

A Semantic Web Ontology for expressing GDPR compliance over consent & data lifecycles

Harshvardhan J. Pandit

Supervised by: Dave Lewis
Co-Supervised by: Declan O'Sullivan

Theme E
ADAPT Centre
Trinity College Dublin

The ADAPT Centre is funded under the
SFI Research Centres Programme (Grant 13/RC/2106)
and is co-funded under the European Regional Development Fund.

Presentation Index

12~13mins ; 25 slides

- 1) Motivation
- 2) Research Question & Contribution
- 3) State of the Art
- 4) Work done till date
- 5) Completion Plan
- 6) Overview ; Summary

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

GDPR

General Data Protection Regulation

- Needs 'valid' given consent
- Fines 4% of global turnover
- Record of processing activity
- Data Protection Officer to monitor compliance
- Demonstrate compliance → *Past*
- Plan & Maintain compliance → *Future*

Research

Area and Domain

- Express legal obligations → ODRL
- Infrastructure for GDPR compliance
- Metadata modeling, storing, and querying

Provenance Metadata

- Activity and Entity
- i.e. Consent and Personal Data lifecycles

Research Question

To what extent can **GDPR obligations** be expressed and evaluated using **compliance queries** over **provenance of consent & data lifecycles** expressed using semantic web ontologies?

Research Objectives

- 1) State of the art in representing provenance of consent & data lifecycles
- 2) Demonstrate effectiveness of SPARQL as a query mechanism for retrieving compliance related information from provenance graph
- 3) Explore how constraints over provenance graphs can be used as a mechanism for expressing and evaluating GDPR obligations

Research Contribution

Major Contribution

Creation of an ontology to express GDPR compliance based on obligations evaluated over provenance graphs of consent and data lifecycles

Minor Contributions

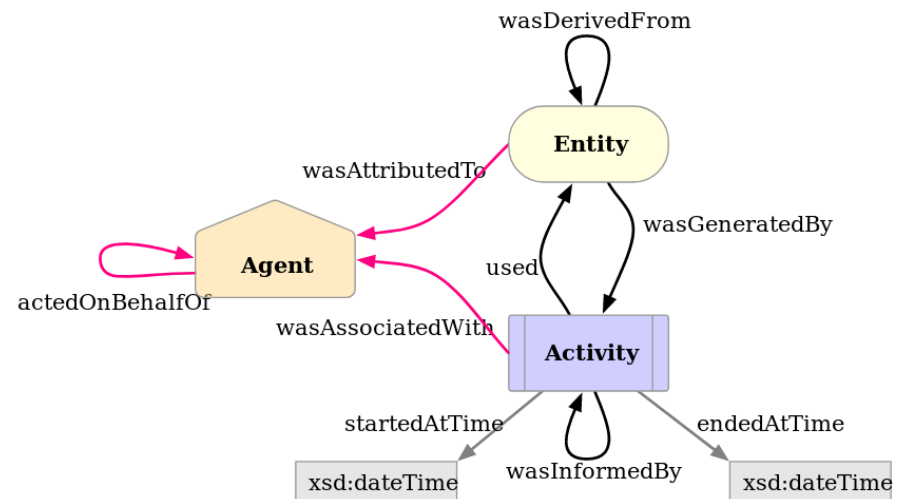
- Ontology to express provenance
- Constraints to express compliance obligations
- GDPR as a resource

State of the Art

Provenance - PROV Ontology (PROV-O)

<https://www.w3.org/TR/prov-o/>

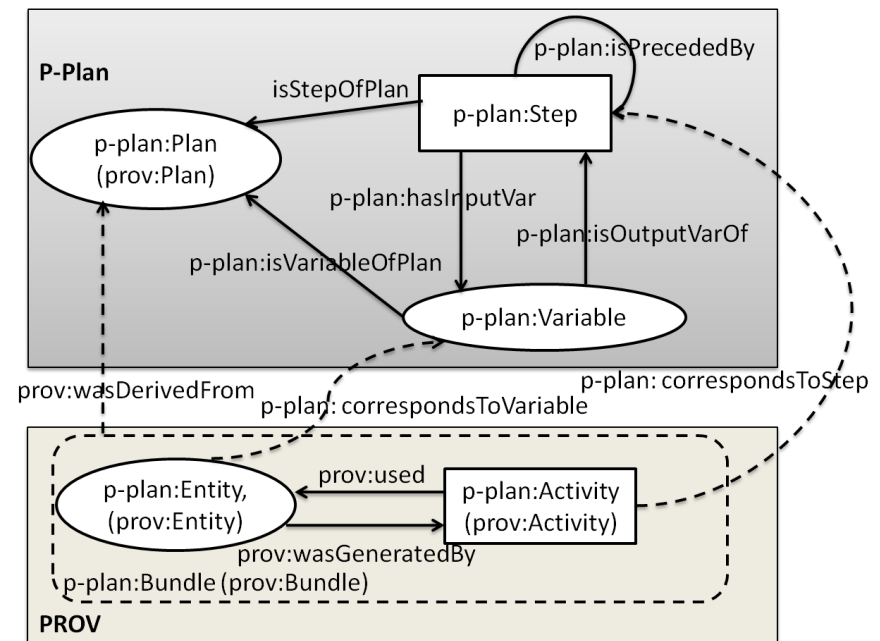
- OWL2 ontology to express provenance
- W3C Recommendation 30-APR-2013
- Interaction between Activity, Entity, Agents
- Record history (*past*)



