Engaging Content
Engaging People

A World Leading SFI Research Centre

Science Foundation Ireland SFI For what's next

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

PhD Viva

Harshvardhan J. Pandit

HOST INSTITUTION

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

HOST INSTITUTION

DCU
Ollscoil Chathair Bhaile Átha Cliath
Dublin City University

PARTNER INSTITUTIONS

TU Dublin
UCD Dublin
CIT Cork Institute of Technology
Institiúid Teicneolaíochta Chorcaí
AIT
Maynooth University
National University of Ireland Maynooth

FUNDED BY:

Ireland's European Structural and Investment Funds Programmes 2014-2020
Co-funded by the Irish Government and the European Union

European Union
European Regional Development Fund

Science Foundation Ireland SFI For what's next

Background & Motivation

- Technological solutions for compliance with GDPR [3] face challenges regarding [12-18]:
  - Algorithmic interpretation of legal compliance requirements
  - Formal investigations require information associated with clauses
  - Interoperability between stakeholders
- GDPR compliance is associated with:
  - Processing of personal data
  - Legal Bases (in particular consent)
  - Data Governance (e.g. Controller – Processor)
  - Rights Management

Background & Motivation

- Business Process Management is an existing field
  - Represents and works with processes / activities (e.g. BPMN[1]) in Ex-ante (planning), real-time (during), and Ex-post (verification) stages
  - Legal Compliance is a secondary operation on top of these [12]
  - GDPR imposes additional requirements on existing information management practices
- Semantic Web technologies have proved useful in legal compliance [15,16] and for addressing GDPR
  - Commercial solutions e.g. Top Quadrant, Thomas Reuters, Signatu
  - H2020 projects such as SPECIAL [26], MIREL [27], BPR4GDPR [29]
  - Use is growing in legal domain for information representation, querying, reasoning, and interoperability [25]

1 https://www.omg.org/spec/BPMN/2.0/

Research Question

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

---

1)Representation of information

2)Querying

3)Validation

4)Assessment/Evaluation

A)Information required in order to conduct evaluation of compliance

B)Information about compliance evaluated

i. State of system at a given time
ii. Stakeholders

State of the Art

Engaging Content
Engaging People

Table 3.1 (cropped)

Work	Type	Clause	Onto	Ex-ante	Ex-post	Activities	Consent	Compliance	Req.	OA
SPECIAL	PRJ		✓	✓	✓	✓	✓	✓		✓
SERAMIS	PRJ	✓	✓					✓	✓	
Vos et al	RES		✓					✓	✓	✓
CitySPIN	PRJ		✓	✓	✓	✓	✓	✓		✓
MIREL	PRJ	✓	✓	✓	✓	✓		✓	✓	
DAPRECO	PRJ	✓	✓	✓	✓					
BPR4GDPR	PRJ			✓	✓					
Elluri et al.	RES		✓							
Ujcich et al.	RES		✓		✓					
PICS	RES				✓					
AdvoCATE	RES		✓							
Geko & Tjoa	RES		✓							
LPL	RES		✓							
Lodge et al	RES		✓		✓					
Peras	RES		✓							
Tom et al	RES		✓	✓						
Coletti et al	RES		✓	✓						
Corrales et al	RES							✓	✓	
LUCE	RES					✓		✓		

- 29 approaches
- 9 Research Projects (EU/H2020/National)
- 3 model GDPR clauses
- 20 model some form of ontological representation
- 13 model activities associated with GDPR
- 12 represent Ex-ante, 10 represent Ex-post
- 9 model consent
- 18 evaluate compliance
- 9 provide remedies / suggestions
- 4 are completely open-access

SotA – Gaps & Opportunities (Sec. 3.8)

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 and ex-post phases
4) Representation of consent information
5) Demonstration using authoritative compliance queries
6) Validating information for compliance evaluation

Research Objectives

RO1 Identify the **subset of GDPR** relevant for activities associated with processing of personal data and consent regarding compliance

major -> ***RO2*** **Identify information** required to represent activities associated with processing of personal data and consent in investigation of GDPR compliance.

