Engaging Content
Engaging People

Towards Generating Policy-compliant Datasets

Christophe Debruyne ◆ Harshvardhan J. Pandit ◆ Dave Lewis ◆ Declan O'Sullivan
ADAPT, Trinity College Dublin, Dublin 2, Ireland

Context and Problem

- Datasets are created and used for a specific purpose, but such data processing is increasingly the subject of various internal and external regulations – e.g., GDPR.
- One particular aspect of GDPR is informed consent, which must by given for these purposes.
- SOTA focuses on compliance analysis of processes; either by analyzing the processes before execution or post-hoc analysis of logs.
- Our hypothesis is that compliance verification can be facilitated by generating datasets "on demand".

Research Question

- Can we generate datasets for a specific purpose "just in time" that complies with informed consent?

Goal

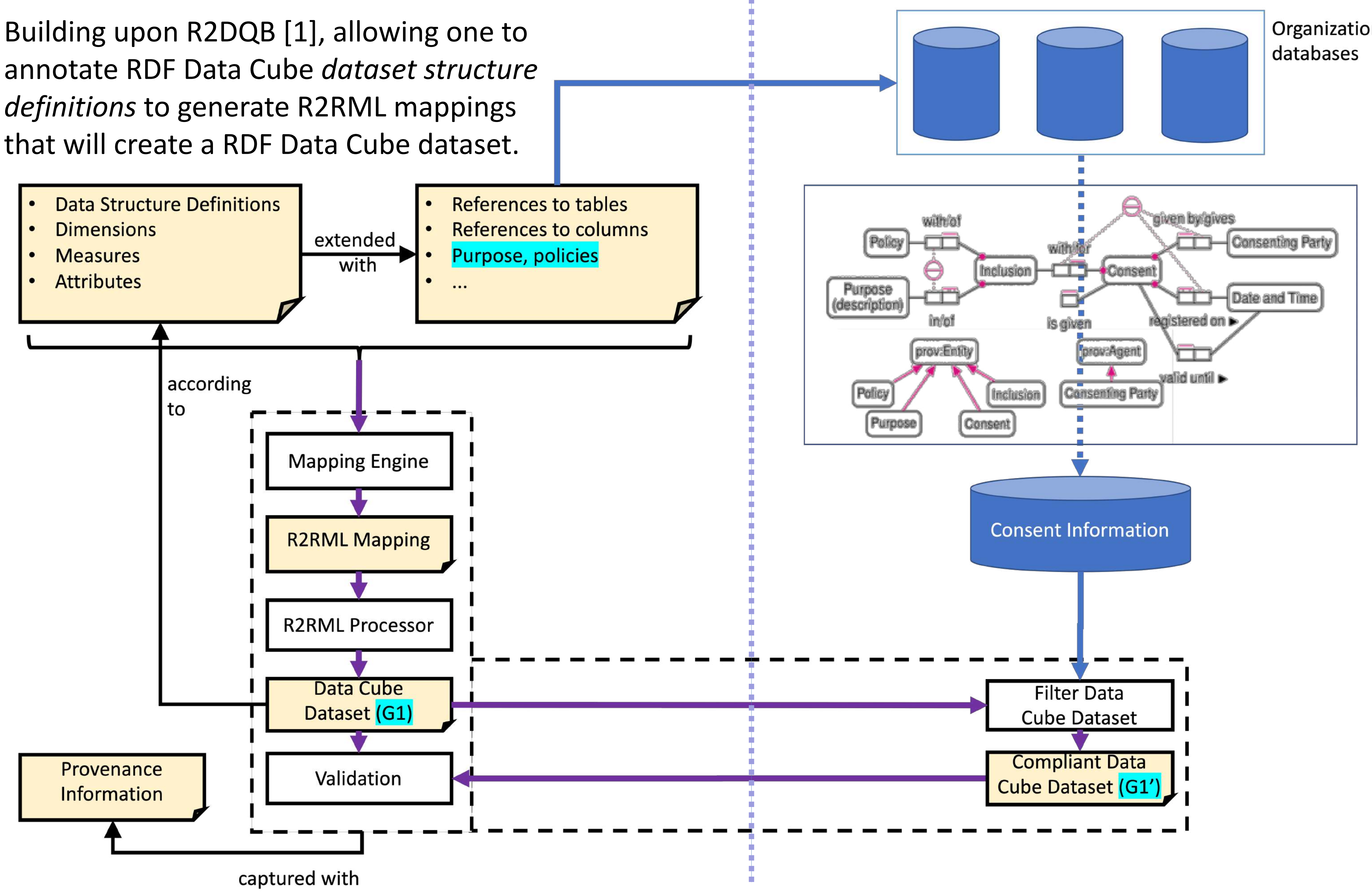
- To propose a method for generating datasets that are fit for a specific purpose and taking into account the ever evolving informed consent of people in a declarative manner, availing of semantic technologies.

