GDPRov: provenance for GDPR

Vyacheslav Tykhonov
Senior Information Scientist
Research and Innovation department (DANS)

CLARIAH provenance workshop,
03.09.2018
General Data Protection Regulation (GDPR)

Specific requirements that important for scientific workflows:

Article 4: definition of Personal Data

Article 7: condition of consent for data collections

Article 9: processing of special categories and Personal Data

Article 30: Records of processing activities
Personal data questions

- Where? (storage)
- What? (metadata)
- Who? (owner)
- Why? (purpose)
- When? (date)
- How? (content)
GDPR compliance

You need to implement specific workflows to be GDPR compliant:

• Data management (practice of organizing and maintaining data processes)

• Data governance (overall management of the availability, usability, integrity and security of data)
Maintaining “right to be forgotten”

“The right to erasure or right to be forgotten grants data subjects a possibility to have their personal data deleted if they don’t want them processed anymore and when there is no legitimate reason for a data controller to keep it” (article 17)

Main challenges:

- problematic for highly connected or interlinked data
- Not always possible to execute data erasure (for example, if data stored in blockchain)
The role of consent in GDPR

• Consent is an permission or agreement to keep Personal Data, it can change its status in time

• In the case of change of consent, the previous consent must be archived

• Using provenance metadata, it can be determined which version of terms and conditions was used to get consent from user
Data storage according to GDPR

- Consent must be obtained with intended duration of storage for data, indefinite storage isn’t allowed any more
- Data controller must renew the content to continue storing data
- Provenance metadata must include anonymisation (Amnesia tool as example) and archival activities
GDPRov - GDPR Provenance Ontology

“GDPRov is an OWL2 ontology for representing the provenance of consent & data lifecycles with a view towards GDPR compliance. The ontology aims to provide a high-level abstraction of the interactions of processes over consent & data.”

GDPRov was introduced in 2017 and extends PROV-O standard and Ontology for Provenance and Plans called P-Plan.
P-Plan ontology for activities

- Extends PROV-O to define software workflows or pipelines as plans and link them to their past executions
- Uses “steps” and “variables” to explain how something is happening
- Allows to create templates for execution with defined inputs and outputs
- P-Plan usage helps to track changes in activities along how they interact with consent and data
GDPRov vocabulary hierarchy
PREFIX GDPRov: <http://purl.org/adaptcentre/openscience/ontologies/gdprov#>

SELECT ?consent ?template ?toc
WHERE {
    ?consent a GDPRov:ConsentAgreement .
    ?template a GDPRov:ConsentAgreementTemplate .
    ?toc a GDPRov:TermsAndConditions .
    ?step a GDPRov:ConsentAcquisitionStep .
}

GDProv SPARQL query with consent
Questions?