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 be 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.

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/cisc2019/

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