A Semantic Specification for Data Protection Impact Assessments (DPIA)

Harsh(vardhan J. Pandit)

Research Fellow @ ADAPT Centre, Trinity College Dublin

Email: pandith@tcd.ie | Twitter: @coolharsh55

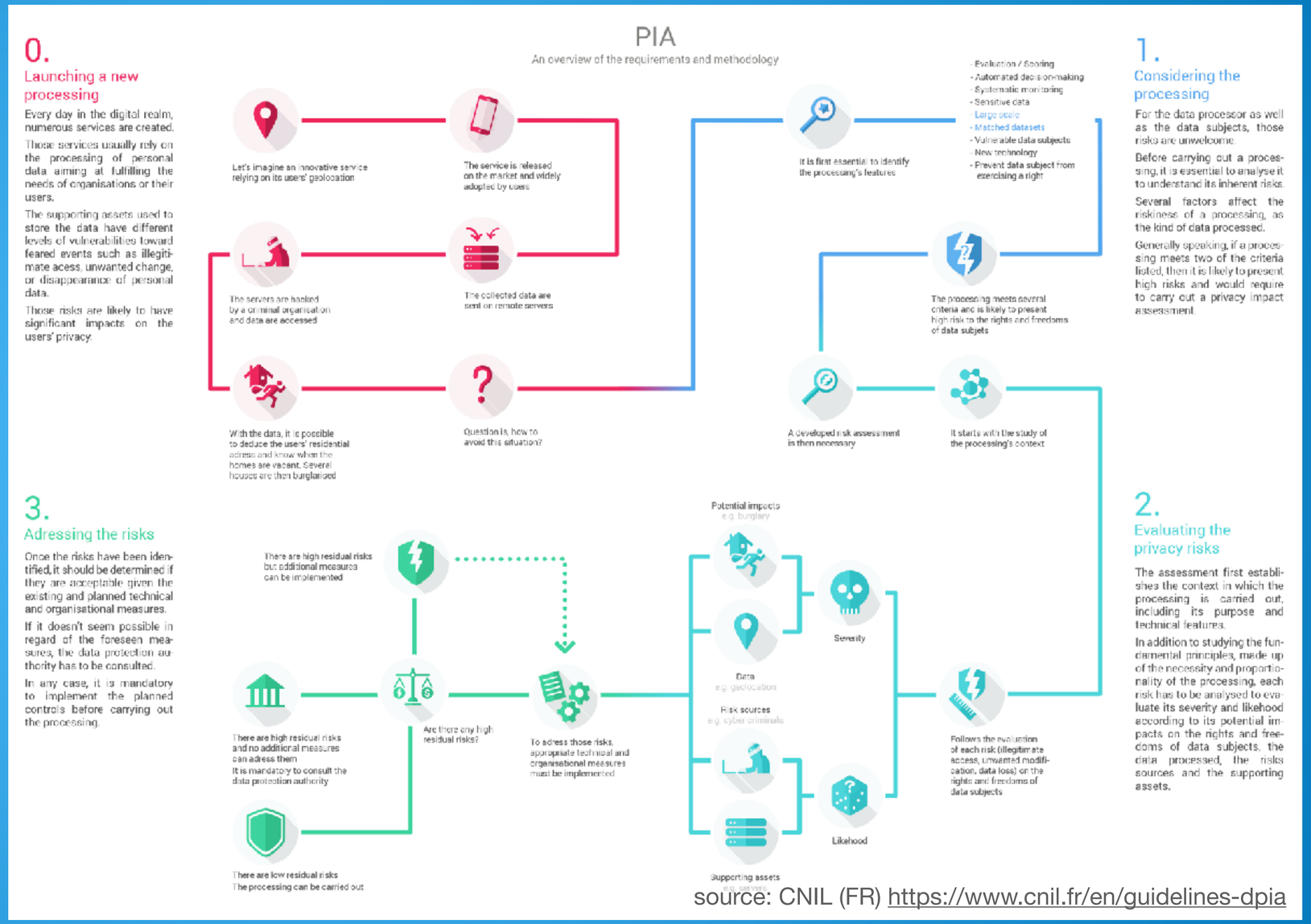
https://harshp.com/research/presentations

