Engaging Content
Engaging People

# Test-driven approach towards GDPR compliance

**Harshvardhan J. Pandit**, Declan O’Sullivan, Dave Lewis
ADAPT Centre, School of Computer Science & Statistics, Trinity College Dublin, Ireland
email: pandith@tcd.ie
website: http://openscience.adaptcentre.ie/
resources: https://w3id.org/GDPRep/semantic-tests

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

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 Structure

1. Aims and Goals
2. Use-case: Consent mechanism on a website
3. Information requirements
4. Creating dataset in RDF
5. Some interesting findings
6. GDPR requirements expressed as constraints
7. Testing constraints using SHACL
8. Results of compliance process
9. Discussion
10. Q&A

# Domain / Scope

A typical scenario is to start from a process model or workflow of how the service/operations take place, assess it for compliance based on obligations and requirements obtained from interpretation of specific clauses of the law, and tweak it as required based on the outcome to become compliant.

This requires a way to express information about:

1) Process model / Workflow + Provenance Log
2) Constraints / Obligations / Requirements
3) Links to specific aspects of law

# Aim – Testing process workflows for compliance

**Aim:** Test and record compliance information for process workflows

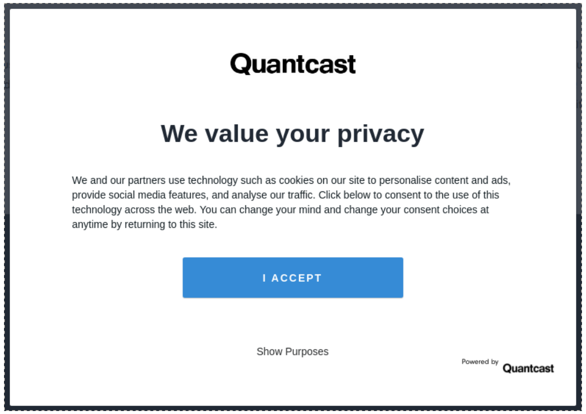
**Objective:** information (including compliance) can be

1) persisted 2) queried 3) validated 4) linked to legal requirements

We choose Semantic Web because:

a) Linked Data
b) Interoperable Standards (RDF, OWL, SPARQL, SHACL)
c) Creating Knowledge Graph i.e. embedding semantics
d) Extensible based on further use-cases as needed

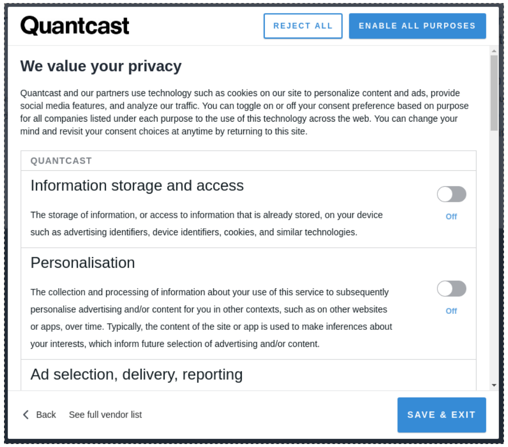
# Use-case: consent mechanism on Quantcast.com



We value your privacy

We and our partners use technology such as cookies on our site to personalise content and ads, provide social media features, and analyse our traffic. Click below to consent to the use of this technology across the web. You can change your mind and change your consent choices at anytime by returning to this site.

(a)



We value your privacy

Quantcast and our partners use technology such as cookies on our site to personalize content and ads, provide social media features, and analyze our traffic. You can toggle on or off your consent preference based on purpose for all companies listed under each purpose to the use of this technology across the web. You can change your mind and revisit your consent choices at anytime by returning to this site.

QUANTCAST

Information storage and access — Off

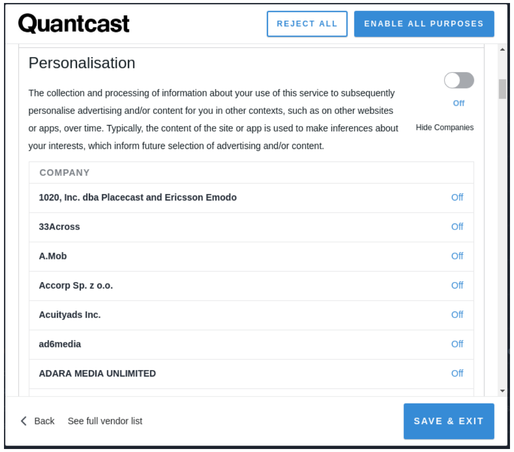
The storage of information, or access to information that is already stored, on your device such as advertising identifiers, device identifiers, cookies, and similar technologies.

Personalisation — Off

The collection and processing of information about your use of this service to subsequently personalise advertising and/or content for you in other contexts, such as on other websites or apps, over time. Typically, the content of the site or app is used to make inferences about your interests, which inform future selection of advertising and/or content.

Ad selection, delivery, reporting

(b)

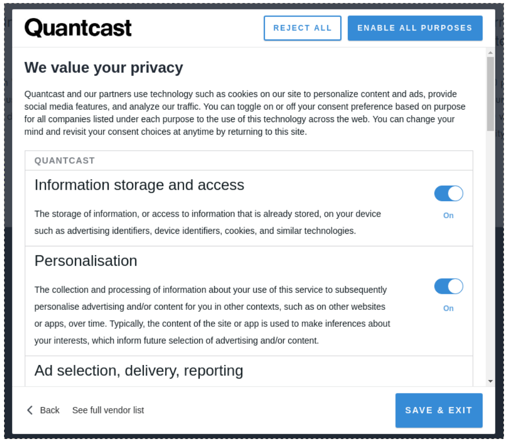


Personalisation — Off

The collection and processing of information about your use of this service to subsequently personalise advertising and/or content for you in other contexts, such as on other websites or apps, over time. Typically, the content of the site or app is used to make inferences about your interests, which inform future selection of advertising and/or content.

| COMPANY | |
|---|---|
| 1020, Inc. dba Placecast and Ericsson Emodo | Off |
| 33Across | Off |
| A.Mob | Off |
| Accorp Sp. z o.o. | Off |
| Acuityads Inc. | Off |
| ad6media | Off |
| ADARA MEDIA UNLIMITED | Off |

(c)



We value your privacy

Quantcast and our partners use technology such as cookies on our site to personalize content and ads, provide social media features, and analyze our traffic. You can toggle on or off your consent preference based on purpose for all companies listed under each purpose to the use of this technology across the web. You can change your mind and revisit your consent choices at anytime by returning to this site.

QUANTCAST

Information storage and access — On

The storage of information, or access to information that is already stored, on your device such as advertising identifiers, device identifiers, cookies, and similar technologies.

Personalisation — On

The collection and processing of information about your use of this service to subsequently personalise advertising and/or content for you in other contexts, such as on other websites or apps, over time. Typically, the content of the site or app is used to make inferences about your interests, which inform future selection of advertising and/or content.

