

Modelling provenance for GDPR compliance using linked open data vocabularies

Harshvardhan J. Pandit

Dave Lewis

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

- 1) Background on GDPR
- 2) Compliance for GDPR
- 3) Provenance requirements
- 4) Semantic Web Vocabularies
- 5) GDPRov vocabulary
- 6) SPARQL queries for compliance
- 7) Future Outlook

GDPR

- Greater responsibility & accountability
- Needs 'valid' given consent
- Anonymisation & Pseudo-Anonymisation
- Rights to Data Subjects
- Fines up to 4% of global turnover
- Record of processing activity
- Data Protection Officer to monitor compliance
- Record and notify about Data Breaches

Article 6(1)

Lawfulness of Processing

- Given consent
- Contract
- Legal obligation
- Protect vital interests
- Public interest / Official authority
- Legitimate interests

Controller / Processor

- Controller
 - Determines purposes of processing
 - More responsibilities
 - Appoints Processor(s)
- Processor
 - Processes data on instruction of Controller
 - Cannot act outside given instructions

Compliance

A) Demonstrate Compliance

- i. PAST
- ii. Activities that already have happened
- iii. Records / Archives / Documentation

B) Monitor Compliance (on-going basis)

- i. FUTURE
- ii. Activities yet to happen
- iii. Planning / Modeling

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.”

Provenance

- Flexible in expression
 - Open
 - Shareable
 - Extendable
 - Queriable
 - Can be adapted for GDPR
- } Linked Open Data

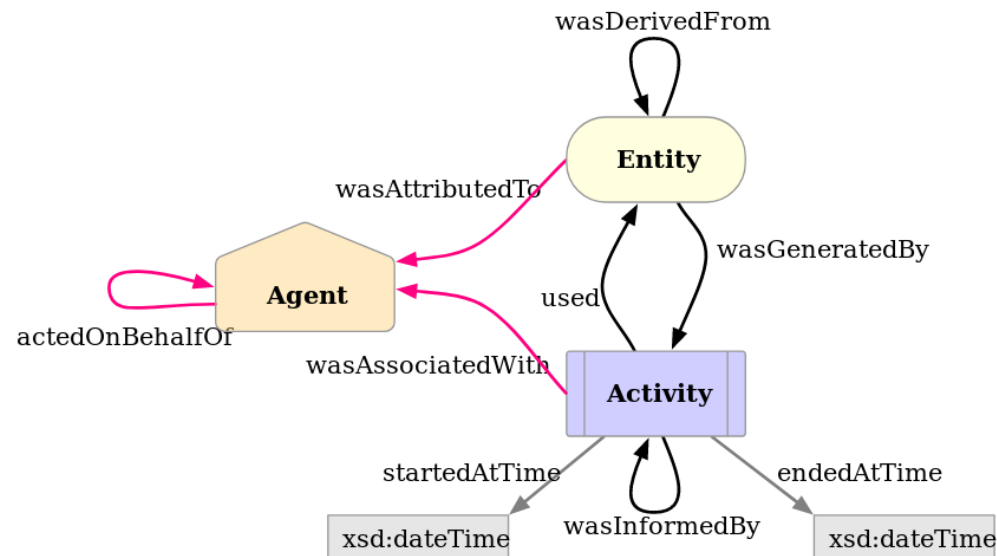
Semantic Web Vocabularies

- Express → RDF
- Model → OWL
- Query → SPARQL
- Collaborate → Open World Assumption eases creating abstract common model

PROV Ontology (PROV-O)

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

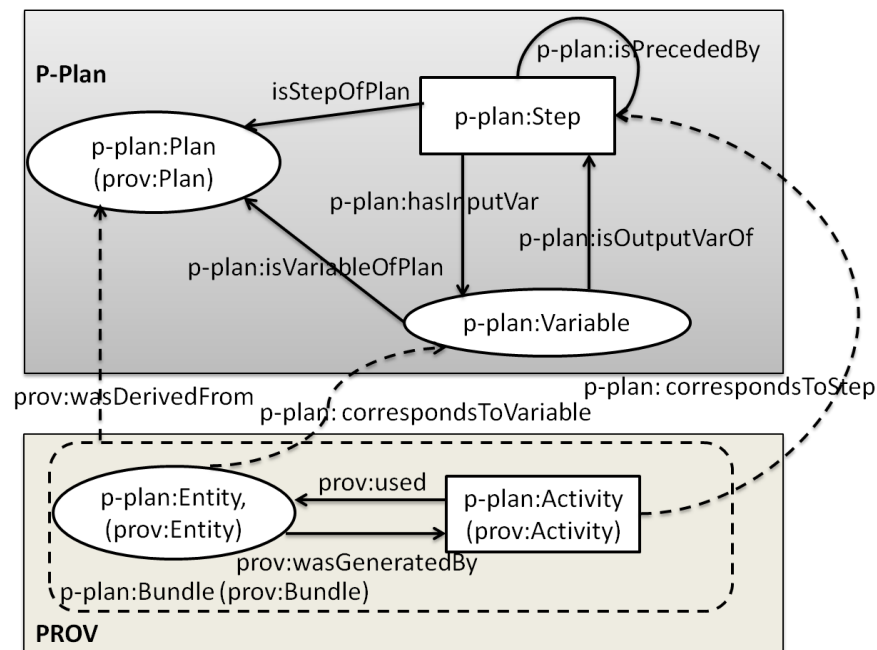
- OWL2 ontology to express provenance
- W3C Recommendation 30-APR-2013



