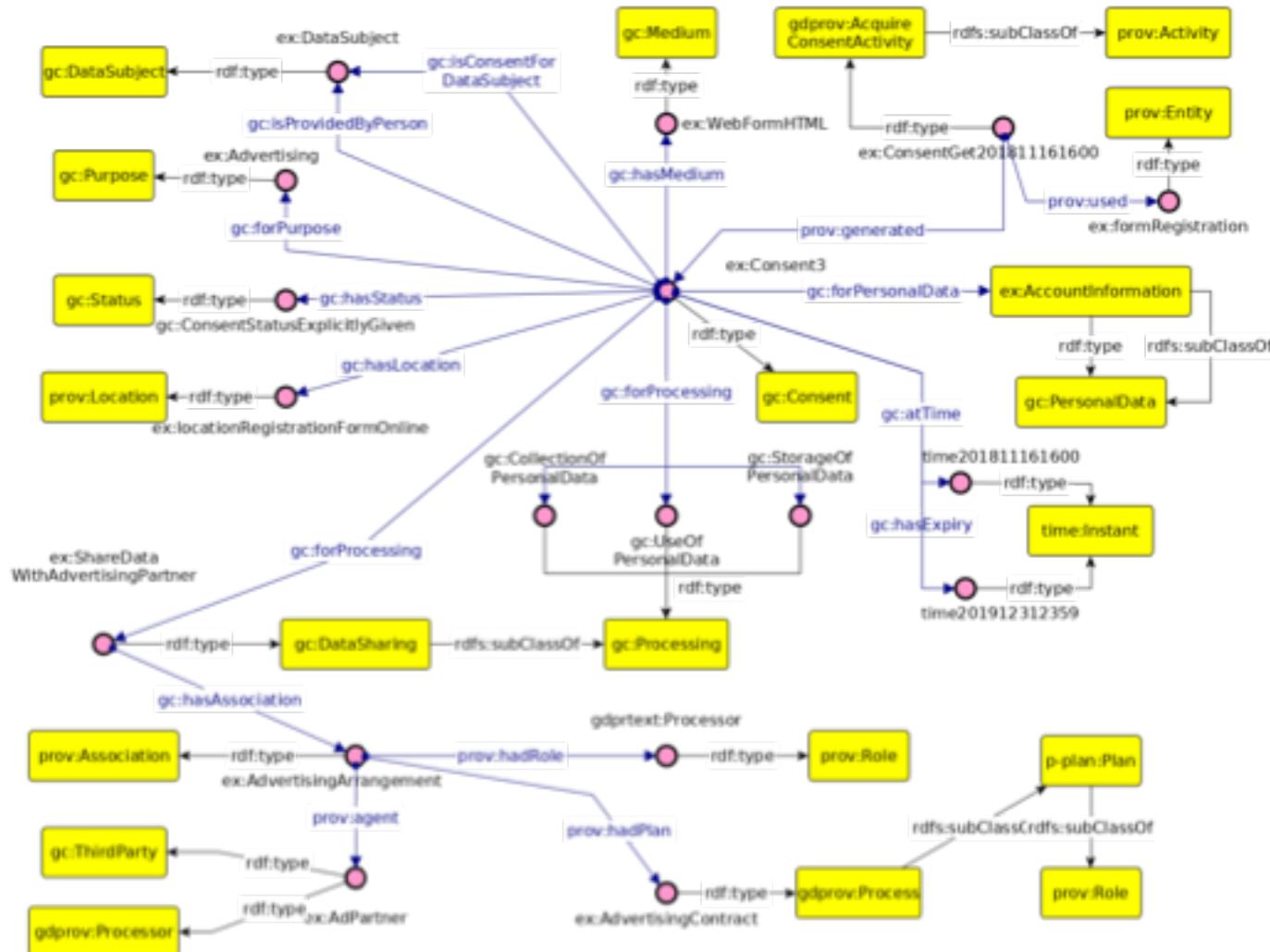
<https://openscience.adaptcentre.ie/>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# GConsent – GDProv

<https://opensource.adaptcentre.ie/>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