State of the Art Provenance - P-Plan

<http://vocab.linkeddata.es/p-plan/>

- Extension of PROV-O
- Represent 'plan' that guided execution
- Model execution that is yet to happen (*future*)
- Common template
- Individual instantiations using PROV-O



State of the Art

Constraints - Shapes Constraint Language (SHACL)

<https://www.w3.org/TR/shacl/>

- W3C Recommendation
20-JUL-2017
- Describe and validate
RDF graphs
- What should the graph
'look like' to be valid?
- *SHACL-SPARQL* to define
constraint using SPARQL

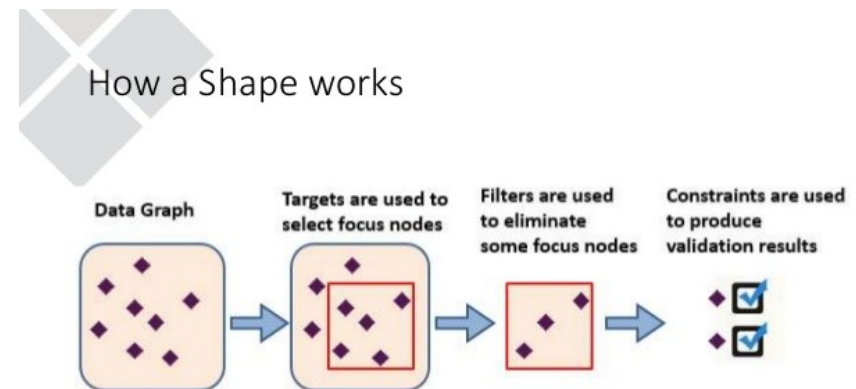


Diagram: Dimitris Kontokostas

State of the Art

Graph patterns / fragments

WHY?

- Different use-cases
- Common requirements
- Easier to query
- Evaluate constraints
- Check information exists

HOW?

- Pattern detection
- Extracting fragments
- Isomorphism
- Reduction
- Abstraction
- Change detection

State of the Art Legislation & Privacy

- GDPR
 - Bartolini et. al.
 - Model GDPR
 - Based on draft
 - Log / Blockchain [4, 56]
 - Impact Assessment
 - Other legislation
HIPAA, PIPPA, DPD, ...
- Privacy Policy
 - UsablePrivacy Project
<https://usableprivacy.org/>
115 annotated policies
- EuroPrise
<https://www.european-privacy-seal.eu/>
GDPR certification
through audit