Presentation for SEMANTiCS 2022

GDPR & DPIA

1) Art.35 Data Protection Impact Assessments
2) Check if there may be high-risk to data subjects
3) Risk assessed for/to *rights and freedoms*

DPIA 3-step Process
1) Assess need for DPIA
2) Conduct the DPIA
3) Change Processing



source: CNIL (FR) https://www.cnil.fr/en/guidelines-dpia

To conduct a DPIA, you need:

- Details of how personal data is processed
- Risks involved or possible for the operations
- Security and other risk mitigation measures applied
- A risk assessment methodology and process
- Documentation and Records of the Above

unique for use-case

common elements

automate

machine-readable

interoperable format

queryable

verifiable

reuse standards & guidelines

RQ: How to assist with this?

share information

simplify risk assessment

assist with other GDPR compliance

practical application

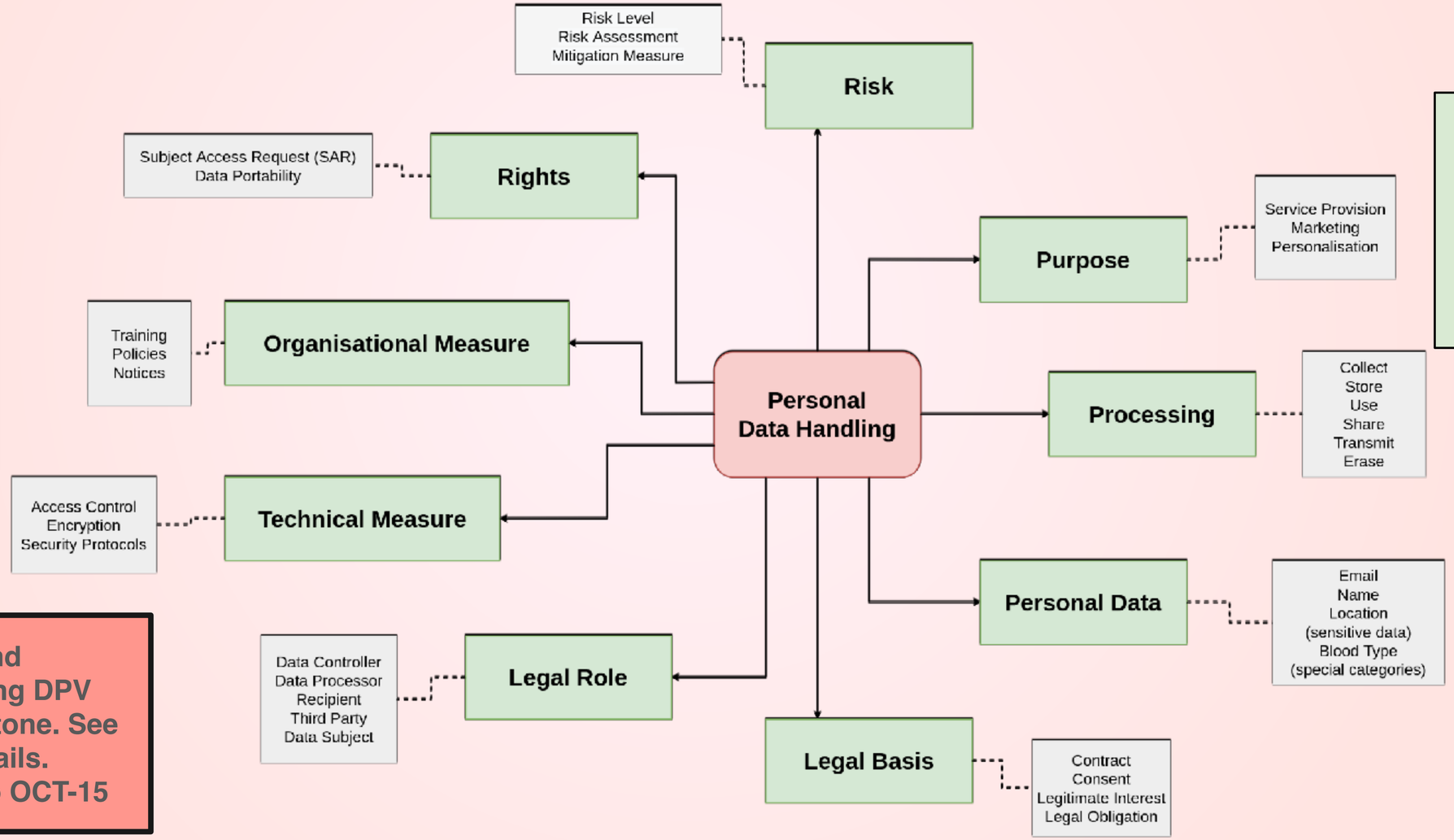
Variance in Knowledge

Guidelines by Data Protection Authorities
GDPR Case Law / Court Arguments and Opinions
