Harmonising FAIR data sharing with Legal Compliance

Harshvardhan J. Pandit
pandith@tcd.ie | @coolharsh55

FAIRPoints Events | 23 March 2022 | Online/Virtual
Slides available at: https://harshp.com/research/presentations

CC BY NC

CC BY NC

Harsh(vardhan J. Pandit)

An Introduction

https://harshp.com/research

- Postdoctoral Researcher at Trinity College Dublin, IE
- Current Project: creating a knowledge graph of privacy risks for DPIA
- PhD in Computer Science (2020) - Representation of activities involving personal data and consent for GDPR compliance
- Chair of W3C Community Groups: Data Privacy Vocabularies and Controls Community Group (DPVCG) and Consent (ConsentCG)

https://www.w3.org/community/dpvcg/

https://www.w3.org/community/dpvcg/

CC BY NC

F: Findable
A: Accessible
I: Interoperable
R: Reusable

Harmonising FAIR data sharing with Legal Compliance

CC BY NC

Harmonising FAIR data sharing with Legal Compliance

CC BY NC

F: Findable
A: Accessible
I: Interoperable
R: Reusable

Harmonising FAIR data sharing with Legal Compliance

- Copyright Law
- Privacy Regulation
- Data Protection Regulation
  - GDPR
- AI Act
- Digital Markets Act
- Data Act
- Digital Services Act

CC BY NC

Legally Compliant FAIR =
Legally Compliant Data Sharing
= Legally Compliant Value

F: Findable
A: Accessible
I: Interoperable
R: Reusable

Harmonising FAIR data sharing with Legal Compliance

- Copyright Law
- Privacy Regulation
- Data Protection Regulation
  - GDPR
- AI Act
- Digital Markets Act
- Data Act
- Digital Services Act

CC BY NC

General Data Protection Regulation[1]

Applies when FAIR data involves or alludes to Processing of Personal Data

1. What is meant by Personal Data ?
2. What is meant by Processing ?
3. How is data is being processed? (what/how/where...)
4. Who is involved? (whose data, processed by whom)
5. How to check processing is following the rules of GDPR?

[1] https://eur-lex.europa.eu/eli/reg/2016/679/oj

CC BY NC

FAIR + GDPR

Where to begin?

- FAIR
  - Data Catalog
  - Actors, Agents, Entities
  - Licenses
- GDPR
  - Controllers, Data Subjects
  - Legal Basis e.g. consent
  - Sensitive, Special Category Data

image: freesvg.org @openclipart

To ensure better “handling” of data, we need better “metadata”

image: pixels.com @rodnae-prod

CC BY NC

Personal Data

Some "definitions" from across the globe

'personal data' means **any information relating to an identified or identifiable natural person** ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

GDPR Art.4(1)

any information that (a) **can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal**

ISO 29100:2011

"Personal information" means information that **identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly**, with a particular consumer or household.

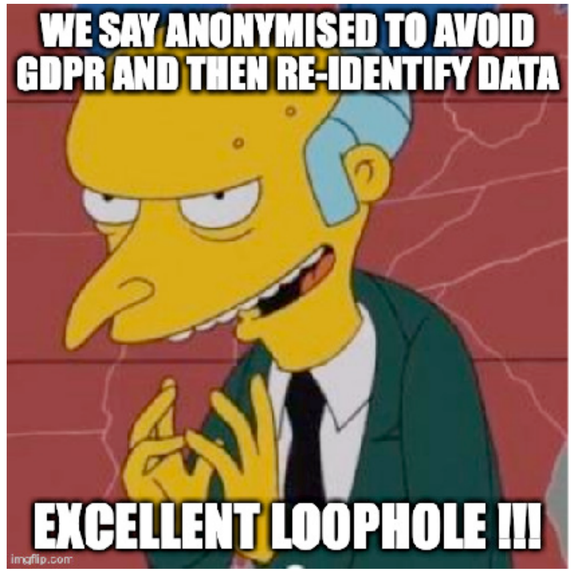
CCPA 1798.140 (o)(1)

CC BY NC

Q: When is Anonymised Data not Anonymised?