Work done till date

Gathering Requirements

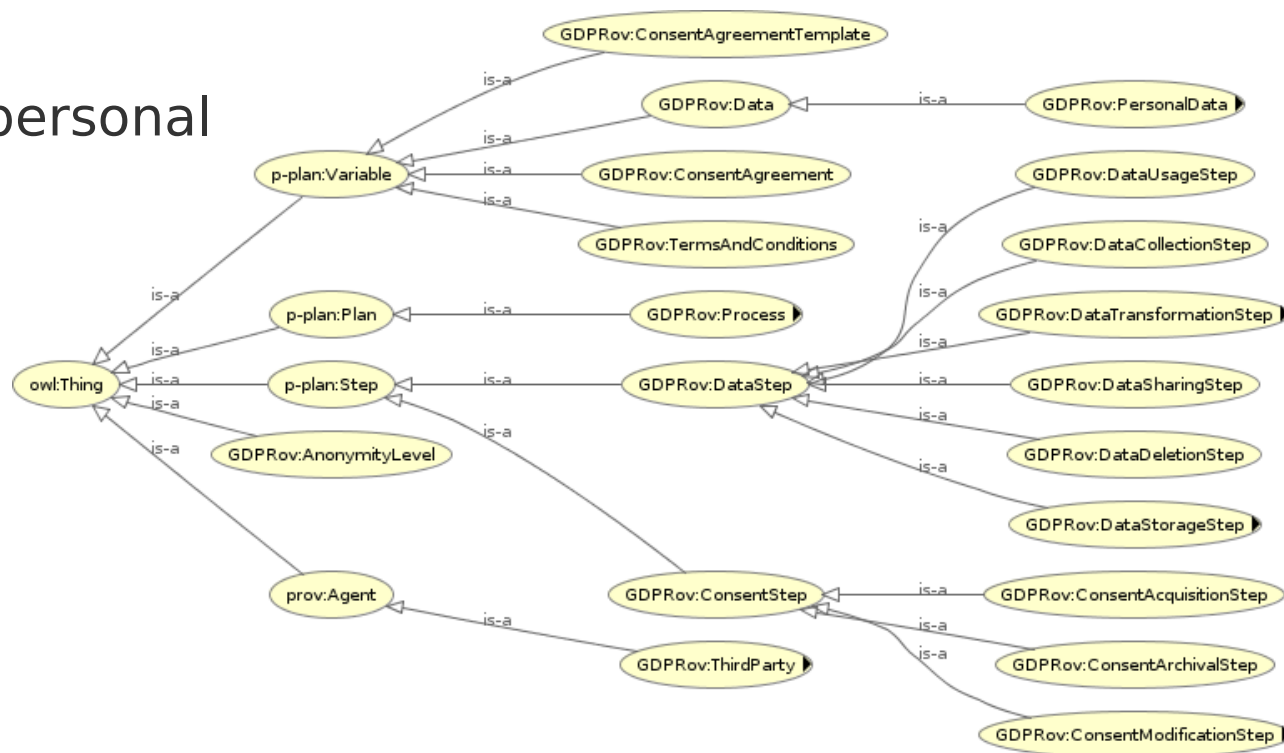
Goal: Representing provenance of consent & data lifecycles using semantic web vocabularies

- Identify requirements by reading GDPR
- Evaluate existing work
- PROV-O and P-Plan identified to be suitable
- Extend with relevant terminology

Work done till date

GDPRov - GDPR Provenance Ontology

- Separation between personal data and consent activities and entities
- GDPR terminology
- Published at PrivOn workshop co-located with ISWC 2017 [5]



Work done till date

GDPRov - query using SPARQL

```
PREFIX GDPRov:
  <https://openscience.adaptcentre.ie/ontologies/GDPRov#>

SELECT ?data ?sharestep ?isAnonymised ?anonymisationStep
WHERE {
  ?data a GDPRov:Data .
  ?sharestep a GDPRov:DataSharingStep .
  ?sharestep GDPRov:sharesData ?data.
  BIND (
    EXISTS { ?data a GDPRov:AnonymisedData . }
    as ?isAnonymised ) .
  OPTIONAL {
    ?anonymisationStep
    GDPRov:generatesAnonymisedData ?data .
  }
}
```