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

(a): concepts and text of GDPR

(b): activities associated with processing of personal data and consent

(c): consent required to determine compliance

minor -> ***RO4*** Represent compliance questions using **SPARQL to query information** about activities associated with processing of personal data and consent

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

Methodology

Engaging Content
Engaging People

RO1 RO2 **GDPR Analysis (Sec. 4.2.1)**

1) Study authoritative sources (official text, Article 29 Working Party and European Data Protection Board opinions, court cases, legal experts)
2) Identification of requirements and stakeholders – identify actors and information flows for data governance
3) Develop ‘compliance questions’ to retrieve and assess information for GDPR compliance

RO3 **Ontology Development (Sec. 5.1)**

1) Use compliance questions as Competency Questions [35] and develop in iterative manner using NeOn [34] and UPON Lite [36]
2) Ontology quality [37,38] and documentation [38] best practices
3) Disseminate (Open-Access, advertise, publish)
3) Evaluation on above criteria

RO5 **Validation Framework (Sec. 6.2.1)**

1) Identify role of information validation and its persistence for compliance
2) Develop framework to validate information and persist results as KG
3) Develop validation constraints from compliance questions
4) Implement validation constraints using SHACL
5) Demonstrate using real-world example/use-case
6) Evaluate

RO4 **Querying (Sec 6.1.2)**

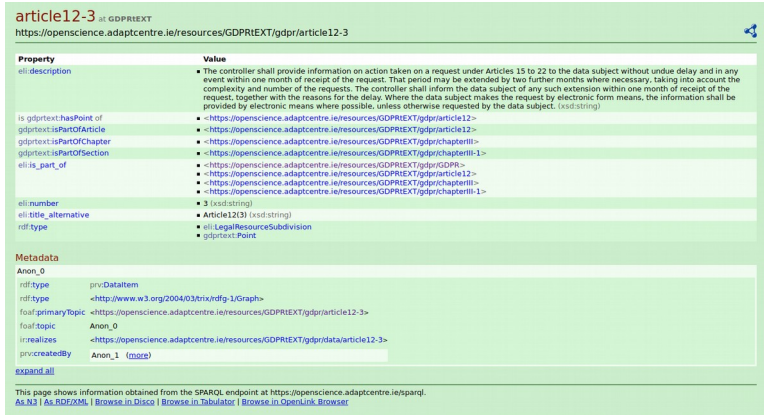
1) Represent SPARQL queries using developed ontologies
2) Demonstrate using real-world questions/requirements
3) Evaluate

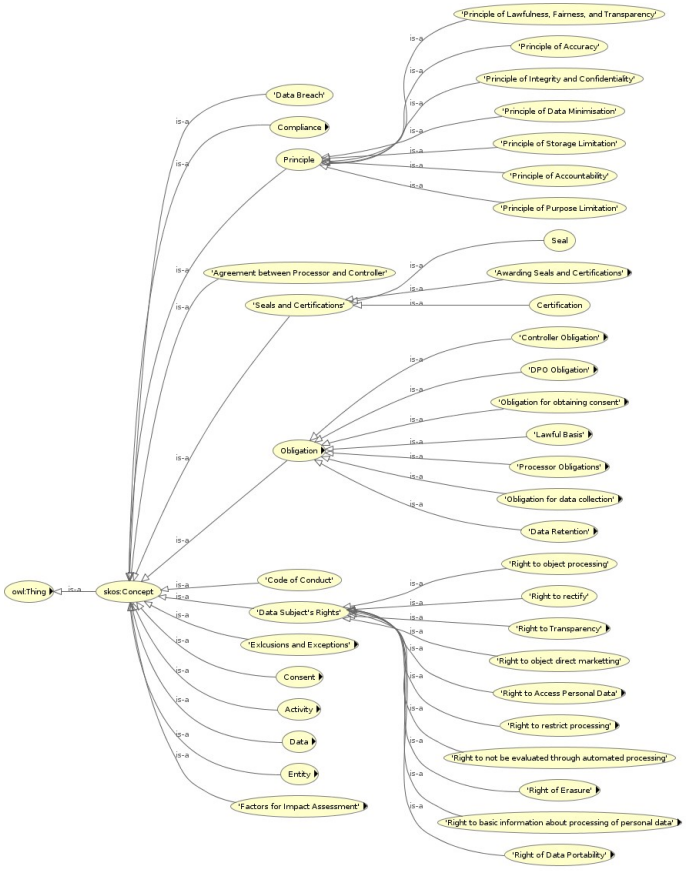
Representing Information about GDPR – GDPRtEXT (Sec. 5.2)