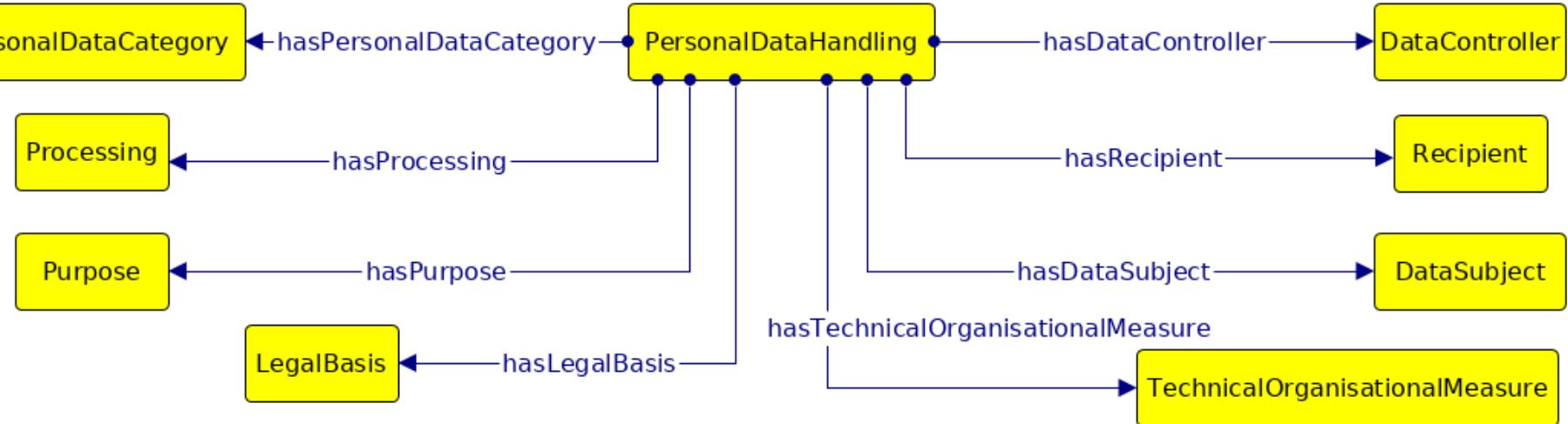
# GConsent – Compare with SotA

<https://openscience.adaptcentre.ie/>

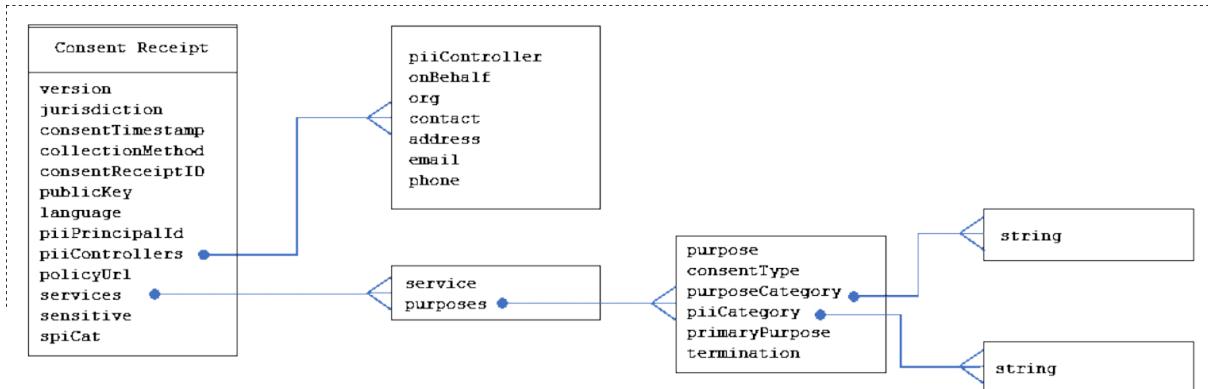
Table 5.7: Comparison of GConsent with SotA

Work	PD	Pu	Pr	Sh	St	Rp	S	W	D	SE	Ct	T
GConsent	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SPECIAL	✓	✓	✓	✓	✓	✓		✓				
SPL+CitySPIN	✓	✓	✓	✓	✓	✓		✓				
Lodge et al	✓	✓										
Peras	✓	✓	✓	✓	✓			✓				
Coletti et al	✓	✓						✓	✓			
AdvoCATE	✓	✓				✓	✓			✓	✓	
RestAssured	✓	✓	✓	✓	✓	✓	✓					
OPERANDO	✓	✓	✓	✓			✓					
PoSEID-on	✓						✓					
MHMD	✓											
DECODE	✓	✓			✓							
Consent Receipt	✓	✓								✓	✓	