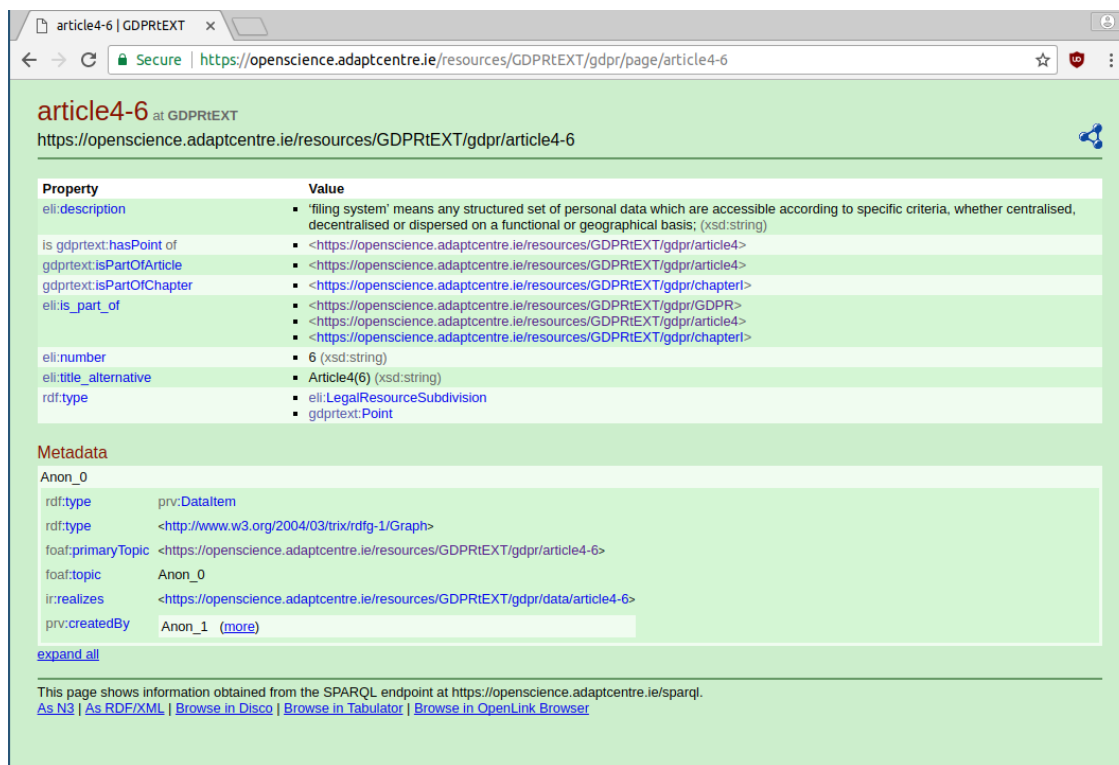
Query to retrieve data shared with third parties, whether that data was anonymised, and if yes, then using which process

data	shareStep	isAnonymised	anonymiserStep
productsSold	productAnalytics	false	NULL
billingInfo	billingAnalytics	false	NULL
customerInfo	profiling	true	anonymiseUsers

Work done till date

GDPRtEXT - GDPR as a linked data resource

- RDF dataset
- Annotated HTML version
- SPARQL endpoint
- DCAT catalog
- Refer to specific points within GDPR
- Publication opportunity ESWC 2018 resource track



article4-6 at GDPRtEXT
<https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/page/article4-6>

Property	Value
eli:description	▪ 'filing system' means any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis; (xsd:string)
is gdprtext:hasPoint of	▪ https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article4-6
gdprtext:isPartOfArticle	▪ https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article4-6
gdprtext:isPartOfChapter	▪ https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapter1
eli:is_part_of	▪ https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/GDPR ▪ https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article4-6 ▪ https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/chapter1
eli:number	▪ 6 (xsd:string)
eli:title_alternative	▪ Article4(6) (xsd:string)
rdf:type	▪ eli:LegalResourceSubdivision ▪ gdprtext:Point

Metadata

Anon_0

rdf:type	priv:DataItem
rdf:type	http://www.w3.org/2004/03/trix/rdg-1/Graph
foaf:primaryTopic	https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/article4-6
foaf:topic	Anon_0
ir:realizes	https://openseience.adaptcentre.ie/resources/GDPRtEXT/gdpr/data/article4-6
priv:createdBy	Anon_1 (more)

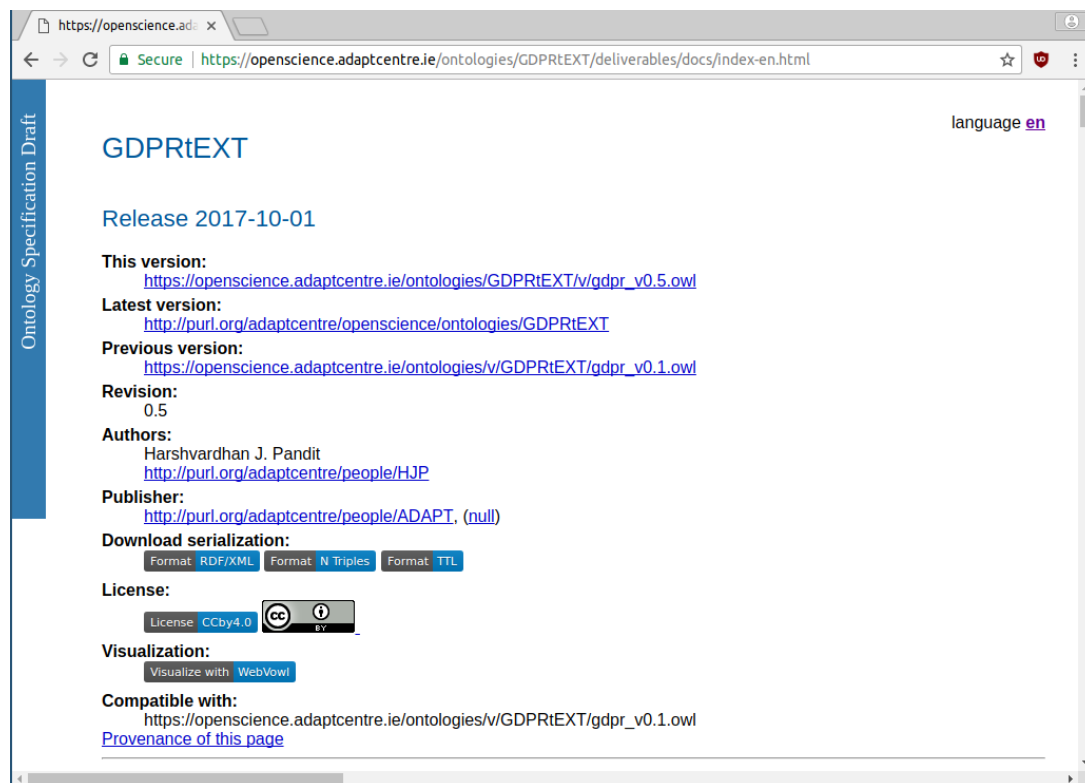
[expand all](#)