Legal Consultancy Reports
DPIAs published in the public domain
ISO 31000 series for Risk Management
FAIR (Factor Analysis of Information Risk)

Data Privacy Vocabulary (DPV)

Description of Personal Data Processing https://w3id.org/dpv

:Taxonomies:
Purpose
Personal Data
Processing
Legal Basis
Legal Role
Tech/Org Measures
Risk
Rights



Serialisations
RDF (canonical)
CSV, JSON

Semantics
RDFS+SKOS
OWL2

We invite comments and feedbacks for publishing DPV v1 - a significant milestone. See DPV spec for more details. Comment period: up to OCT-15

What was missing?

In DPV and State of the Art

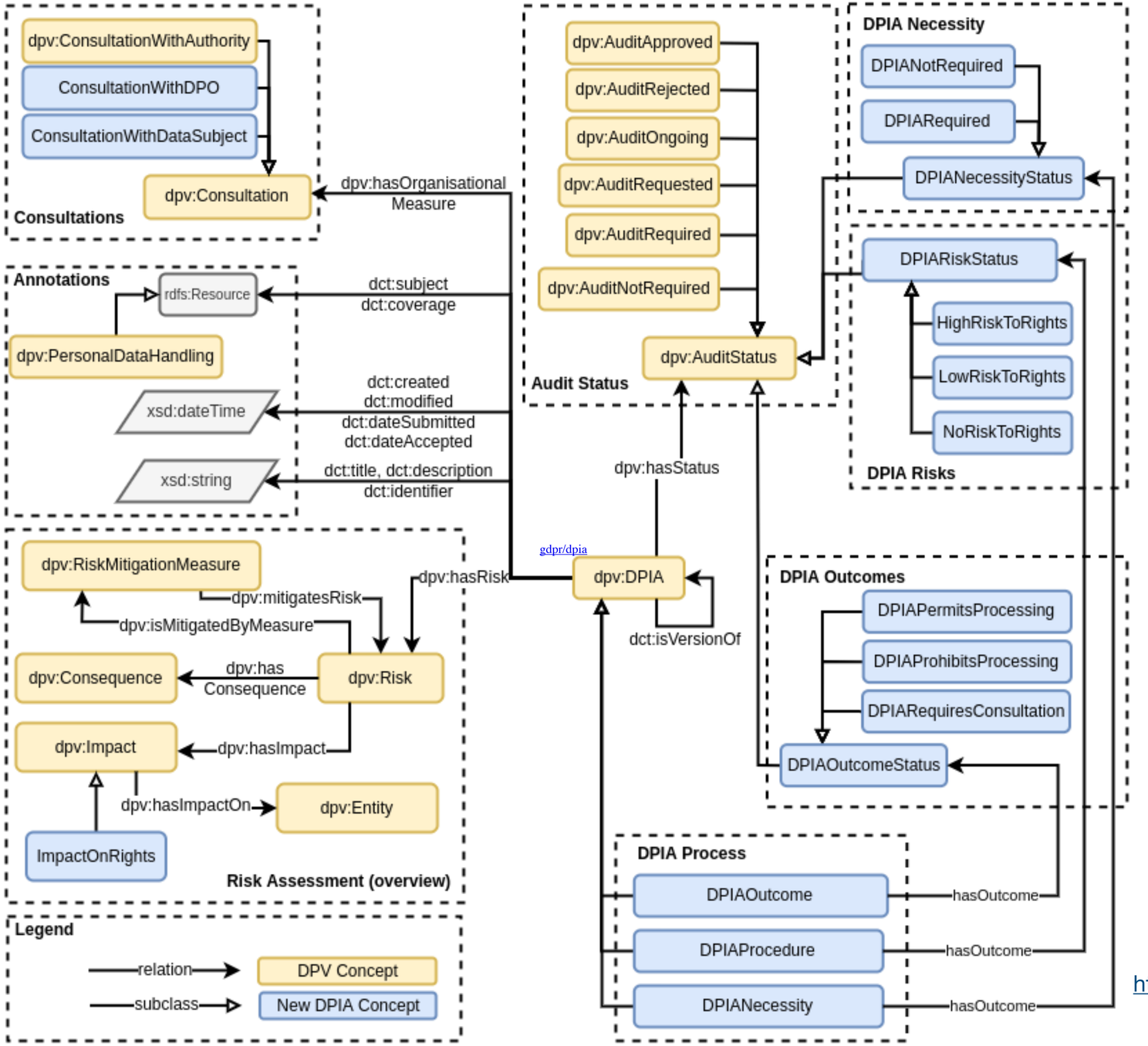
- DPIA considered an isolated process, not part of the larger processing activities and knowledge
- Granularity of DPIA as a process e.g. necessity assessment, risk assessment, changes made to processing
- Recording outcomes of each step
- Concepts required for determining high-risk criteria (e.g. scale of data subjects, volume of data) - and their analysis
- Risk Ontology / Semantic Risk specification