P-Plan

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

- Extension of PROV-O
- Represent 'plan' that guided execution



Research Hypothesis

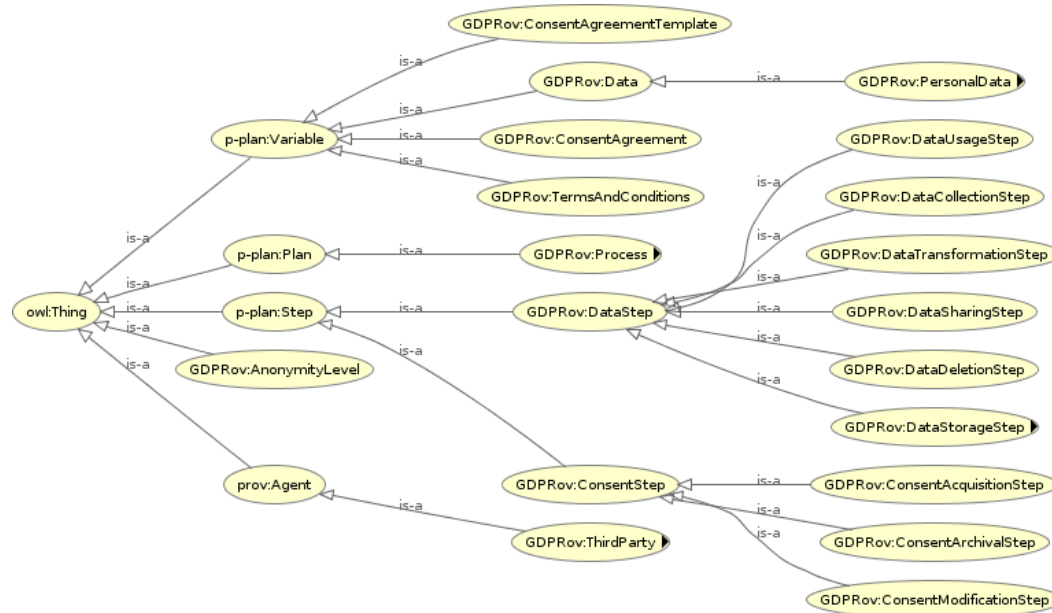
- Provenance of activities and entities
- Track / Model with relevance to compliance
- Uses obligations specified by GDPR
- Uses semantic web ontologies
- Can be queried using SPARQL

QUESTION: to what extent?

GDPRov

GDPR Provenance Ontology

- PROV-O → past executions
- P-PLAN → future activities



GDPROv - Data

- Personal Data
 - Sensitive personal data
 - (Pseudo-)Anonymous data
- Data Step
 - Use / Share / Delete / Transform / Modify
 - p-plan:Step → p-plan:Activity | prov:Activity

GDPROV - Consent

- Given Consent
 - ConsentAgreement (given consent)
 - ConsentAgreementTemplate (choices given)
- Consent Step
 - Acquisition / Modification / Archival
 - p-plan:Step → p-plan:Activity | prov:Activity

GDPROv - Process

- Combine steps into a cohesive activity
- Uses p-plan:Plan | prov:Plan
- GDPR Rights
 - Data erasure, Consent Withdrawal, Data Rectification, etc.
- Data Breach
 - Notify authority and data subject, record breach extent and actions, etc.