This page shows information obtained from the SPARQL endpoint at <https://openseience.adaptcentre.ie/sparql>.
[As N3](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

Work done till date

GDPRtEXT - GDPR ontology

- Defines terms using *skos:Concept*
- Link related terms
- 200+ concepts for GDPR
 - Rights
 - Compliance
 - Obligations
 - Data & Consent



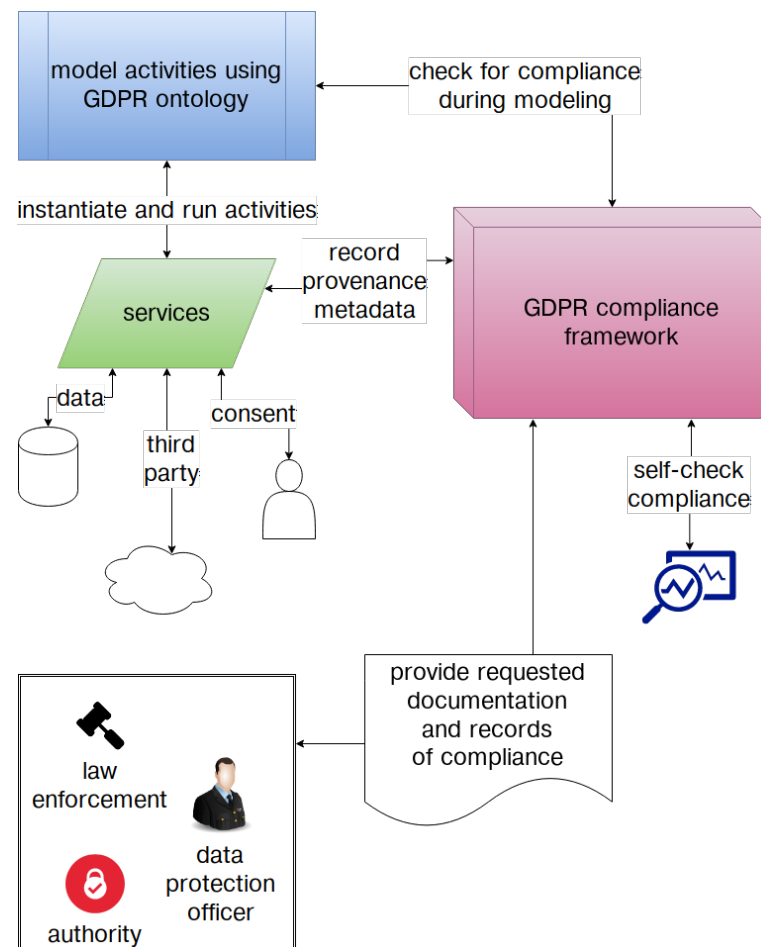
The screenshot shows a web browser displaying the GDPRtEXT ontology page. The page title is "GDPRtEXT" and it is identified as an "Ontology Specification Draft". The page includes the following information:

- Release:** 2017-10-01
- This version:** https://opscience.adaptcentre.ie/ontologies/GDPRtEXT/v/gdpr_v0.5.owl
- Latest version:** <http://purl.org/adaptcentre/opscience/ontologies/GDPRtEXT>
- Previous version:** https://opscience.adaptcentre.ie/ontologies/v/GDPRtEXT/gdpr_v0.1.owl
- Revision:** 0.5
- Authors:** Harshvardhan J. Pandit, <http://purl.org/adaptcentre/people/HJP>
- Publisher:** <http://purl.org/adaptcentre/people/ADAPT>, (null)
- Download serialization:** Buttons for "Format: RDF/XML", "Format: N Triples", and "Format: TTL".
- License:** CC BY 4.0
- Visualization:** Button for "Visualize with: WebVowl"
- Compatible with:** https://opscience.adaptcentre.ie/ontologies/v/GDPRtEXT/gdpr_v0.1.owl
- [Provenance of this page](#)

Work done till date

Collaborative Work

- Consent and Data Management Framework [9]
 - Consent ontology
 - Model GDPR compliance activities and processes
- Data Protection Rights Language [3]
 - ODRL extension
 - Model data sharing agreements between third-parties



PhD Completion Plan

Phase I - Compliance queries for provenance lifecycles

- Corpus of use-cases using *GDPROv*
- UsablePrivacy to understand data-related activities
- EuroPriSe as use-case
- Identify information needed for compliance
- Express query to retrieve this information
- *SPARQL + SPIN* → *capture SPARQL queries as RDF data*

Hypothesis: Provenance information pertaining to GDPR obligations can be retrieved using SPARQL queries

PhD Completion Plan

Phase I - Compliance queries for provenance lifecycles

Evaluation criteria

- Difficult to measure as data cannot be evaluated as qualitative nor quantitative
- Focus on question -
“how much information is present?”
- Which obligations can be expressed as queries?
- What % of GDPR can be covered using these queries?

PhD Completion Plan

Phase II - Expressing constraints over provenance lifecycles

- Gather use-cases from real-world compliance
- Take obligations from Phase I, and express as constraints using *SHACL*
- Evaluate the constraints → provides a metric
- Evaluation can be both the presence as well as structure of information

Hypothesis: GDPR obligations can be expressed as constraints over provenance graphs and evaluated as compliance queries

PhD Completion Plan

Phase II - Expressing constraints over provenance lifecycles

- *SHACL-SPARQL* can use SPARQL to model query for constraint ; re-use queries from Phase I
- Compare two approaches to find benefits and drawbacks ; see which model fits which obligation
- To make evaluation easier, use graph-based approaches to reduce amount of complexity
- An ontology to express compliance information based on the evaluation of these approaches

Summary

Publications

- 1) Modelling provenance for GDPR compliance using linked open data vocabularies** (2017 Workshop) *Harshvardhan J. Pandit, Dave Lewis*. Society, Privacy and the Semantic Web - Policy and Technology (PrivOn), co-located with ISWC 2017
- 2) Compliance through Informed Consent: Semantic Based Consent Permission and Data Management Model** (2017 Workshop) *Kaniz Fatema, Ensar Hadziselimovic, Harshvardhan J. Pandit, Dave Lewis*. Society, Privacy and the Semantic Web - Policy and Technology (PrivOn), co-located with ISWC 2017
- 3) Linked Data Contracts to Support Data Protection and Data Ethics in the Sharing of Scientific Data** (2017 Workshop) *Ensar Hadziselimovic, Kaniz Fatema, Harshvardhan J. Pandit, Dave Lewis*. Sharing of Scientific Data in Enabling Open Semantic Science (SemSci), co-located with ISWC 2017.
- 4) Utilising Semantic Web Ontologies to publish Experimental Workflows** (2017 workshop) *Harshvardhan J. Pandit, Ensar Hadziselimovic, Dave Lewis*. Joint Proceedings of the 1st International Workshop on Scientometrics and 1st International Workshop on Enabling Decentralised Scholarly Communication co-located with 14th Extended Semantic Web Conference (ESWC 2017)
- 5) The Use of Open Data to Improve the Repeatability of Adaptivity and Personalisation Experiment** (2016 position) *Harshvardhan J. Pandit, Ramisa Hamed, Seamus Lawless, David Lewis*. Published in the proceedings of the 1st EvalUMAP workshop - Towards comparative evaluation in user modeling, adaptation and personalization, held in conjunction with the 24th Conference on User Modeling, Adaptation and Personalization, UMAP 2016.