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>



## Consent Receipt v1.1



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
 PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
 CC-BY-NC 4.0 | [pandith@tcd.ie](mailto:pandith@tcd.ie) | @coolharsh55 | <https://harshp.com/research/phd-thesis>

- GDPRtEXT
  - DPV does not provide GDPR glossary but provides related concepts
  - Does not model text of GDPR
  - Uses a more recent iteration of URL format (not formalized)
  - GDPRtEXT can be additional glossary for DPV
- GDPRov
  - DPV does not model provenance records (not in scope)
  - DPV does not have phases of compliance
  - GDPRov can add on DPV concepts (e.g. purpose → steps)
- GConsent
  - DPV models consent fields (based on GConsent + Consent Receipt)
  - Not all fields are modeled
  - Consent States not incorporated
  - DPV can utilize more fields from GConsent

# Querying with SPARQL

<https://openscience.adaptcentre.ie/>

- SotA does not provide SPARQL queries for reuse (as resource)
- SPARQL queries are highly use-case dependant
- Chosen use-case: GDPR Readiness Guide (2017) by Irish DPC
  - One of first such guides published, authoritative
  - In English, clear questions, meant for organisations
  - Also used to develop compliance questions for RO1 and RO2
- Method:
  - Identify information required by questions
  - Analyse questions based on information requirements
  - Interpret SPARQL queries using ontologies (GDPRov + GDPRtEXT)
  - Create synthetic use-case (test data)
  - Execute SPARQL queries
  - Evaluate a) extent of queries b) suitability of queries

# Querying with SPARQL - Queries

<https://openscience.adaptcentre.ie/>

- Questions categorized based on information dependency
  1. Demonstrative – retrieved information is sufficient
  2. Assistive – retrieved information needs to be evaluated
  3. Evaluative – more information needs to be retrieved
- Argument: SPARQL nested queries? Or compliance evaluation approach?
- Questions do not relate to ex-ante and ex-post
  - but implicitly refer to plans and processing activities
  - Categorise based on implicit information
- GDPRov evolved after this exercise
- Thesis also reflects capabilities of current iteration of GDPRov
- Validation use-case better choice for demonstration of queries
- Provides a way to link information to GDPR
  - SPARQL CONSTRUCT queries an alternative
  - But does not permit validation of information

# Querying with SPARQL - Evaluation

<https://openscience.adaptcentre.ie/>

- 32 of 63 questions represented as SPARQL queries
- 31 of 63 not implemented: 20 not in scope, 3 lacking concepts in GDPRov
  - G8: data retention/storage periods/duration
  - S5, S7: technical measures
- Exercise undertook before GConsent was developed
- Since this guide is based on organization's plans, it was not repeated with GConsent
- Online demo  
(alt implementation using RDF-JS, Communica possible to permit anyone to upload data graphs and query)

Legal basis for each processing purpose (non-special categories of personal data)

For each purpose that personal data is processed, list the legal basis on which it is based e.g. consent, contract, legal obligation (Article 6).

Q5. legal basis for processing

```
SELECT DISTINCT ?process ?legal WHERE {
  ?data a ?data_type .
  ?data_type rdfs:subClassOf* ?process .
  ?process a ?process_type .
  ?process_type rdfs:subClassOf* ?legal .
  ?process gpo:isUsedData ?data .
  ?data gpo:isPartOfProcess ?process .
  OPTIONAL { ?process gpo:hasLegalBasis ?legal } .
  OPTIONAL { ?process gpo:hasLegalBasis ?legal } .
}
ORDER BY ?process
```

process	legal
1 this.HandleRightDataPortability	
2 this.HandleSAR	
3 this.NewUserSignUpProcess	gdpr:ext:GivenConsent
4 this.OrderProcess	gdpr:ext:LegitimateInterest
5 this.RemoveUserAccountProcess	