SPARQL query

```
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 what process?

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

GDPR linked data resource

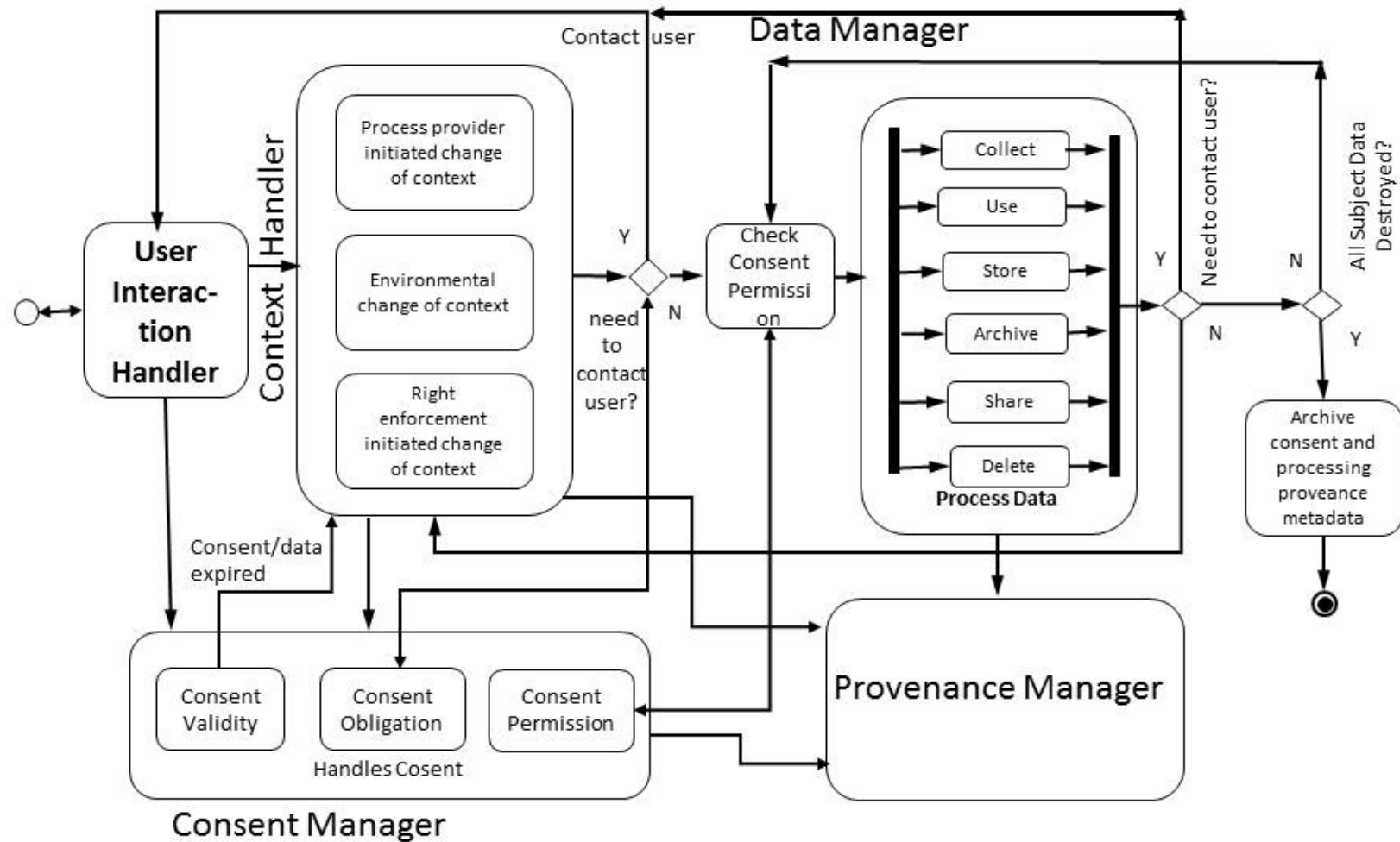
- References to canonical text are needed to -
 - Associate provenance
 - Cover obligations
 - Track compliance queries
- GDPR as linked data resource
 - Use SKOS to model concepts
 - Use European Legislation Identifier (ELI)

GDPRtEXT

GDPR text EXTensions

- GDPR text as RDF & Annotated HTML
- Define concepts from GDPR as *skos:Concept*
- Relates terms to where they occur
- Relates terms to each other
- Accessible online
<https://openscience.adaptcentre.ie>

GDPR - the BIGGER picture



Quick Recall - Contributions

- Expressing Provenance of Entity/Activity by extending PROV-O / P-Plan
- Modelling ontology on GDPR obligations and terms
- Using SPARQL to retrieve provenance information relevant to compliance
- Creation of GDPRtEXT resource

Future Outlook

- Model use-cases for provenance and compliance queries
- UsablePrivacy project → annotated privacy policies that describe data collection process
- Create a vocabulary to express GDPR terms
- Keep a lookout for privacy policies updated to reflect requirements of GDPR
- Google Cloud recently published their GDPR privacy policy

Let's Discuss!!!

- GDPR
- Provenance
- Semantic Web Vocabularies
- GDPRov model
- SPARQL queries for compliance
- GDPRtEXT resource
- Collaboration
me + you = ideas