Summary

Short term and Long term Plan

Next 6 months

- Create use-cases
- *SPARQL* queries
- GDPR obligations
- GDPRtEXT resource
- Publication: ESWC 2018

Next 18 months

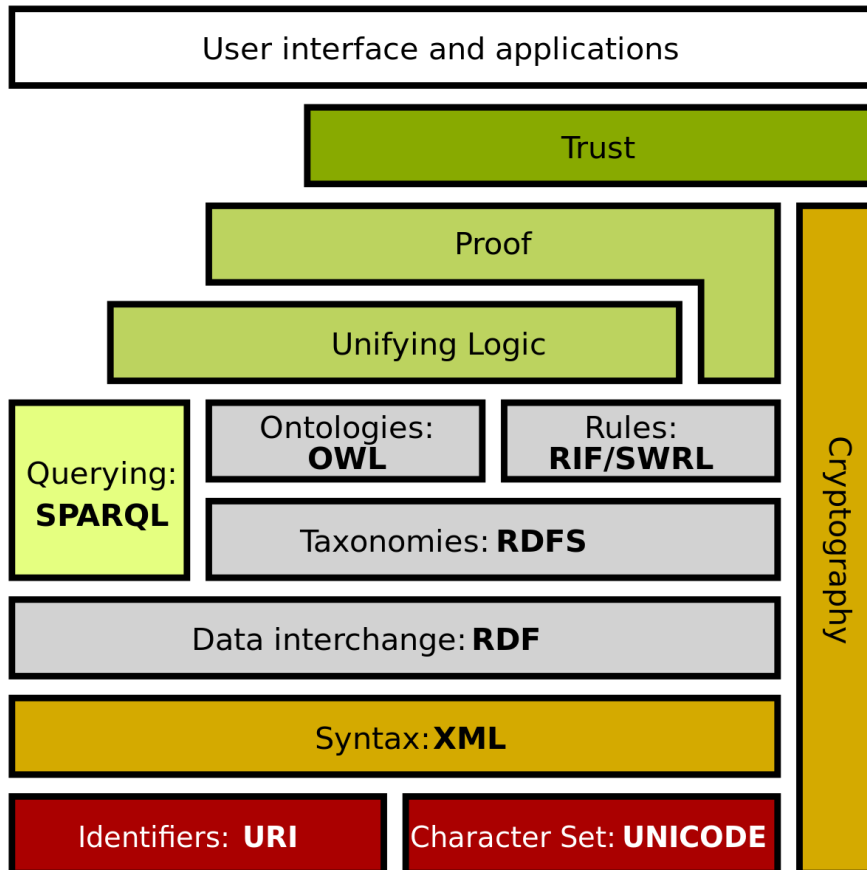
- Evaluate *SPARQL* queries
- Express *SHACL* constraints
- Explore graph techniques
- GDPR impact
- Publication:
ISWC2018, ESWC2019

A Semantic Web Ontology for expressing GDPR compliance over
consent & data lifecycles