Ad selection, delivery, reporting

(d)

Why Quantcast?
- It is one of the largest consent dialogue providers
- Option to change consent

Screenshots show:
(a) first dialogue
(b) second set of options
(c) list of third parties
(d) allows to change consent in subsequent visits

Also investigate:
1. Privacy Policy
2. Subject Access Request
3. Products offered on site

# Modeling the data in RDF

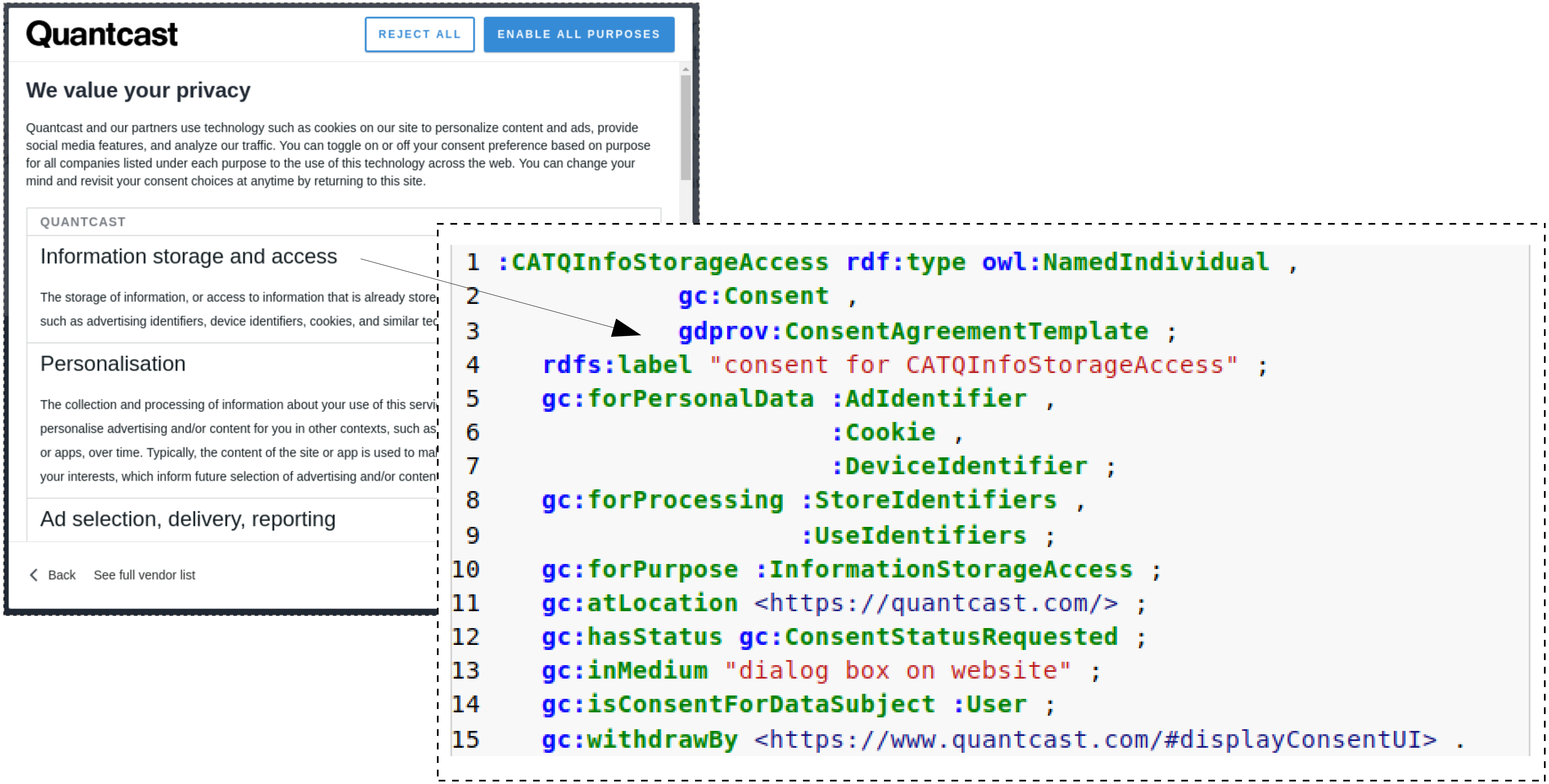
Used vocabularies (also prior work!):

- **GDPRtEXT**: a RDF version of GDPR text, UID for every clause in text, SKOS thesauri of concepts [13] https://w3id.org/GDPRtEXT
- **GDPRov**: extends PROV-O and P-Plan with GDPR specific concept to represent logs and plans/models/templates [16] https://w3id.org/GDPRov
- **Gconsent**: vocabulary for capturing information about consent as per requirements of the GDPR [14] https://w3id.org/GConsent
- SPARQL queries to retrieve information for GDPR compliance [15] https://w3id.org/GDPRep/checklist-demo

Interpret information in the consent dialogue, and also investigate privacy policy and information on the website to get: purpose & categories of processing, personal data categories, legal basis, recipients of data, storage duration

[13] Pandit, H.J. et al.: GConsent - A Consent Ontology based on the GDPR. ESWC 2019
[14] Pandit, H.J. et al.: GDPRtEXT - GDPR as a Linked Data Resource. ESWC 2018
[15] Pandit, H.J. et al.: Queryable Provenance Metadata For GDPR Compliance. SEMANTiCS 2018
[16] Pandit, H.J., Lewis, D.: Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies. PrivOn (ISWC 2017)

# Modeling the data in RDF



We value your privacy

Quantcast and our partners use technology such as cookies on our site to personalize content and ads, provide social media features, and analyze our traffic. You can toggle on or off your consent preference based on purpose for all companies listed under each purpose to the use of this technology across the web. You can change your mind and revisit your consent choices at anytime by returning to this site.

QUANTCAST

Information storage and access

The storage of information, or access to information that is already store
such as advertising identifiers, device identifiers, cookies, and similar tec

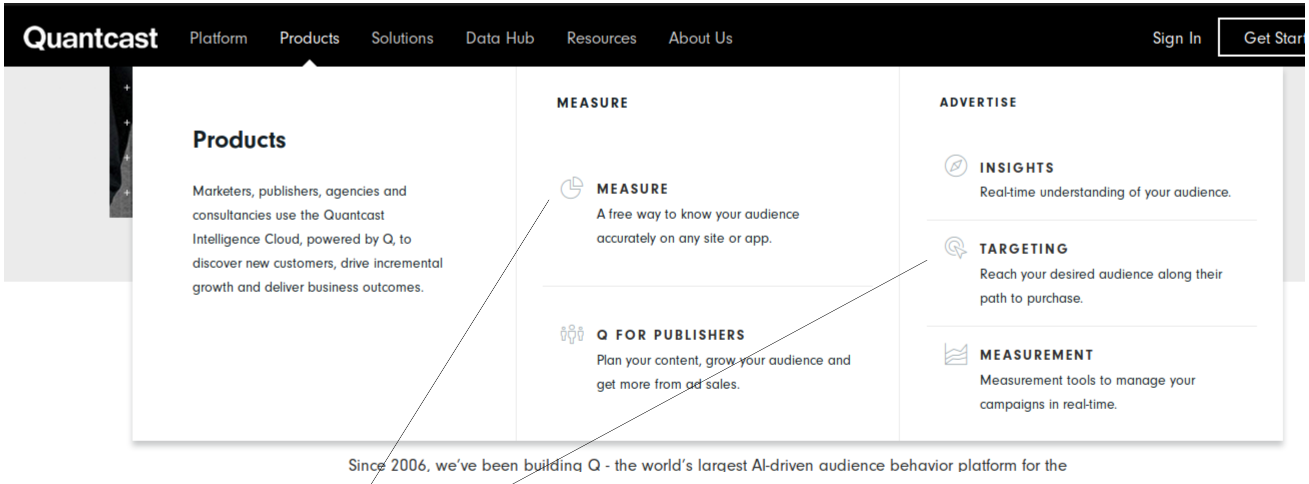
Personalisation

The collection and processing of information about your use of this servi
personalise advertising and/or content for you in other contexts, such as
or apps, over time. Typically, the content of the site or app is used to ma
your interests, which inform future selection of advertising and/or conten