Engaging Content
Engaging People

Structure of GDPR

- Chapters, Articles, Clauses – and their relationship
- Extends European Legislation Identifier (ELI) [39] (ELI is maintained by EU Publications Office)
- Assign IRIs for individual clause for granularity
- Export as RDF dataset





Glossary of GDPR concepts (related to compliance)

- Principles, Rights, Data, Activities, Entities, Obligations
- Relevant clauses: definitions, relations
- Represent using SKOS

Dissemination & Publication

- Ontology+documentation: https://w3id.org/GDPRtEXT/
- ESWC 2018 Resource Track [72]
- Added as dataset on Open Data Portal (Ireland)
- Publications Office expressed interest on work

Representing Information about GDPR – GDPRtEXT (Sec. 5.2)

Comparison with SotA (Table 5.2)

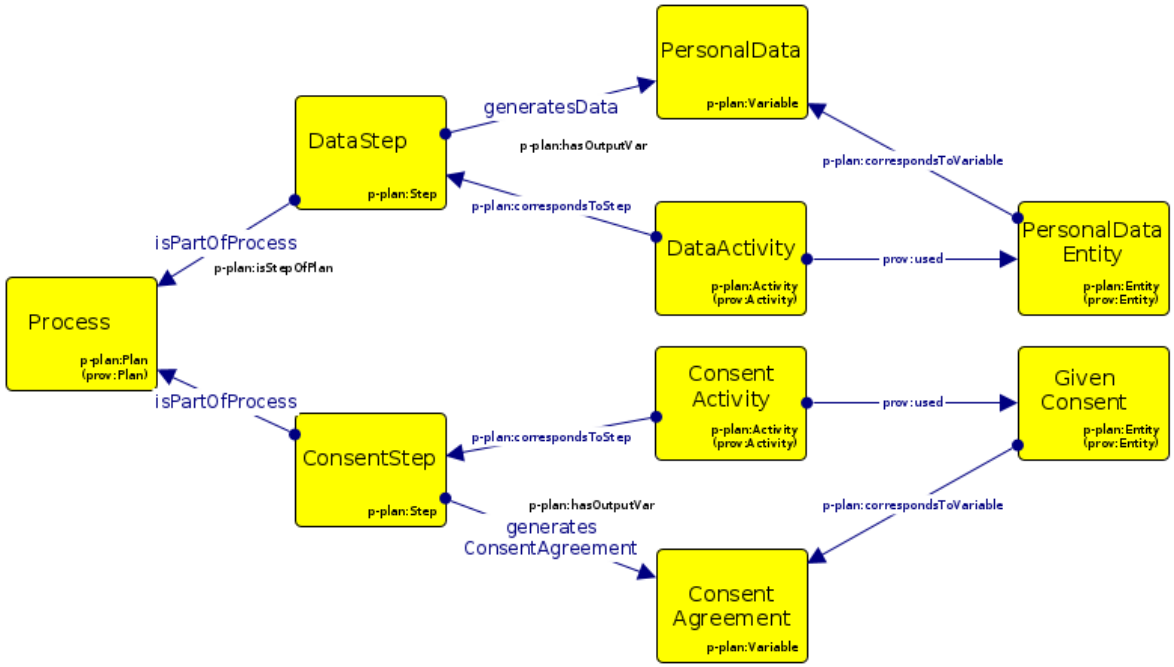
Work	**GDPRtEXT**	ELI	ELI+	Agarwal et al	PrOnto
Vocabulary	ELI	OWL2	OWL2	RDFS	Akoma Ntoso
Granularity	Sub-Paragraph	Legislation	Sub-Paragraph	Paragraph	Sub-Paragraph
Glossary	✓	✗	✓	✗	✗
PID	✓	✓	✓	✗	✗
OA	✓	✓	✓	✗	✗
GDPR text	✓	✗	✓	✗	✓

1) Based on ELI – authoritative representation of EU legislations
2) Glossary of concepts
3) Granularity can refer to clauses
4) Uses best practices (as a semantic web resource)
5) Is Open Access under permissive license

Note: EU Publications Office has indicated their plans to update ELI (indicated as ELI+ in the table) which will incorporate the above features and some additional ones such as markup for specifying concepts of interest. It has also indicated the intention to align ELI with Akoma Ntoso to provide a unified vocabulary.