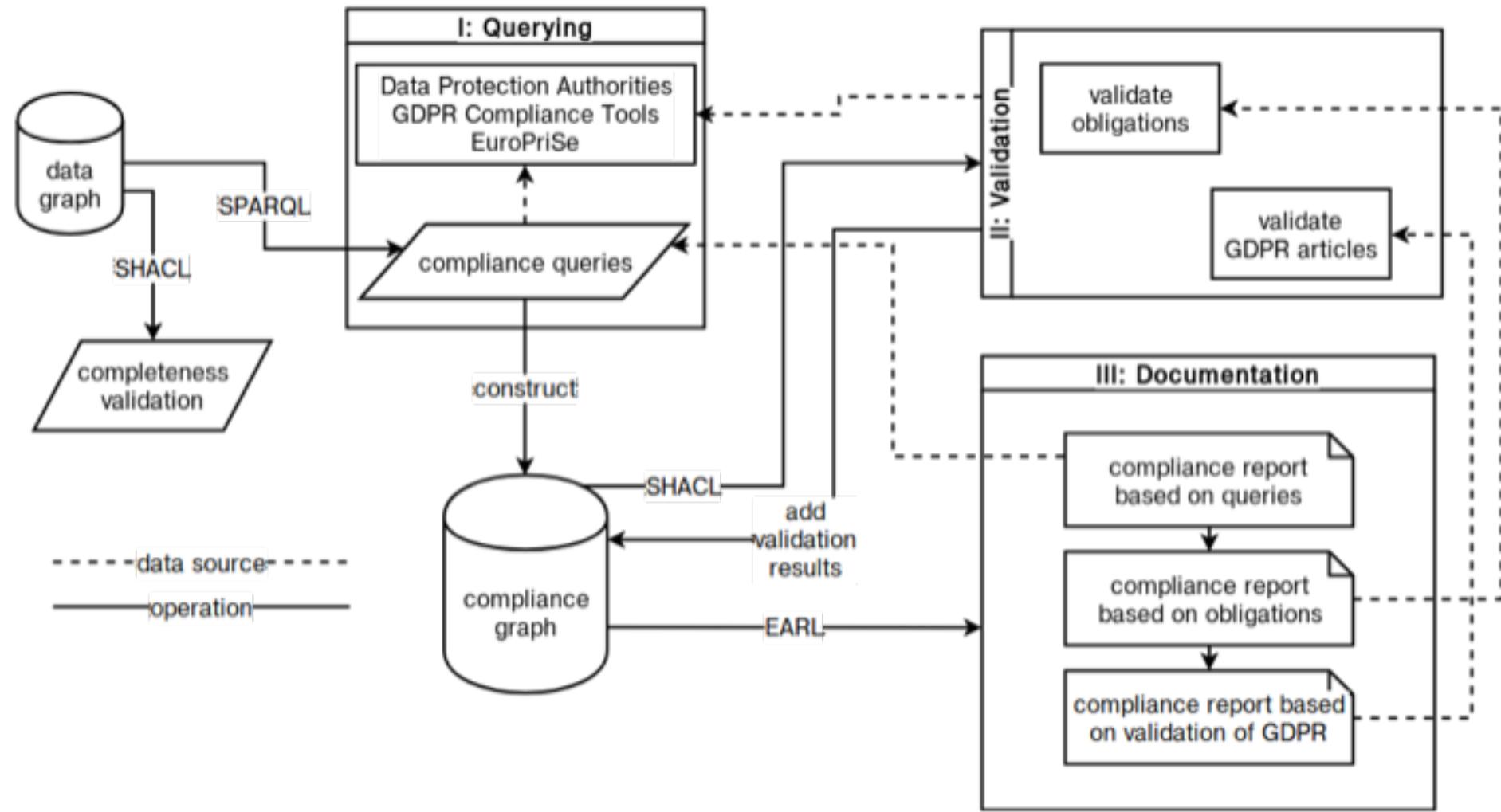
- GDPR Compliance: info satisfies obligations and requirements
- Ensuring correct information i.e. validation for a) correctness b) completeness
- SotA does compliance evaluation, PhD work is about this validation
- If approach is using RDF data, SHACL is nice because:
  1. Persists test results as RDF data
  2. Flexible – build own tests and add components

## PhD → Validation Model

1. Utilises SHACL to validate information for GDPR compliance.
2. Expresses compliance as a test-driven exercise similar to the concept of unit-testing in software engineering.
3. Utilises results of testing ex-ante information for testing of ex-post information in order to reduce the number of tests required.
4. Constructs a compliance graph by storing validation results based on concept of knowledge-graph.
5. Demonstrates use of compliance graph in retrieving and documenting information regarding GDPR compliance

# Validation Model – Model Architecture

<https://openscience.adaptcentre.ie/>



Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# Validation Model – linking to GDPR