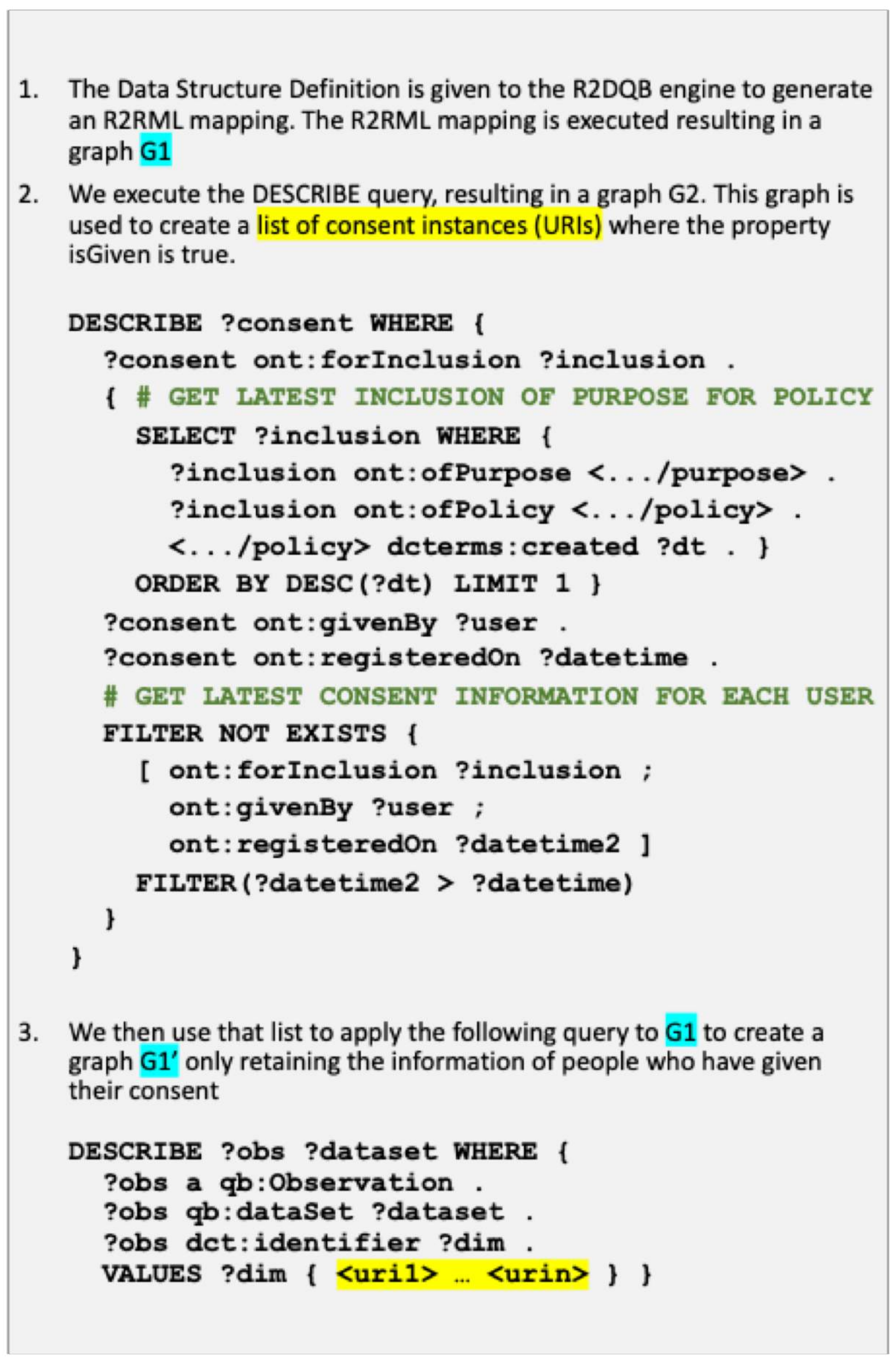
Potential Impact

- *Facilitating* compliance verification as part of data governance best practices within an organization

Approach

Building upon R2DQB [1], allowing one to annotate RDF Data Cube *dataset structure definitions* to generate R2RML mappings that will create a RDF Data Cube dataset.





1. The Data Structure Definition is given to the R2DQB engine to generate an R2RML mapping. The R2RML mapping is executed resulting in a graph G1
2. We execute the DESCRIBE query, resulting in a graph G2. This graph is used to create a list of consent instances (URIs) where the property isGiven is true.

```
DESCRIBE ?consent WHERE {
  ?consent ont:forInclusion ?inclusion .
  { # GET LATEST INCLUSION OF PURPOSE FOR POLICY
    SELECT ?inclusion WHERE {
      ?inclusion ont:ofPurpose <.../purpose> .
      ?inclusion ont:ofPolicy <.../policy> .
      <.../policy> dcterms:created ?dt . }
    ORDER BY DESC(?dt) LIMIT 1 }
  ?consent ont:givenBy ?user .
  ?consent ont:registeredOn ?datetime .
  # GET LATEST CONSENT INFORMATION FOR EACH USER
  FILTER NOT EXISTS {
    [ ont:forInclusion ?inclusion ;
      ont:givenBy ?user ;
      ont:registeredOn ?datetime2 ]
    FILTER(?datetime2 > ?datetime)
  }
}
```

3. We then use that list to apply the following query to G1 to create a graph G1' only retaining the information of people who have given their consent

```
DESCRIBE ?obs ?dataset WHERE {
  ?obs a qb:Observation .
  ?obs qb:dataSet ?dataset .
  ?obs dct:identifier ?dim .
  VALUES ?dim { <uri1> … <urin> } }
```

Demonstration and Results

- We demonstrated the viability of our approach, using a synthetic dataset, though more experiments are called for.
- All intermediate graphs allow one to trace the various steps – traceability and transparency (provenance)

Future Work

- A current limitation is a lack of evaluation beyond the synthetic dataset created for the study.
- We furthermore recognize the opportunities in aligning or integrating our models and approach with related work.

References and Links

1. Christophe Debruyne, Dave Lewis, Declan O'Sullivan: Generating Executable Mappings from RDF Data Cube Data Structure Definitions. OTM Conferences (2) 2018: 333-350

- Ontology: http://openscience.adaptcentre.ie/ontologies/consent-mapping-jit/ontology
- Experiment: https://scss.tcd.ie/~debruync/icsc2019/

See
http://openscience.adaptcentre.ie/
for more of our projects.

Fondúireacht Eolaíochta Éireann
Science Foundation Ireland

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 for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.