Representing Information about Activities associated with Personal Data and Consent – GDPRov (Sec. 5.3)

- **GDPRov** ontology for representing provenance information based on GDPR requirements
- PROV-O (W3C standard for provenance) and P-Plan (extends PROV-O to represent Plans as Scientific Workflows)
- Uses GDPRtEXT to indicate source of concepts



Models:

- Actors and Agents involved
- Details of processing activities
- Lifecycle of processing activities
- Consent activities and their similarity to data activities
- Similarly activities: rights, data breach reporting
- Ex-ante (plans) and Ex-post (verification)

Dissemination & Publication

- Ontology+documentation: http://w3id.org/GDPRov
- ISWC 2017 PrivOn workshop [66]

Representing Information about Activities associated with Personal Data and Consent – GDPRov (Sec. 5.3)

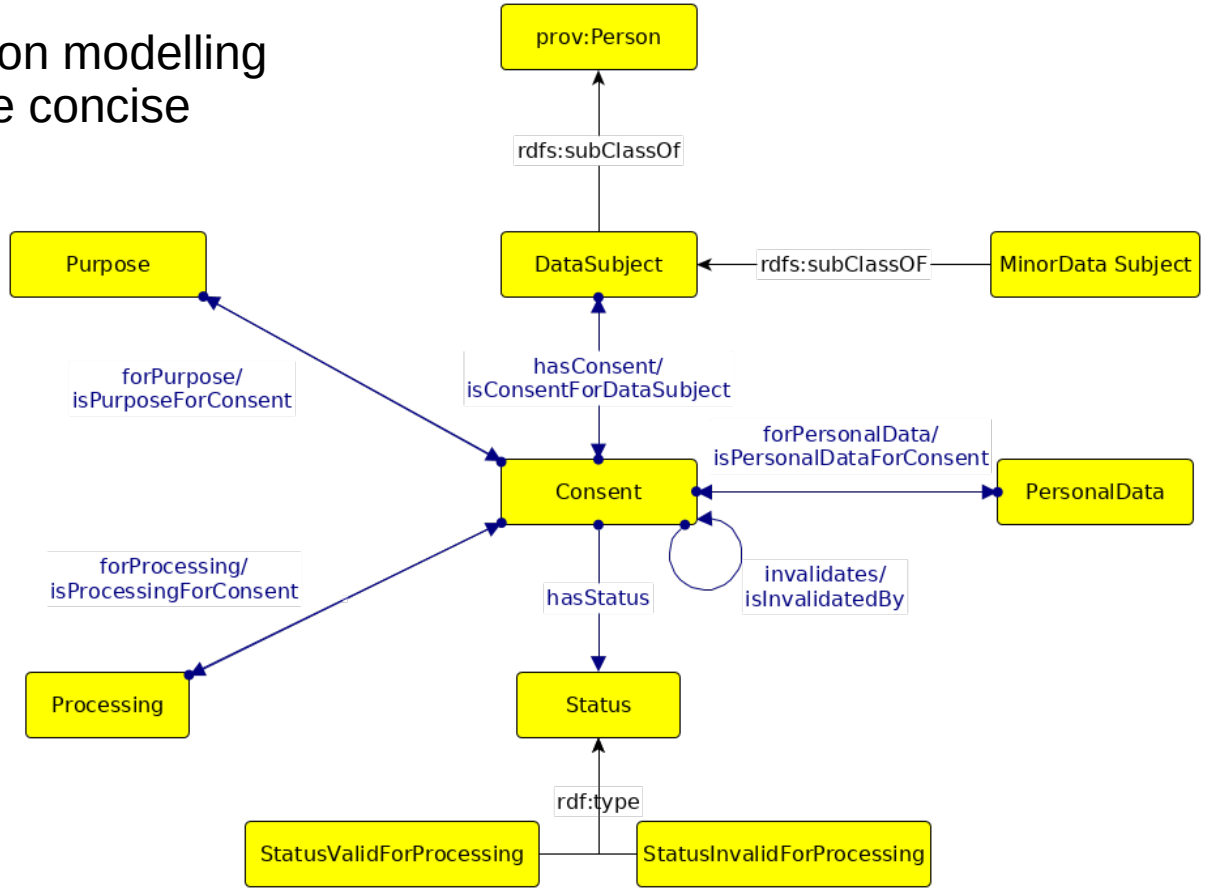
Comparison with SotA (Table 5.4)