Consultations
- Authority
- DPO
- Data Subject

DPIA as Audit
- Approved
- Not Approved
- Requested
- Rejected
- Required
- Not Required
- Ongoing

Risk
- Mitigation
- Consequence
- Impact
- Level
- Severity

Rights and Freedoms Taxonomy from EU Charter



DPIA Necessity
- Required
- Not Required

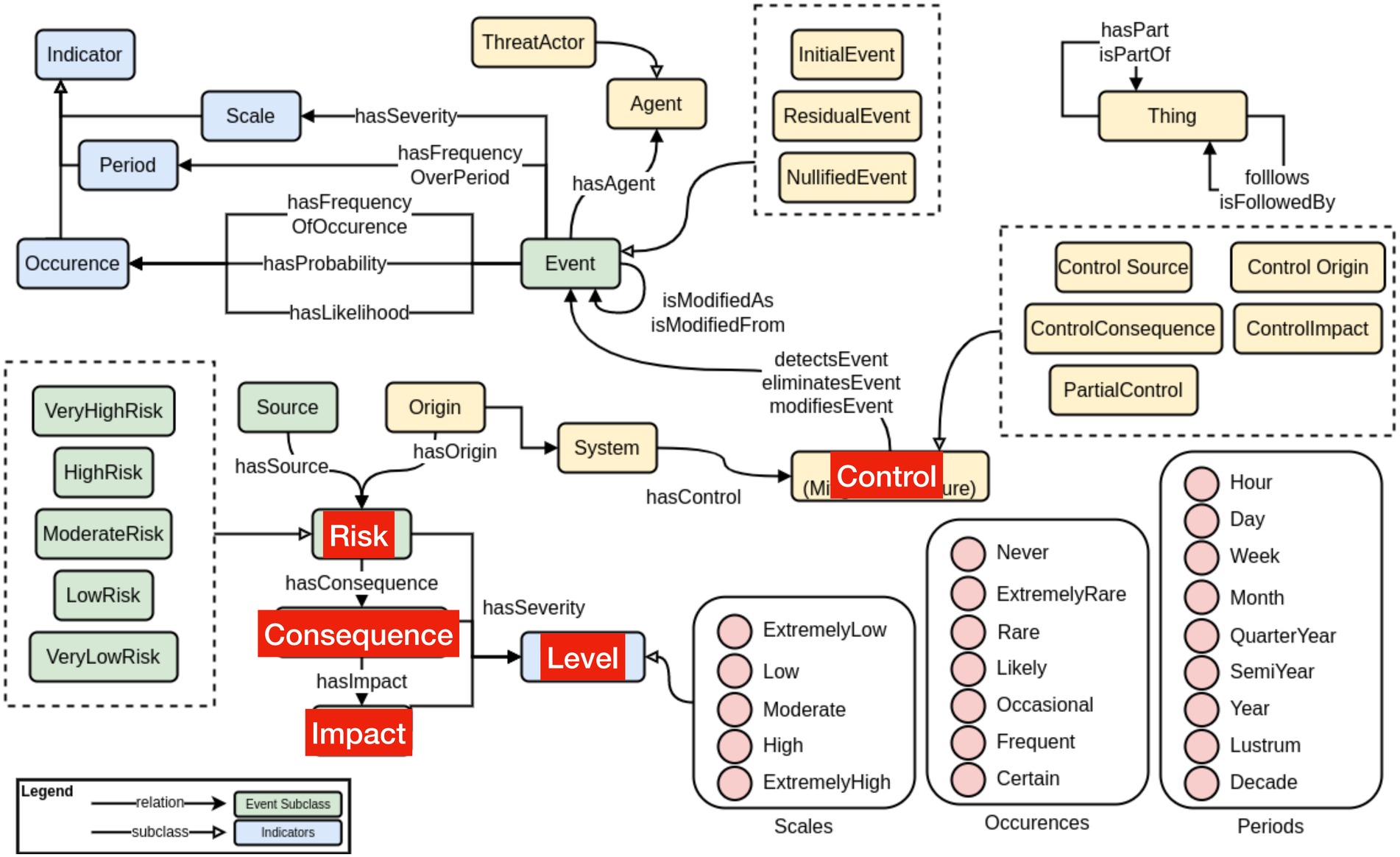
Risk Status
- High Risk
- Low Risk
- No Risk