<https://openscience.adaptcentre.ie/>

- Associate test with GDPR to specify validation of related information
- Add property using GDPRtEXT
- Enables annotation of compliance evaluation and testing back to GDPR
- Permits querying of tests/information based on GDPR links
- To extend for other laws, create a generic property referring to legal clauses

```
1 :Constraint rdfs:subClassOf sh:NodeShape ;
  rdfs:label "Constraint" .
2 :AutomaticallyCheckedConstraint rdfs:subClassOf :Constraint, sh:NodeShape ;
  rdfs:label "Automatically Checked Constraint" .
3 :ManuallyCheckedConstraint rdfs:subClassOf :Constraint, sh:NodeShape ;
  rdfs:label "Manually Checked Constraint" .
4
5
6
7
8 :linkToGDPR a rdfs:Property ;
  rdfs:range eli:LegalResourceSubdivision ;
  rdfs:label "linkToGDPR" .
```

# Validation Model – Ex-ante → Ex-post

<https://openscience.adaptcentre.ie/>

- GDPR requires organisations to plan activities i.e. ex-ante plans
- GDPR requires organisations to ensure plans compliance i.e. ex-ante compliance
- GDPR requires organisations to ensure processing activities are compliant i.e. ex-post
- However, ex-post processing activities are based on ex-ante plans
- Reuse ex-ante test data
  - if common elements are already tested in ex-ante
  - only unique elements need testing again in ex-post
- E.g. Consent Dialogue
- Shown to all users of websites i.e. ex-ante information
- Given consent unique to individual i.e. ex-post information
- To test consent as per GDPR
  - Test consent dialogue was valid (common for all users)
  - Test given consent was valid (specific to individual)
- Test common information once, persist test results as data
- Now only single check to ensure common tests were passed
- Therefore, test complexity is reduced in ex-post testing
- (common sense approach – used commonly!)

# Validation Model – Quantcast

<https://openscience.adaptcentre.ie/>

- Quantcast is the largest consent dialogue provider on the internet
- Part of the IAB framework
- Claims high number of data points available for analytics as a service
- Consent Dialogue → RDF data using GDPRov + GConsent + GDPRtEXT
- Develop SHACL constraints based on compliance questions to validate data
  - Complexity: basic, low, simple
  - Does not do compliance evaluation, only ensures information is correct & complete
- Consent dialogue tested in Ex-ante phase, results saved
- Given consent tested in Ex-post phase, also tests Ex-ante saved results
- More efficient tests in this process
  - Without splitting across phases: 59 tests for each given consent
  - Ex-ante tests: 57
  - Ex-post tests: 3 (2+1 for ex-ante saved results)
  - After splitting across phases: 3 tests for each given consent

# Validation Model – Quantcast SPARQL queries

<https://openscience.adaptcentre.ie/>

Table 6.3: SHACL validation report linked to GDPR

Name	Type	GDPR	Result	Node
Consent ≠ Inactivity	M	R32	P	
Consent ≠ Pre-ticked Boxes	M	R32	P	
Consent ≠ Silence	M	R32	P	
Consent → Data Subject	A	A4-11	P	
Consent → Given To	A		P	
Consent → Location	A		P	
Consent → Medium	A	A7-2	P	
Consent → Personal Data	A	A4-11,R32	P	
Consent → Processing	A	A4-11,R32	P	
Consent → Provided By	A	A7-2	P	
Consent → Purpose	A	R32,R42	P	
Consent → Status	A		P	
Consent → Timestamp	A		F	Q:Consent20190415120753
Consent → Timestamp	A		F	Q:Consent20190415140000
Consent ≡ Choice	M		P	
Consent ≡ Freely Given	M	A4-11	P	
Consent ≡ Specific	M	A4-11	P	
Consent ≡ Statement of Clear Action	M	A4-11	P	
Consent ≡ Unambiguous	M	A4-11	P	
Consent Generating Activity	A		P	

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# Validation Model – Compare with SotA