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				✓	✓	✓	✓	✓		

1) Separation of provenance and plans using P-Plan
2) Open Access under permissive license
3) Larger scope of concepts
4) Models activities associated with GDPR
5) Associates legal basis with activities
6) Uniform modelling of activities associated with GDPR
7) Permits capturing state of system to indicate compliance at time

Representing Information about Consent - GConsent (Sec. 5.4)

- **GConsent** ontology for representing information about consent based on GDPR requirements
- GDPRov models consent as a provenance-related entity and represents its lifecycle
- GConsent presents alternate/companion modelling of information about consent with more concise association between concepts



Models:

- Context (e.g. location, medium)
- Provenance chaining of consent (i.e. evolution of consent)
- Consent States (e.g. given, refused, requested, invalidated, withdrawn)

Dissemination & Publication:

- Ontology+documentation http://w3id.org/GConsent
- ESWC 2019 [71]

Representing Information about Consent - GConsent (Sec. 5.4)

1) Larger coverage of GDPR requirements for consent
2) Models context
3) Consent states
4) Delegation
5) Significant effects of Processing
6) Relation between consent instances (lifecycle)
7) Open Access under permissive license

Comparison with SotA (Table 5.7)

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	✓	✓									✓	✓

Querying information using SPARQL (Sec. 6.1)

- Two types of questions associated with compliance:
  1) Retrieve relevant information e.g. what purposes do you use personal data for?
  2) Assess compliance e.g. does every purpose have a legal basis?
- Goal: represent questions as SPARQL queries
- Use-case: GDPR readiness guide published by Data Protection Commission (Ireland)
- Model information using GDPRtEXT, GDPRov
- https://w3id.org/GDPRep/checklist-demo

- Demonstrate use of SPARQL in retrieving information relevant to GDPR compliance
- Application on authoritative questions
- Show dependency on underlying data representations (i.e. ontologies used)
- Published SEMANTiCS 2018 [57]

Elements of personal data included within each data category

List each type of personal data included within each category of personal data e.g. name, address, banking details, purchasing history, online browsing history, video and images.

G2. Types of Personal Data

```
SELECT DISTINCT ?data ?type where {
  ?data a ?type .
  ?type rdfs:subClassOf gdprov:PersonalData .
  FILTER(regex(str(?data), "http://example.com/ontology/shoppingapp#")) .
} ORDER BY ?data ?type
```

RAW RESPONSE | TABLE | PIVOT TABLE | GOOGLE CHART | Search: | Show 50 entries

	data	type
1	this:AnonymisedUserProfile	gdprov:AnonymisedData
2	this:CustomerAddress	this:CustomerInfo
3	this:CustomerBankAC	gdprov:SensitiveData
4	this:CustomerCardDetails	gdprov:SensitiveData
5	this:CustomerContactNo	this:CustomerInfo
6	this:CustomerEmail	this:CustomerInfo
7	this:CustomerName	this:CustomerInfo

Showing 1 to 7 of 7 entries

Source of the personal data

List the source(s) of the personal data e.g. collected directly from individuals; from third parties (if third party identify the data controller as this information will be necessary to meet obligations under Article 14).

G3. data sources

```
SELECT DISTINCT ?data ?step ?agent ?agent_type where {
  ?data a ?data_type .
  ?data_type rdfs:subClassOf gdprov:PersonalData .
  ?step a gdprov:DataCollectionStep .
  ?step gdprov:collectsData ?data .
  ?step gdprov:collectsDataFromAgent ?agent .
  ?agent a ?agent_type .
  FILTER(regex(str(?agent_type), "http://example.com/ontology/shoppingapp#")) .
} ORDER BY ?data ?step ?agent
```