Ad selection, delivery, reporting

```
:CATQInfoStorageAccess rdf:type owl:NamedIndividual ,
            gc:Consent ,
            gdprov:ConsentAgreementTemplate ;
   rdfs:label "consent for CATQInfoStorageAccess" ;
   gc:forPersonalData :AdIdentifier ,
                      :Cookie ,
                      :DeviceIdentifier ;
   gc:forProcessing :StoreIdentifiers ,
                    :UseIdentifiers ;
   gc:forPurpose :InformationStorageAccess ;
   gc:atLocation <https://quantcast.com/> ;
   gc:hasStatus gc:ConsentStatusRequested ;
   gc:inMedium "dialog box on website" ;
   gc:isConsentForDataSubject :User ;
   gc:withdrawBy <https://www.quantcast.com/#displayConsentUI> .
```

# Modeling the data in RDF



**Products**

Marketers, publishers, agencies and consultancies use the Quantcast Intelligence Cloud, powered by Q, to discover new customers, drive incremental growth and deliver business outcomes.

MEASURE

MEASURE
A free way to know your audience accurately on any site or app.

Q FOR PUBLISHERS
Plan your content, grow your audience and get more from ad sales.

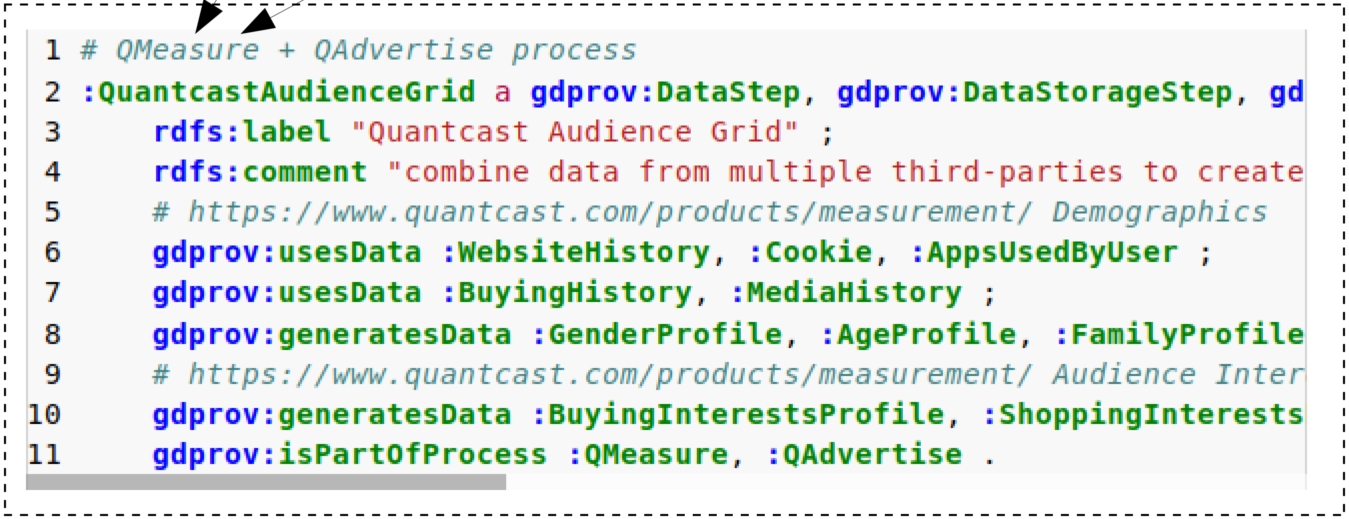
ADVERTISE

INSIGHTS
Real-time understanding of your audience.

TARGETING
Reach your desired audience along their path to purchase.

MEASUREMENT
Measurement tools to manage your campaigns in real-time.

Since 2006, we've been building Q - the world's largest AI-driven audience behavior platform for the



```
# QMeasure + QAdvertise process
:QuantcastAudienceGrid a gdprov:DataStep, gdprov:DataStorageStep, gd
    rdfs:label "Quantcast Audience Grid" ;
    rdfs:comment "combine data from multiple third-parties to create
    # https://www.quantcast.com/products/measurement/ Demographics
    gdprov:usesData :WebsiteHistory, :Cookie, :AppsUsedByUser ;
    gdprov:usesData :BuyingHistory, :MediaHistory ;
    gdprov:generatesData :GenderProfile, :AgeProfile, :FamilyProfile
    # https://www.quantcast.com/products/measurement/ Audience Inter
    gdprov:generatesData :BuyingInterestsProfile, :ShoppingInterests
    gdprov:isPartOfProcess :QMeasure, :QAdvertise .
```