DPIA Outcomes
- Permit
- Prohibit
- Consultation

DPIA Process
- Necessity
- Procedure
- Outcome

https://w3id.org/dpv/dpv-gdpr/dpia

Risk Ontology based on ISO 31000 series

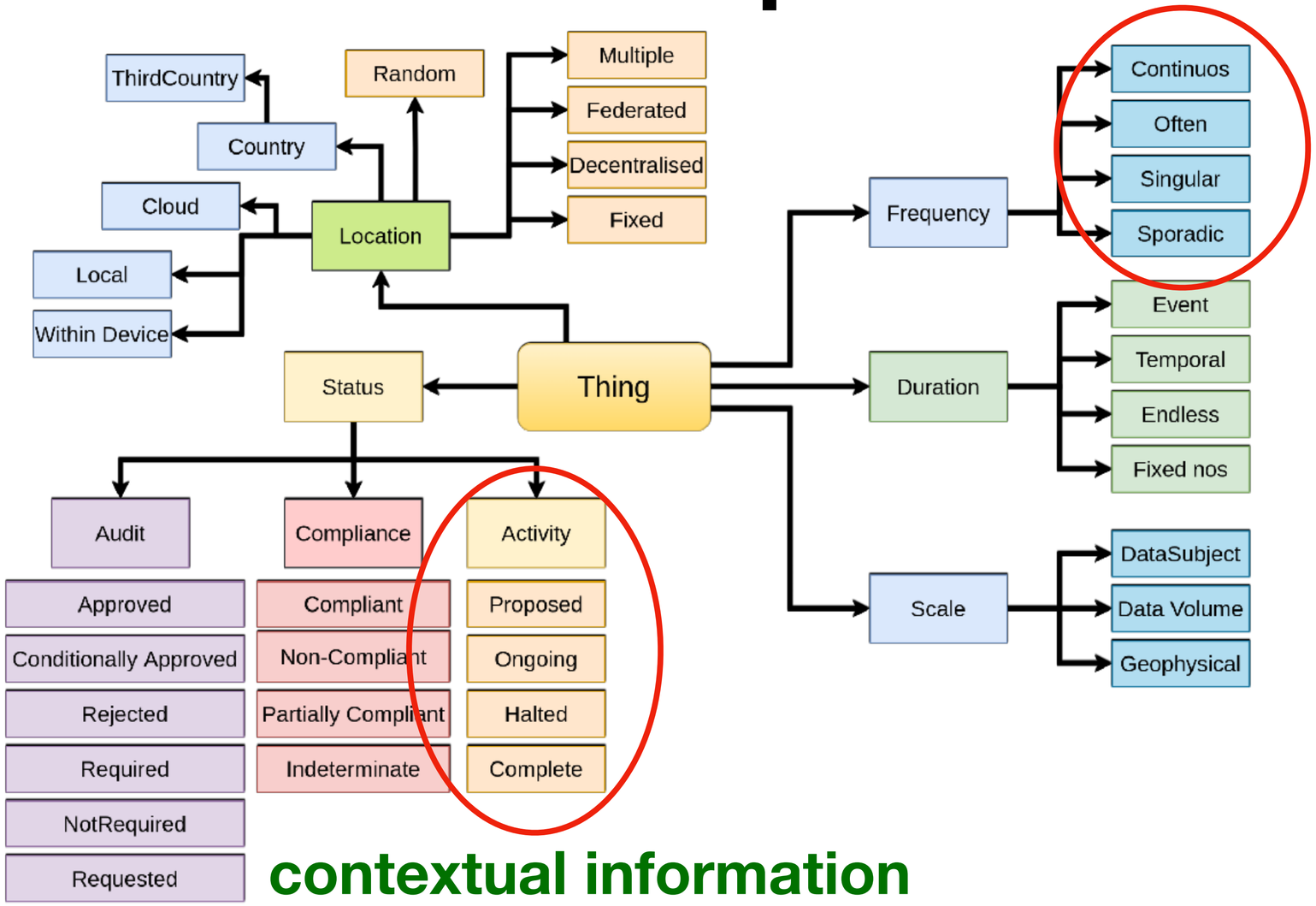


Risk Frameworks are too diverse for consistency