RAW RESPONSE | TABLE | PIVOT TABLE | GOOGLE CHART | Search: | Show 50 entries

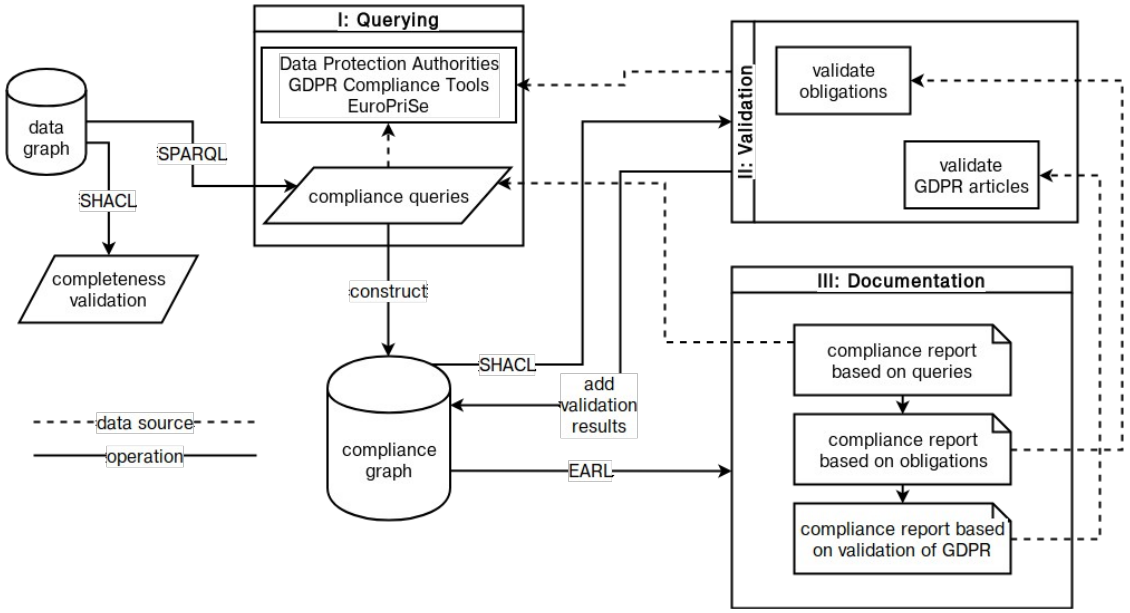
Validating information using SHACL (Sec. 6.2)

- Organisations required to document and maintain information associated with compliance
- Concept of ‘Compliance graph’ distinct from ‘data graph’
- Information validation as precursor to compliance evaluation
- Create validation ‘constraints’ from compliance questions
- Use SHACL to i) validate information ii) persist results iii) link with GDPR

```
:Constraint rdfs:subClassOf sh:NodeShape ;
    rdfs:label "Constraint" .
:AutomaticallyCheckedConstraint rdfs:subClassOf :Constraint, sh:NodeShape ;
    rdfs:label "Automatically Checked Constraint" .
:ManuallyCheckedConstraint rdfs:subClassOf :Constraint, sh:NodeShape ;
    rdfs:label "Manually Checked Constraint" .

:linkToGDPR a rdfs:Property ;
    rdfs:range eli:LegalResourceSubdivision ;
    rdfs:label "linkToGDPR" .
```

- Tests need to be carried out in ex-ante and ex-post phase
- Use ex-ante test results to reduce number of ex-post tests (i.e. abstract common constraints)



Validating information using SHACL (Sec. 6.2)

- Use-case: consent dialogue on Quantcast.com
- ‘data graph’: GDPRtEXT, GDPRov, and Gconsent
- Validate results and save into ‘compliance graph’
- Use SPARQL to ‘query’ and generate ‘dashboards’
- Published in SEMANTiCS 2018 [59], 2019 [60], ISWC 2018 CKG Workshop [58]
- Demonstrated to DPC – Ireland

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:CATTPInfoStorageAccess
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	

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	✓		

Comparison with SotA (Table 6.4)

- Use of SHACL
- Separation of ex-ante and ex-post phases
- Reuse results
- Link to GDPR
- Indicate violating constraints/nodes
- Compliance results persisted
- SPARQL queries as ‘dashboard’
- Open-Access implementation https://w3id.org/GDPRep/semantic-tests

Fulfilment of Research Objectives (Sec. 7.1)

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) presented in Chapter 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)