# Interesting finding

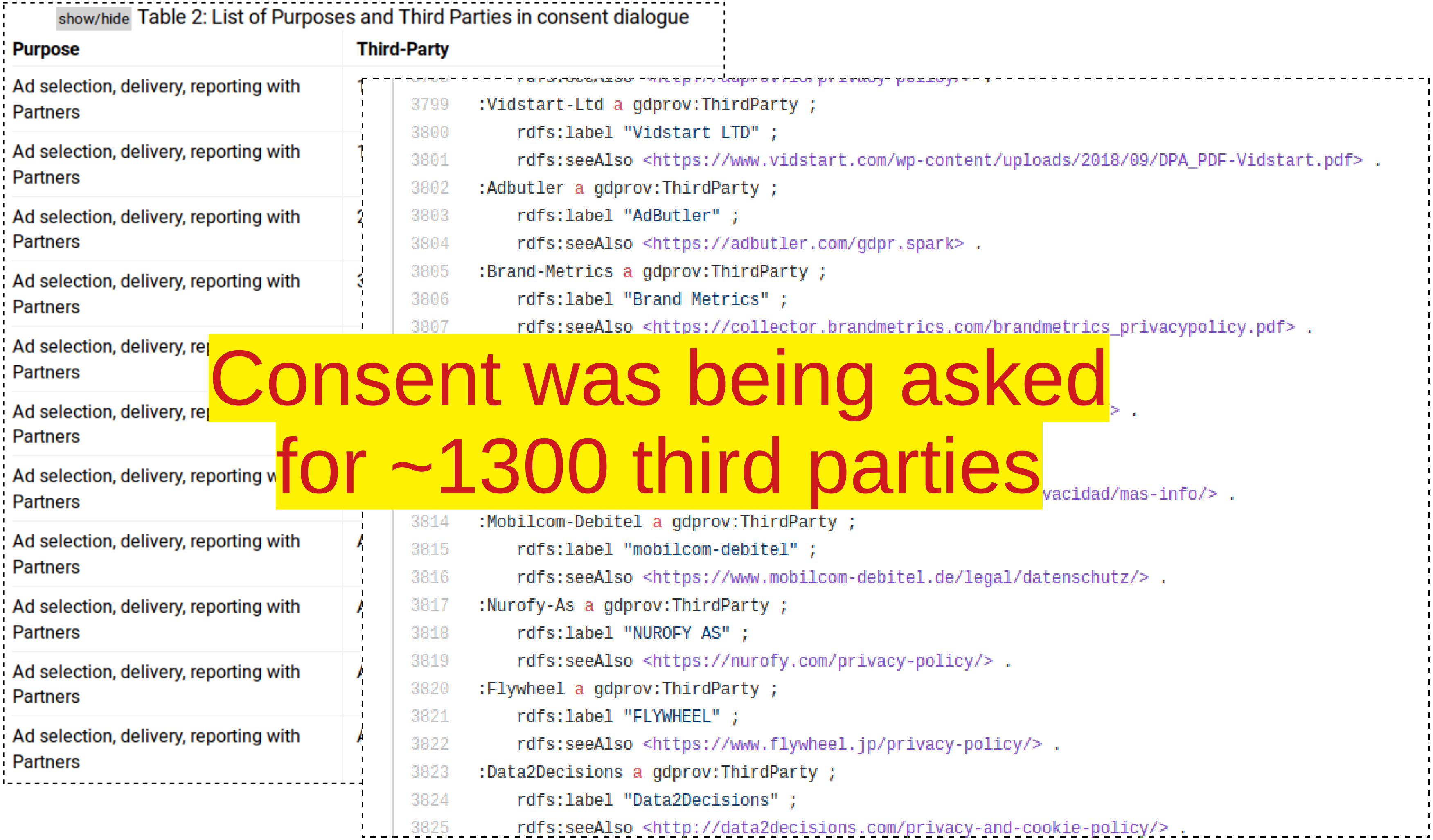


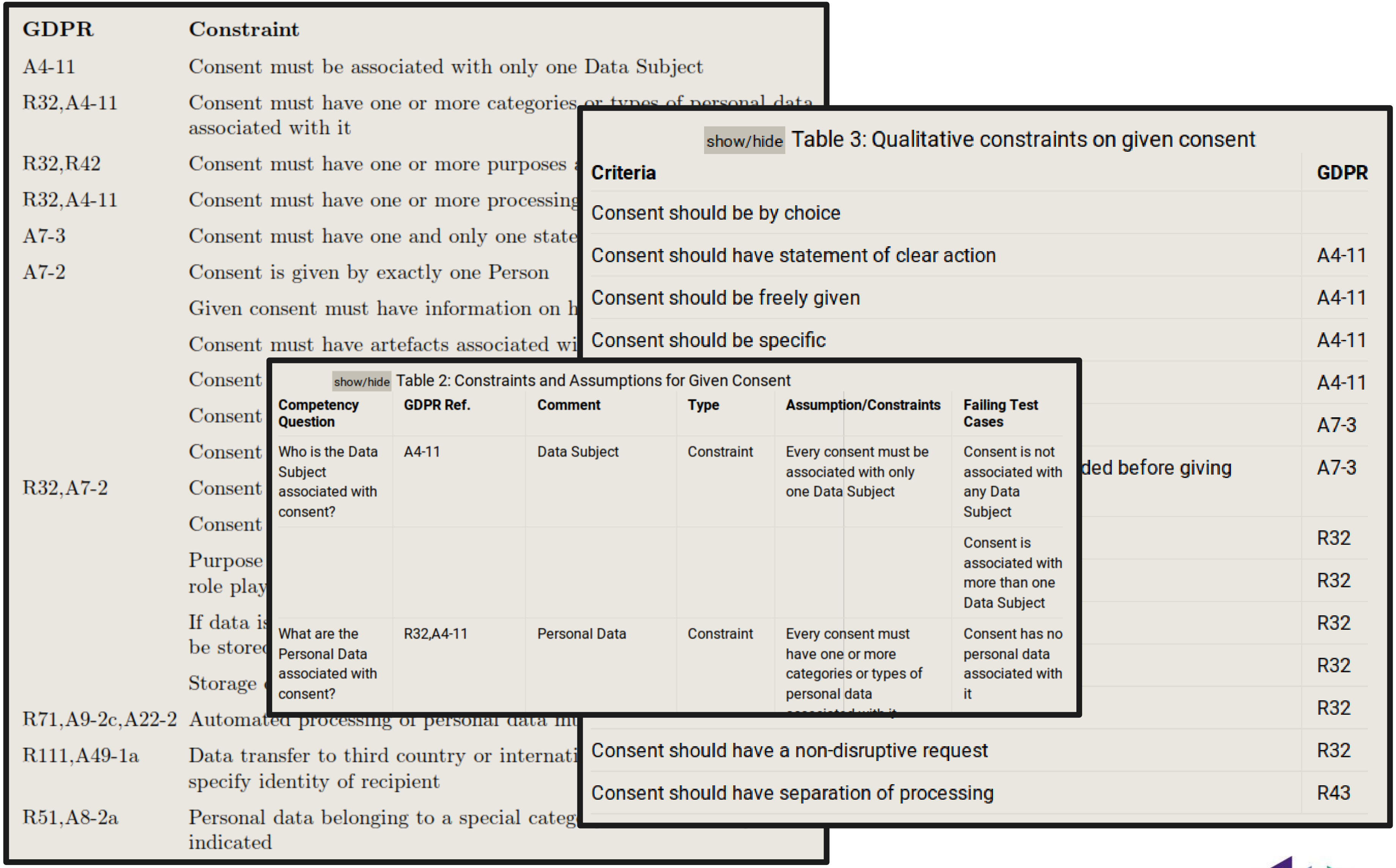
Table 2: List of Purposes and Third Parties in consent dialogue

| Purpose | Third-Party |
|---|---|
| Ad selection, delivery, reporting with Partners | 1 |
| Ad selection, delivery, reporting with Partners | 1 |
| Ad selection, delivery, reporting with Partners | 2 |
| Ad selection, delivery, reporting with Partners | 3 |
| Ad selection, delivery, re | |
| Ad selection, delivery, re | |
| Ad selection, delivery, reporting w | |
| Ad selection, delivery, reporting with Partners | A |
| Ad selection, delivery, reporting with Partners | A |
| Ad selection, delivery, reporting with Partners | A |
| Ad selection, delivery, reporting with Partners | A |

```
:Vidstart-Ltd a gdprov:ThirdParty ;
    rdfs:label "Vidstart LTD" ;
    rdfs:seeAlso <https://www.vidstart.com/wp-content/uploads/2018/09/DPA_PDF-Vidstart.pdf> .
:Adbutler a gdprov:ThirdParty ;
    rdfs:label "AdButler" ;
    rdfs:seeAlso <https://adbutler.com/gdpr.spark> .
:Brand-Metrics a gdprov:ThirdParty ;
    rdfs:label "Brand Metrics" ;
    rdfs:seeAlso <https://collector.brandmetrics.com/brandmetrics_privacypolicy.pdf> .
                                                                                  > .
                                                            vacidad/mas-info/> .
:Mobilcom-Debitel a gdprov:ThirdParty ;
    rdfs:label "mobilcom-debitel" ;
    rdfs:seeAlso <https://www.mobilcom-debitel.de/legal/datenschutz/> .
:Nurofy-As a gdprov:ThirdParty ;
    rdfs:label "NUROFY AS" ;
    rdfs:seeAlso <https://nurofy.com/privacy-policy/> .
:Flywheel a gdprov:ThirdParty ;
    rdfs:label "FLYWHEEL" ;
    rdfs:seeAlso <https://www.flywheel.jp/privacy-policy/> .
:Data2Decisions a gdprov:ThirdParty ;
    rdfs:label "Data2Decisions" ;
    rdfs:seeAlso <http://data2decisions.com/privacy-and-cookie-policy/> .
```