Challenge: How to create an interoperable ontology?

this work only focused on GDPR and DPIA

Developments made in DPV



contextual information
for all kinds of processes and scales

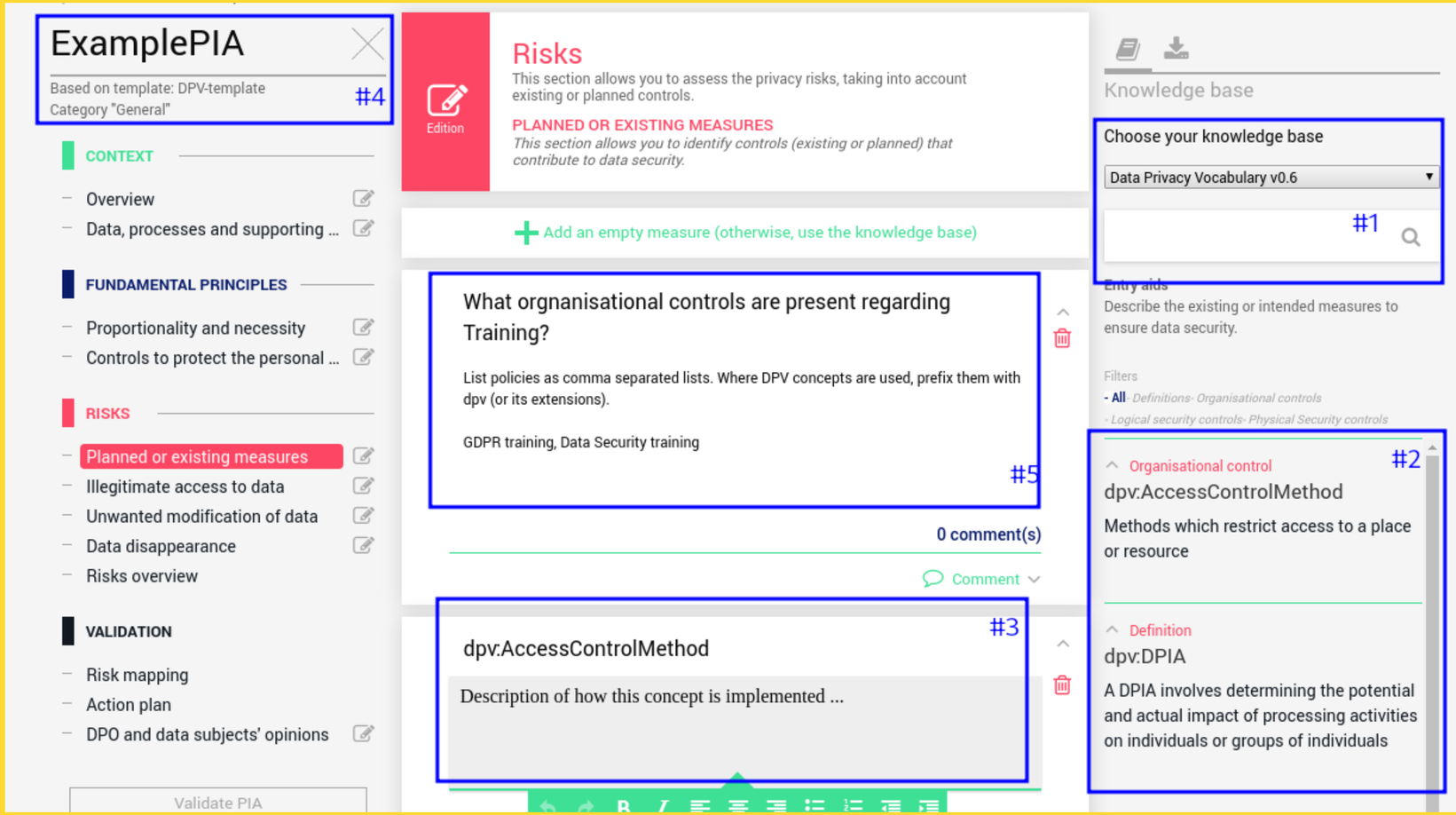
risk extension
https://w3id.org/dpv/risk