<https://openscience.adaptcentre.ie/>

Table 6.4: Comparison of SHACL validation with SotA

Approach	Evaluation method	Scope	Machine-readable result?	Provides remedies?	Links results to GDPR?
Pandit	SHACL	RDF data	✓	✓	✓
SPECIAL	OWL	Consent	✓		
SPL+SERAMIS	ODRL	Obligations	✓	✓	✓
SPL+Vos et al.	OWL, ASP	Obligations	✓	✓	
SPL+CitySPIN	OWL	Consent	✓		
MIREL	RuleML	Obligations	✓	✓	✓
MRL+DAPRECO	RuleML	Obligations	✓	✓	✓
BPR4GDPR	OWL	Process Flows		✓	
Lodge et al	SDK	Process Flows	✓	✓	
Tom et al	BPMN	Process Flows	✓	✓	
Corrales et al	Questionnaire	Obligations			
LUCE	Smart Contracts	Data Sharing	✓		
AdvoCATE	Smart Contracts	Consent	✓		
Sion et al	UML, DFD	Process Flows	✓	✓	
privacyTracker	Access Control	Data Sharing	✓		
Robol et al	STS	Process Flows	✓		
GuideMe	Questionnaire	Process Flows		✓	
Basin et al	Algorithm	Process Flows			
RestAssured	XACML	Process Flows	✓		
DEFeND	Questionnaire	Obligations	✓		
OPERANDO	Access Control	Process Flows	✓		
PoSEID-on	Smart Contracts	Data Sharing	✓		
DECODE	Smart Contracts	Consent	✓		

# Ontology Engineering Methodology

<https://openscience.adaptcentre.ie/>

- Ontology quality - Design, Semantics; OOPS!, Use-cases
- Documentation – WIDOCO
- Dissemination – FAIR principles
  - W3ID – persistent identifier; Linked Open Vocabularies (LOV); GitHub, Zenodo
- Evaluation – competency question, application to use-cases
- NeOn: identify requirements and use-cases
- UPON Lite: iteratively develop ontological representation

Table 1.1: Summary of Evaluation Methods

Method	GDPRtEXT Ontology	GDPRov Ontology	GConsent Ontology	Querying using SPARQL	Validation using SHACL
Fulfilment of Competency Questions	✓	✓	✓	N/A	N/A
Semantic reasoner logical consistency	✓	✓	✓	✓	✓
OOPS! common pitfalls detection	✓	✓	✓	N/A	N/A
Documentation metadata and quality	✓	✓	✓	N/A	N/A
Demonstrate application to use-case	✓	✓	✓	✓	✓
External use-case	✗	✓	✓	✓	✓
Comparison with SotA	✓	✓	✓	✓	✓
Analysis of citations	✓	✓	N/A	✓	N/A
Dissemination of work (for providing transparency)					
Peer-reviewed publication	✓	✓	✓	✓	✓
Reproducibility (open access resources)	✓	✓	✓	✓	✓

Representing Activities associated with Processing of Personal Data and Consent using Semantic Web for GDPR Compliance  
PhD Viva / Thesis Defense of Harshvardhan J. Pandit, PhD Candidate, SCSS, Trinity College Dublin  
CC-BY-NC 4.0 | pandith@tcd.ie | @coolharsh55 | <https://harshp.com/research/phd-thesis>

# Timeline of Research

<https://openscience.adaptcentre.ie/>

- 2016: Started work in March-September 2016
  - Relevant work existing: preliminary work in DAPRECO
- 2017:
  - Published GDPRov by September 2017
  - Preliminary work on consent ontology September 2017
- 2018:
  - Published GDPRtEXT in January 2018
  - SPECIAL published work on GDPR compliance in June 2018
  - MIREL+DAPRECO published PrOnto in July 2018
  - Adopt GDPR readiness guide for SPARQL queries in August 2018
  - Developed GConsent by September 2018
- 2019
  - H2020 projects on GDPR publish deliverables and papers in January 2019
  - Developed SHACL validation model by March 2019
  - Data Privacy Vocabulary is published in June 2019
- SotA analysis in August-September 2019
- Thesis draft in December 2019, Submitted in February 2020

- ODRL to express GDPR obligations
  - Information needed for resolving activities → provenance
  - Information needed for expanding GDPR activities
  - How to evaluate?
  - SotA: Agarwal et al. and Vos et al. have used ODRL profiles
- BPMN
  - Business process standard for representation
  - Own set of tooling, querying, validating
  - How to operate while being interoperable?
  - How to specify GDPR concepts and activities?
- Major factors:
  - Extending vocabulary with GDPR concepts
  - Querying for information
  - Validation / Tests
  - Linking of information with GDPR