Consent was being asked for ~1300 third parties

# Consent requirements from GDPR



| GDPR | Constraint |
|---|---|
| A4-11 | Consent must be associated with only one Data Subject |
| R32,A4-11 | Consent must have one or more categories or types of personal data associated with it |
| R32,R42 | Consent must have one or more purposes a |
| R32,A4-11 | Consent must have one or more processing |
| A7-3 | Consent must have one and only one state |
| A7-2 | Consent is given by exactly one Person |
| | Given consent must have information on h |
| | Consent must have artefacts associated wi |
| | Consent |
| | Consent |
| | Consent |
| R32,A7-2 | Consent |
| | Consent |
| | Purpose role play |
| | If data i be stored |
| | Storage |
| R71,A9-2c,A22-2 | Automated processing of personal data in |
| R111,A49-1a | Data transfer to third country or internati specify identity of recipient |
| R51,A8-2a | Personal data belonging to a special categ indicated |

show/hide Table 3: Qualitative constraints on given consent

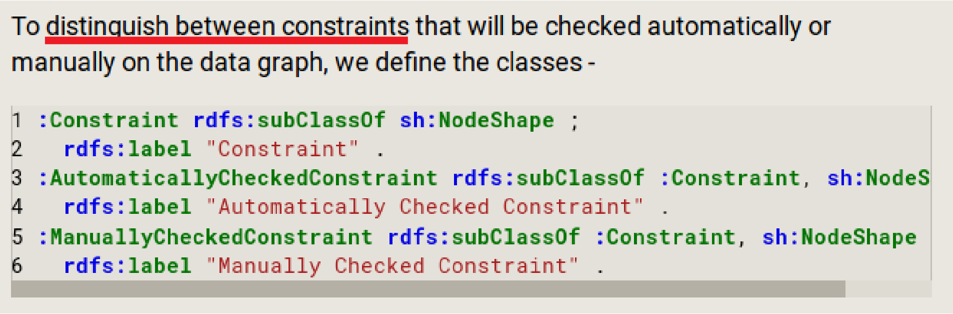
| Criteria | GDPR |
|---|---|
| Consent should be by choice | |
| Consent should have statement of clear action | A4-11 |
| Consent should be freely given | A4-11 |
| Consent should be specific | A4-11 |
| | A4-11 |
| | A7-3 |
| ded before giving | A7-3 |
| | R32 |
| | R32 |
| | R32 |
| | R32 |
| | R32 |
| Consent should have a non-disruptive request | R32 |
| Consent should have separation of processing | R43 |

show/hide Table 2: Constraints and Assumptions for Given Consent

| Competency Question | GDPR Ref. | Comment | Type | Assumption/Constraints | Failing Test Cases |
|---|---|---|---|---|---|
| Who is the Data Subject associated with consent? | A4-11 | Data Subject | Constraint | Every consent must be associated with only one Data Subject | Consent is not associated with any Data Subject |
| | | | | | Consent is associated with more than one Data Subject |
| What are the Personal Data associated with consent? | R32,A4-11 | Personal Data | Constraint | Every consent must have one or more categories or types of personal data associated with it | Consent has no personal data associated with it |

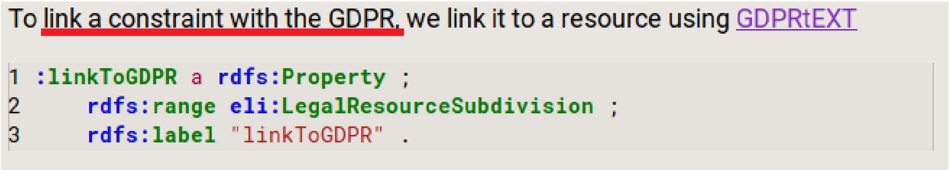
Test-driven approach towards GDPR compliance https://w3id.org/GDPRep/semantic-tests
email: pandith@tcd.ie twitter: @coolharsh55
H. J. Pandit, D. O'Sullivan, D. Lewis. SEMANTiCS 2019, Karlsruhe, Germany.

# SHACL to validate constraints



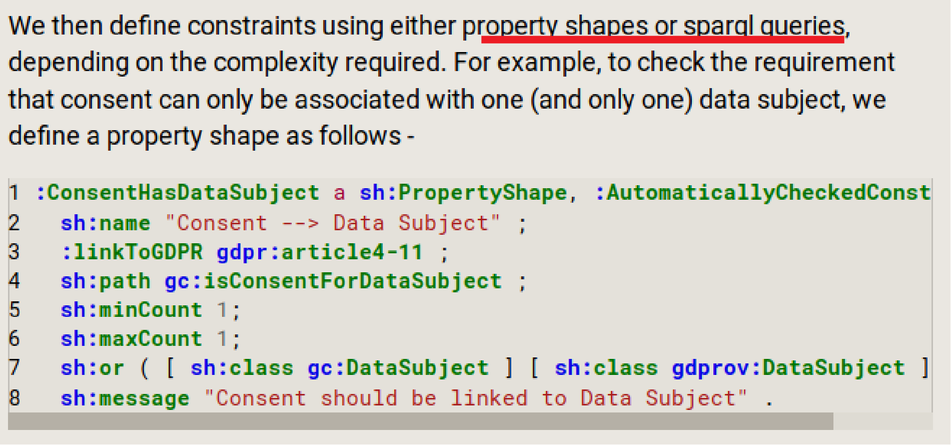
To distinguish between constraints that will be checked automatically or manually on the data graph, we define the classes -

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



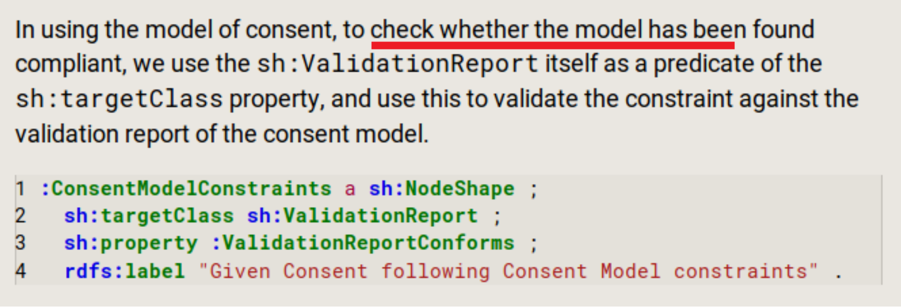
To link a constraint with the GDPR, we link it to a resource using GDPRtEXT

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



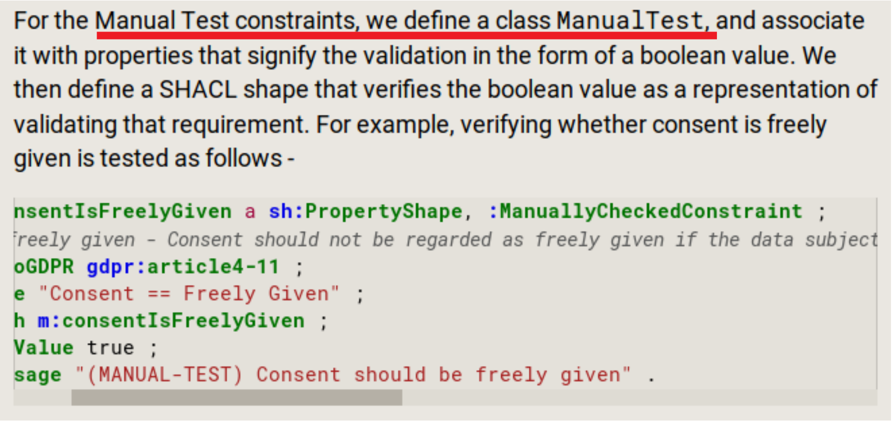
We then define constraints using either property shapes or sparql queries, depending on the complexity required. For example, to check the requirement that consent can only be associated with one (and only one) data subject, we define a property shape as follows -

```
:ConsentHasDataSubject a sh:PropertyShape, :AutomaticallyCheckedConst
  sh:name "Consent --> Data Subject" ;
  :linkToGDPR gdpr:article4-11 ;
  sh:path gc:isConsentForDataSubject ;
  sh:minCount 1;
  sh:maxCount 1;
  sh:or ( [ sh:class gc:DataSubject ] [ sh:class gdprov:DataSubject ]
  sh:message "Consent should be linked to Data Subject" .
```