- **Risk Likelihood, Severity**
- **Risk Levels**
- **Risk Matrix (3x3, 5x5, 7x7)**
- **Controls**
- **Consequences & Impacts**
- **Assessment methods**
- **Methodologies**

Applications

* CNIL's PIA tool for managing and exporting DPIA data and using "knowledge bases"
* Imported DPV and DPIA spec as KB's and reused (semantic) concepts within the PIA tool

Issues
* Difficulties from lack of documentation regarding KBs
* No Reasoning component so no semantics advantage



CNIL's PIA Software https://www.cnil.fr/en/privacy-impact-assessment-pia

Another application: Model DPIA documents published by SURF about Zoom, Office 365, Google Apps
Issues: lack of structure, no vocabulary for some concepts, complexity of services

practicalities

lack of semantic vocabularies - we need *more metadata*

Next steps: look at ISO standards, specifically the 31K and 27K series, ENISA reports and documents, Germany's Standard Data Protection Model (SDM)

knowledge representation requires domain experts

DPIA information is unstructured, in PDFs (at best), and is *very* difficult to convert into semantic vocabulary

Next steps: enrich DPV concepts to reflect as much of the real-world as possible, undertake systematic analysis of ALL guidelines from EU DPAs to identify requirements, create an "ideal DPIA specification" and align inputs to that

shared impact assessments - DPIA, data transfer, etc.

risk sources, causes, controls, consequences, impacts are duplicated across *every* impact assessments

Next steps: create a common template for "impact assessment" and specialise that for specific tasks — also look towards assessments in future regulations

DPV Specification
https://w3id.org/dpv
Github
https://github.com/w3c/dpv/
Joining DPVCG
https://www.w3.org/community/dpvcg/

interested? questions? contact at:

twitter@coolharsh55
pandith@tcd.ie
dpv@harshp.com
public-dpvcg@w3.org

Data Privacy Vocabulary (DPV)
version 0.8

Draft Community Group Report 26 August 2022

Latest published version:
https://w3id.org/dpv

We invite comments and feedbacks for publishing DPV v1 - a significant milestone. See DPV spec for more details. Comment period: up to OCT-15

A Semantic Specification for Data Protection Impact Assessments (DPIA)

https://w3id.org/dpv/dpv-gdpr/dpia