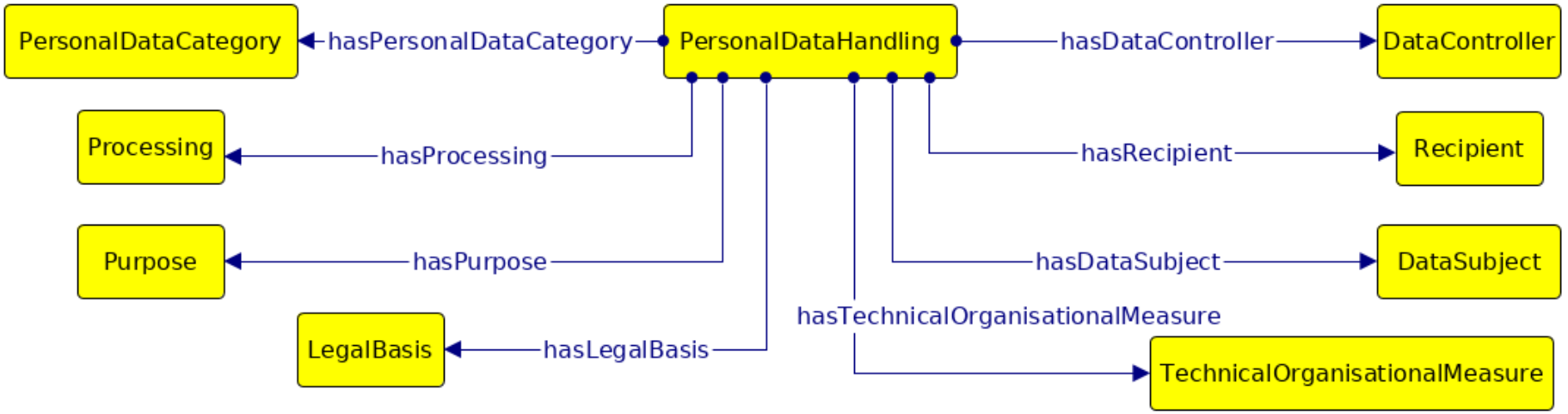
Querying and Validation of information presented in Chapter 6

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)	✓	✓	✓	✓	✓

Table 1.1 Summary of Evaluation Methods

W3C Data Privacy Vocabularies and Controls CG (Sec. 5.5)

- Data Privacy Vocabulary (DPV) [78] for representing information about data handling based on legal requirements – including GDPR
- Represents wider community consensus on information modelling and requirements
- Most comprehensive vocabulary/taxonomy to date
- GDPRtEXT, GDPRov, GConsent were part of SotA analysed
- More generic and abstract e.g. does not model provenance
- GConsent provided input for consent modelling
- Initiated under SPECIAL project [70]
- Specification: http://w3.org/ns/dpv
- Author was co-editor of specification, co-lead author of publication [38] and SPECIAL deliverable 6.5 [70]
- Currently serving as co-chair from Jan 2020



Conclusion

Research Question:
To what extent can
information regarding activities
associated with processing of personal data and consent
be represented using Semantic Web technologies
for GDPR compliance?

1) Linking of information with GDPR
2) Information Representation
3) Querying
4) Validation
5) Compliance Evaluation

Contributions of Thesis

1) GDPRtEXT ontology and resource
2) GDPRov ontology
3) GConsent ontology
4) SPARQL queries for information retrieval
5) SHACL to validate information, link to GDPR
6) Framework for ex-ante and ex-post tests
7) Contribution to DPV(CG)
8) Open-access resources

Student Project Supervision

- Browser extension to record consent (4thyear, 2020)
- Contextual Integrity for GDPR (MSc, 2020)
- Privacy policy generator (4th year, 2020)
- Privacy policy visualisation (4th year, 2019, 2020)

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 in information management – ICSC 2019, ESWC MEPDaW 2018
8) Investigated application on privacy policies – IPAW 2018, ISWC WOP 2018, TELERISE/ADBIS 2018
9) DPV – ODBASE 2019

Future Work

- Align with DPV (as extentions)
- Data Protection Impact Assessments (DPIA)
- Controller – Processor agreements/contracts
- Consolidate court cases and DPA decisions
- Privacy policy representation
- Representation of policy e.g. ODRL

Funding and Opportunities

- Ireland Postdoc Fellowship (2020) – applied
- H2020 IoT project 2020
- NGI funding calls
- ICO (UK) funding call

Formatting Guide

- Footnotes[1] refer to references within slides
- References [1] refer to references in thesis

List of Additional Slides

1) Information about GDPR
2) Scope of Research