In using the model of consent, to check whether the model has been found compliant, we use the `sh:ValidationReport` itself as a predicate of the `sh:targetClass` property, and use this to validate the constraint against the validation report of the consent model.

```
:ConsentModelConstraints a sh:NodeShape ;
  sh:targetClass sh:ValidationReport ;
  sh:property :ValidationReportConforms ;
  rdfs:label "Given Consent following Consent Model constraints" .
```



For the Manual Test constraints, we define a class `ManualTest`, and associate it with properties that signify the validation in the form of a boolean value. We then define a SHACL shape that verifies the boolean value as a representation of validating that requirement. For example, verifying whether consent is freely given is tested as follows -

```
nsentIsFreelyGiven a sh:PropertyShape, :ManuallyCheckedConstraint ;
reely given - Consent should not be regarded as freely given if the data subject
oGDPR gdpr:article4-11 ;
e "Consent == Freely Given" ;
h m:consentIsFreelyGiven ;
Value true ;
sage "(MANUAL-TEST) Consent should be freely given" .
```

Test-driven approach towards GDPR compliance https://w3id.org/GDPRep/semantic-tests
email: pandith@tcd.ie twitter: @coolharsh55
H. J. Pandit, D. O'Sullivan, D. Lewis. SEMANTiCS 2019, Karlsruhe, Germany.

# Categories of constraints

We divide the constraints into 3 parts as follows:

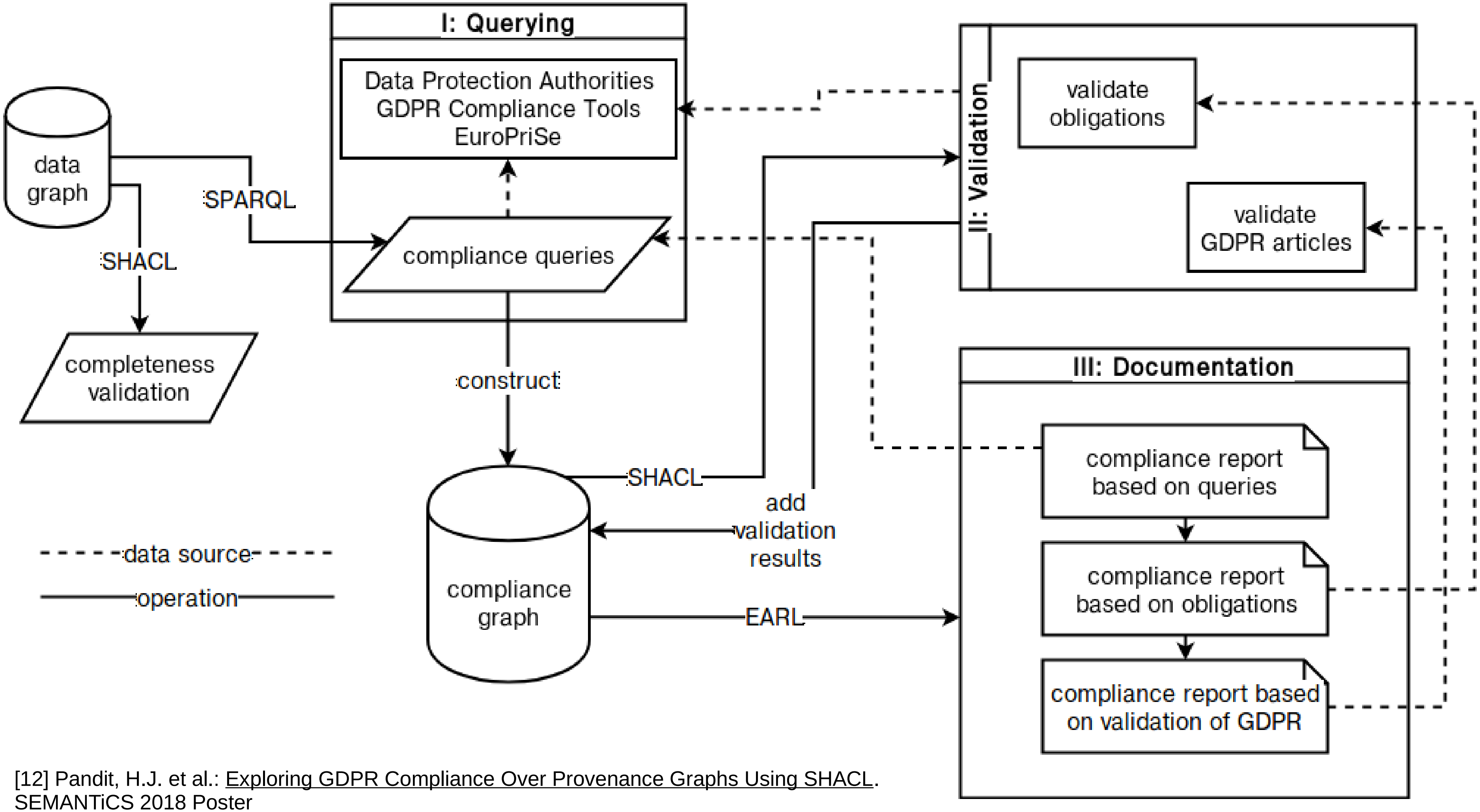
Part A: constraints related to the model of the system
Part B: constraints related to instances of given consent
Part C: common constraints in model + constraints unique for each instances

Part A test requirements such as the presence of DPO and procedures to handle the various rights.

Part B checks requirements directly associated with an instance of given consent. These constraints have to be tested for every instance of given consent.

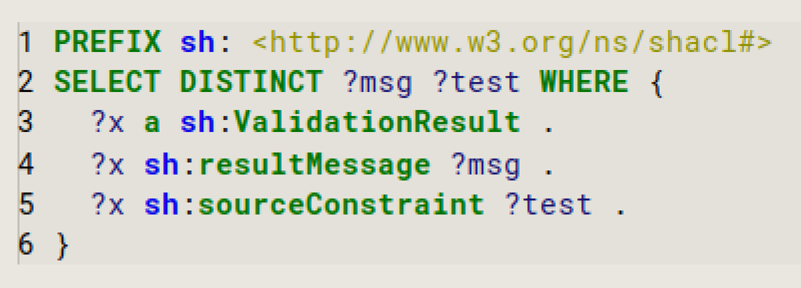
Part C splits the requirements (from Part B) into two parts - one common to all consent and validated against a 'model' or 'template' of consent, and the other validated against the instance of given consent. As most constraints are abstracted away to the model and only need to be checked once, this makes the validation of given consent more efficient.