END OF PRESENTATION

GDPR

Semantic Web Technologies



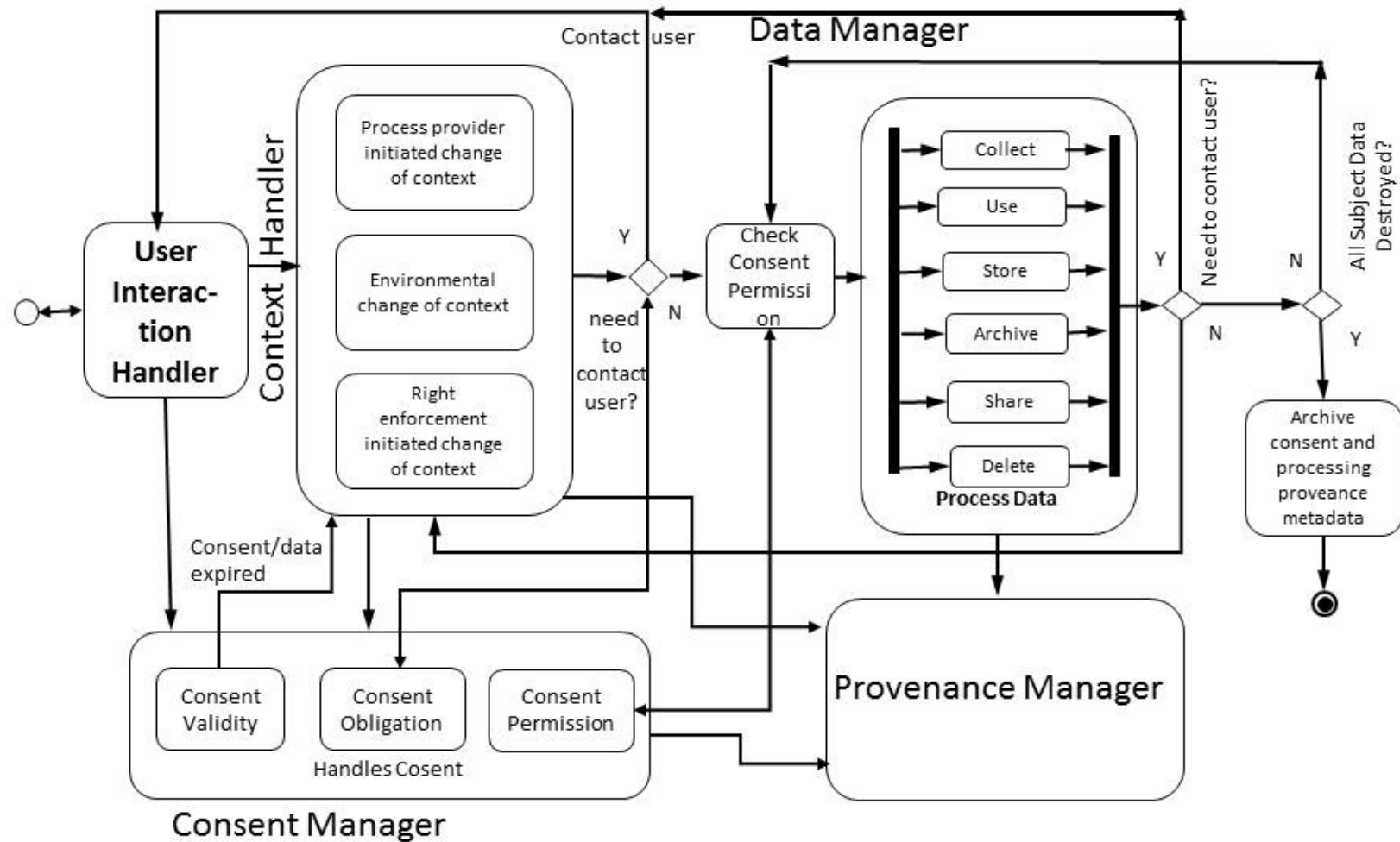
METADATA

- Flexible
- Open
- Shareable
- Extendable
- Queriable

Recital 82

“In order to demonstrate compliance with this Regulation, the controller or processor should **maintain records of processing activities** under its responsibility. Each controller and processor should be obliged to cooperate with the supervisory authority and make those records, on request, available to it, so that it might serve for monitoring those processing operations.”

GDPR - the BIGGER picture



SPARQL query evaluation

Given use-case: personal data is shared with another organisation

Check:

- Where did personal data come from?
- What is the legal justification for using it?
- Does it have given consent?
- Is 3rd party a processor or controller?
- Was the data anonymised? How?

SHACL constraints

- Define constraint
 - Use of data must have legal justification associated → can be via ontology property
 - Personal data must have history of where it came from and under what conditions
 - Sharing of data must be made explicit with reason and identity of entity
 - Can user consent be evaluated within provenance of activities over personal data?

Potential benefits

- Documentation for compliance can be automated to a certain extent
- Some part of legal compliance checking can be automated
- Sharing of information and processes between stakeholders
- Common model for expressing information

PhD Completion Plan

Phase I Build use-cases and compliance queries

- Pre GDPR enforcement
- *SPARQL* queries to retrieve information about consent and data usage required for compliance
- Based on real-world use-cases
- Outcome: compliance queries identified from GDPR

Phase II Express obligations as provenance constraints

- Post GDPR enforcement
- Constraints using *SHACL*
- Takes into factor how organisations have reacted to GDPR w.r.t. privacy policies
- Outcome: ontology to express compliance based on evaluation of constraints

PhD Completion Plan

Phase I - Compliance queries for provenance lifecycles

Evaluation criteria - Compliance score

- Are all obligations of the same severity? Do they have different 'weights'?
- Is absence of information a violation?
- Can we link queries back to specific text in GDPR?

Goal: Create a 'compliance score' based on these questions that can aid in compliance determination

PhD Completion Plan

Collaboration

- ADAPT

- Ensar Hadziselimovic, Gabriel Hogan → GDPR research
- Theme C → personalisation domain

- SPECIAL <https://www.specialprivacy.eu/>

- Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance
- acquisition of consent
- metadata (consent policies, event data, context)
- privacy-aware, secure workflows with access control, transparency and compliance verification