# Compliance testing [12]



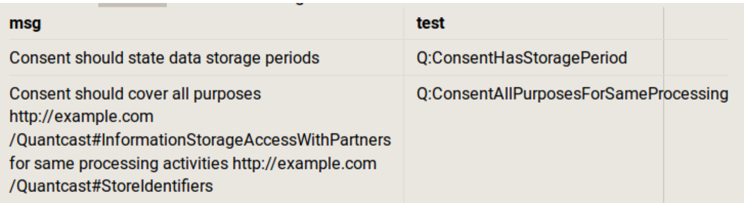
[12] Pandit, H.J. et al.: Exploring GDPR Compliance Over Provenance Graphs Using SHACL. SEMANTiCS 2018 Poster

# SPARQL queries for Compliance results



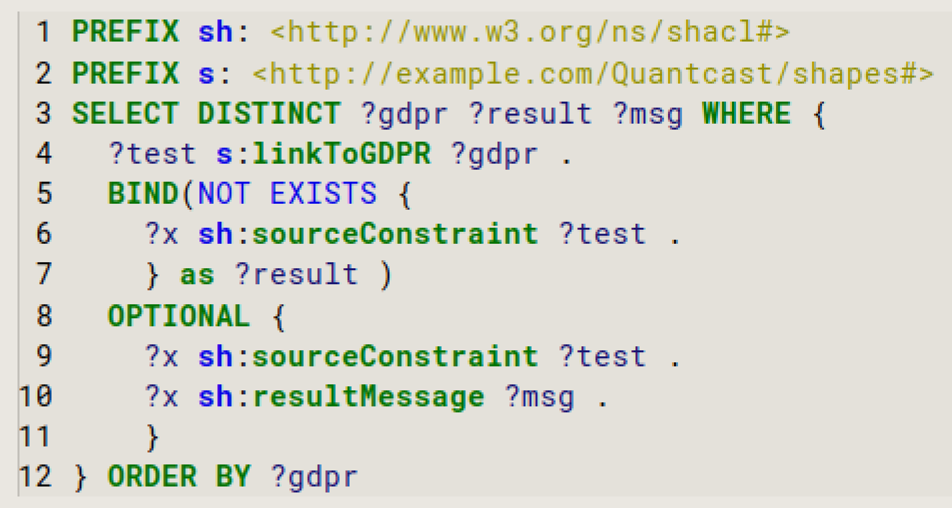
```
PREFIX sh: <http://www.w3.org/ns/shacl#>
SELECT DISTINCT ?msg ?test WHERE {
  ?x a sh:ValidationResult .
  ?x sh:resultMessage ?msg .
  ?x sh:sourceConstraint ?test .
}
```

SPARQL query for retrieving test messages as actionable items



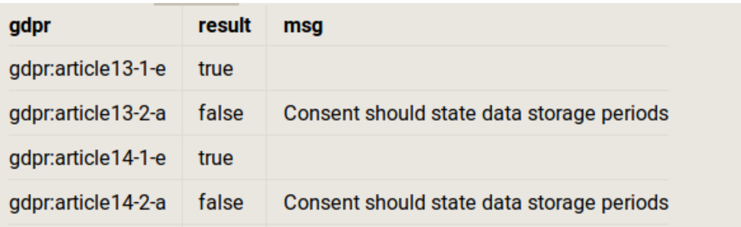
| msg | test |
|---|---|
| Consent should state data storage periods | Q:ConsentHasStoragePeriod |
| Consent should cover all purposes http://example.com /Quantcast#InformationStorageAccessWithPartners for same processing activities http://example.com /Quantcast#StoreIdentifiers | Q:ConsentAllPurposesForSameProcessing |

Messages from tests as actionable items



```
PREFIX sh: <http://www.w3.org/ns/shacl#>
PREFIX s: <http://example.com/Quantcast/shapes#>
SELECT DISTINCT ?gdpr ?result ?msg WHERE {
  ?test s:linkToGDPR ?gdpr .
  BIND(NOT EXISTS {
    ?x sh:sourceConstraint ?test .
    } as ?result )
  OPTIONAL {
    ?x sh:sourceConstraint ?test .
    ?x sh:resultMessage ?msg .
    }
} ORDER BY ?gdpr
```

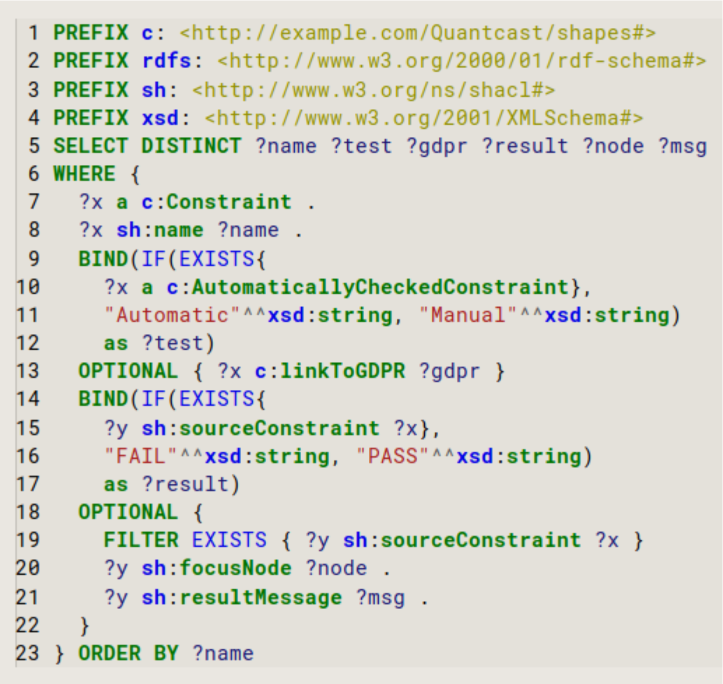
SPARQL query to retrieve GDPR articles and their test results



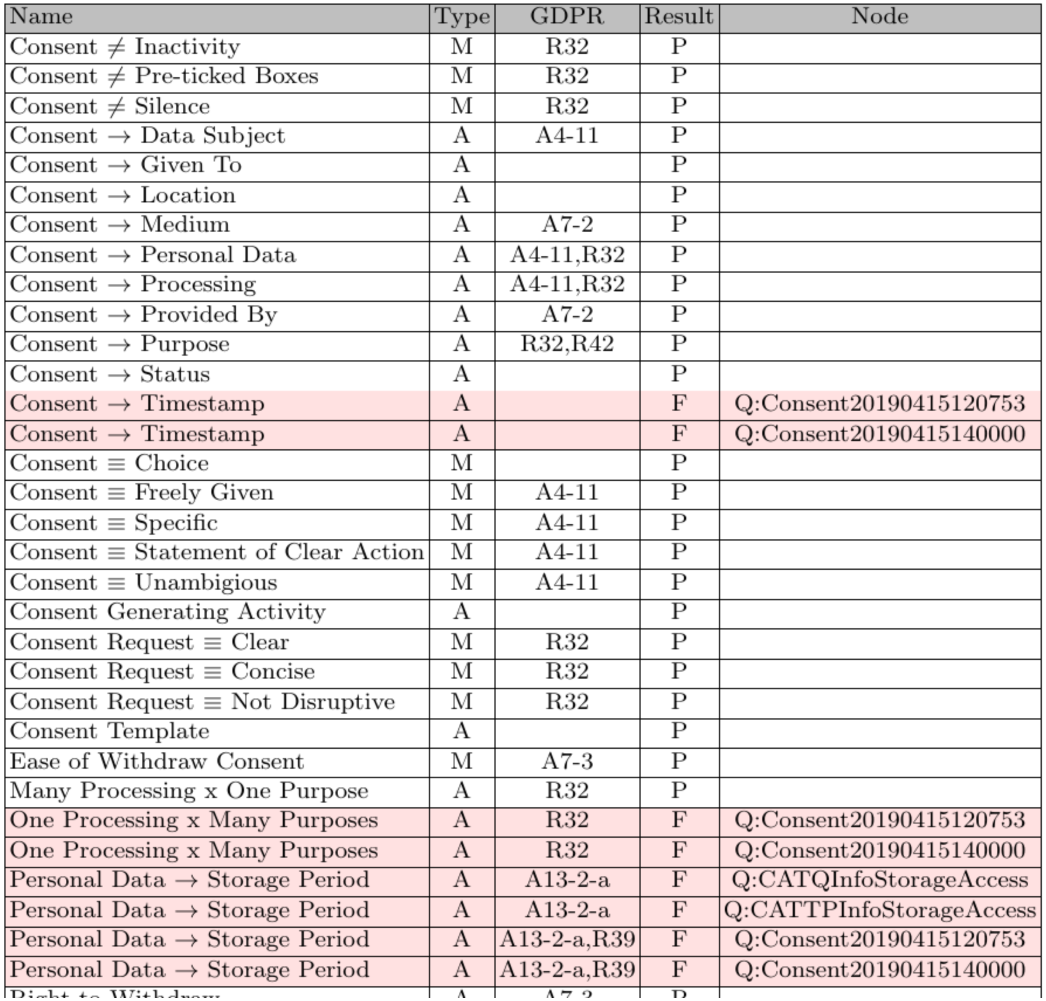
| gdpr | result | msg |
|---|---|---|
| gdpr:article13-1-e | true | |
| gdpr:article13-2-a | false | Consent should state data storage periods |
| gdpr:article14-1-e | true | |
| gdpr:article14-2-a | false | Consent should state data storage periods |

GDPR articles and their test results

# Generating Compliance Report



```
PREFIX c: <http://example.com/Quantcast/shapes#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX sh: <http://www.w3.org/ns/shacl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?name ?test ?gdpr ?result ?node ?msg
WHERE {
  ?x a c:Constraint .
  ?x sh:name ?name .
  BIND(IF(EXISTS{
    ?x a c:AutomaticallyCheckedConstraint},
    "Automatic"^^xsd:string, "Manual"^^xsd:string)
    as ?test)
  OPTIONAL { ?x c:linkToGDPR ?gdpr }
  BIND(IF(EXISTS{
    ?y sh:sourceConstraint ?x},
    "FAIL"^^xsd:string, "PASS"^^xsd:string)
    as ?result)
  OPTIONAL {
    FILTER EXISTS { ?y sh:sourceConstraint ?x }
    ?y sh:focusNode ?node .
    ?y sh:resultMessage ?msg .
  }
} ORDER BY ?name
```

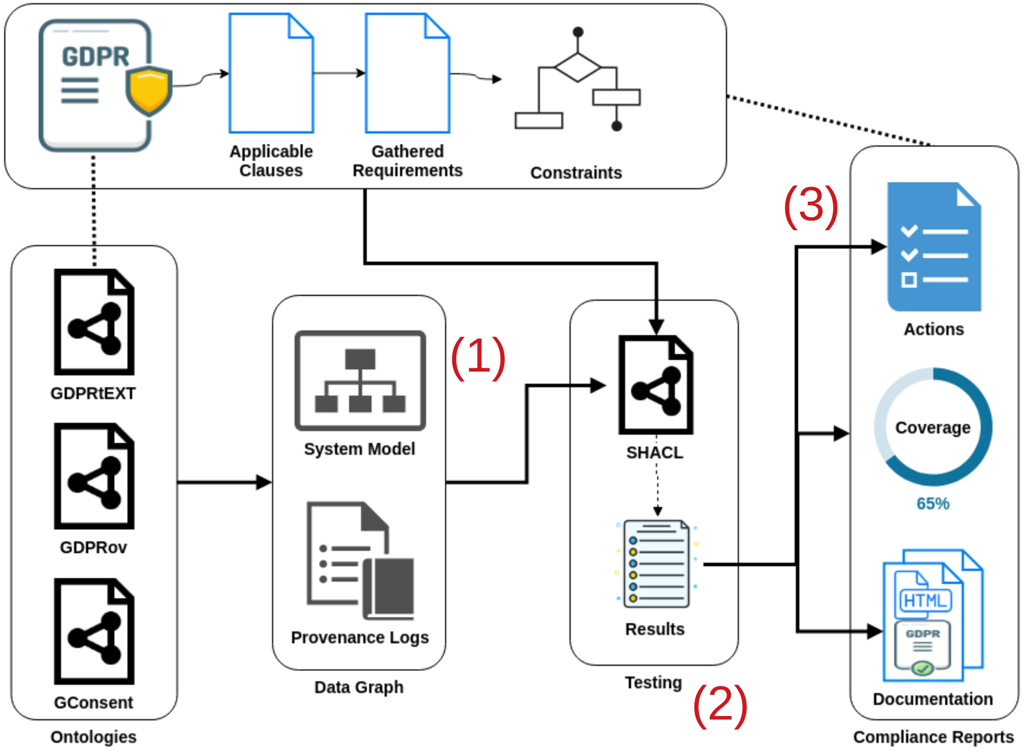


| Name | Type | GDPR | Result | Node |
|---|---|---|---|---|
| Consent ≠ Inactivity | M | R32 | P | |
| Consent ≠ Pre-ticked Boxes | M | R32 | P | |
| Consent ≠ Silence | M | R32 | P | |
| Consent → Data Subject | A | A4-11 | P | |
| Consent → Given To | A | | P | |
| Consent → Location | A | | P | |
| 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 ≡ Unambigious | M | A4-11 | P | |
| Consent Generating Activity | A | | P | |
| Consent Request ≡ Clear | M | R32 | P | |
| Consent Request ≡ Concise | M | R32 | P | |
| Consent Request ≡ Not Disruptive | M | R32 | P | |
| Consent Template | A | | P | |
| Ease of Withdraw Consent | M | A7-3 | P | |
| Many Processing x One Purpose | A | R32 | P | |
| One Processing x Many Purposes | A | R32 | F | Q:Consent20190415120753 |
| 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 | |

SPARQL query and results to generate a report showing constraints, validation results, and link to GDPR

Test-driven approach towards GDPR compliance https://w3id.org/GDPRep/semantic-tests
email: pandith@tcd.ie twitter: @coolharsh55
H. J. Pandit, D. O'Sullivan, D. Lewis. SEMANTiCS 2019, Karlsruhe, Germany.

# Conclusion



(1) Testing a model of a system is more efficient than testing individual instances of processing logs

(2) Persisting results with semantics enables recording and querying compliance information as data

(3) Knowledge can be used to enable systemic information regarding actions for compliance, coverage, and automation in generation of documentation

Engaging Content
Engaging People

# ~ end of presentation ~

## Test-driven approach towards GDPR compliance

**Harshvardhan J. Pandit**, Declan O’Sullivan, Dave Lewis
ADAPT Centre, School of Computer Science & Statistics, Trinity College Dublin, Ireland
email: pandith@tcd.ie
website: http://openscience.adaptcentre.ie/
resources: https://w3id.org/GDPRep/semantic-tests

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

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.