Ans: When it is possible to 're-identify' using any (practical) means possible

- Data is anonymised, i.e. all identifiers like names and emails are removed
- But using a 'combination' of remaining data points, a person is still identified
- Since **re-identification** is possible, its not '**fully anonymised**'
- 'Exploits'
  - Aggregated location — person's routines are unique
  - Voting and voters data
  - Fingerprinting - browser configurations, preferences
- GDPR applies to all the above since it is 'personal data'



CC BY NC

GDPR Prohibits

Processing of Special Categories of Personal Data
and
Requires additional obligations via legal basis in Article. 9

racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited

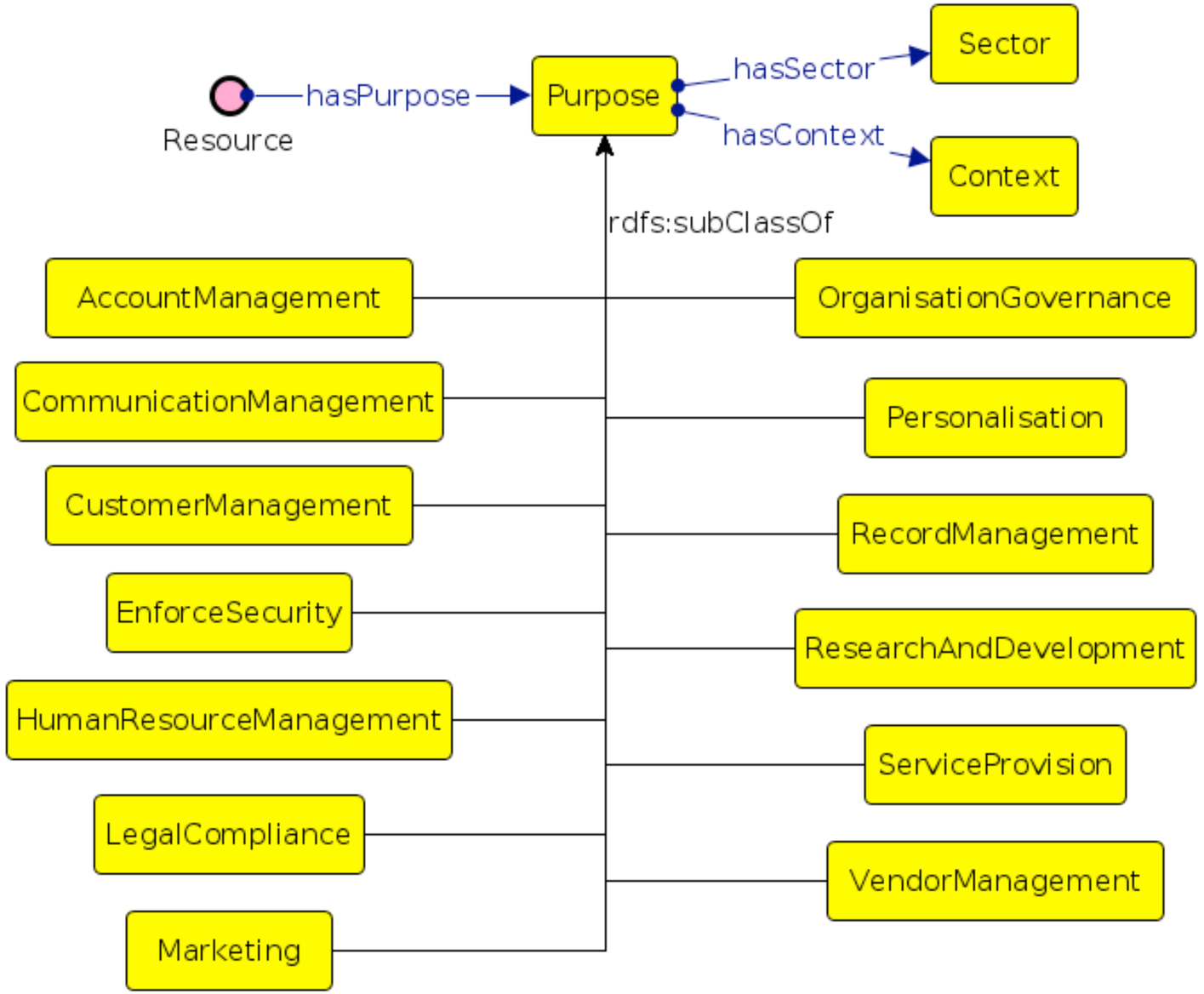
CC BY NC

For FAIR data, the purposes may not be known or wanted to be known (use-as-you-want)

Purposes should be specific and contextual to their use-case

Purposes can be grouped or categorised, but not replaced, e.g. with Marketing for ‘Sending new product emails’

Purposes don’t have to necessarily benefit the data subject e.g. service optimisation



CC BY NC

GDPR's Framework of Legal Basis

A.6(1-b)
Contract

A.6(1-c)
Legal Obligation

A.6(1-e)
Public Interest

A.6(1-d)
Protect vital interests of data subject or other natural person

A.6(1-c)
Official Authority of Controller

A.6(1-a)
Consent

A.6(1-f)
Legitimate Interest of Controller

A.6(1-f)
Legitimate Interest of Third-Party

Widespread Problematic Occurrences

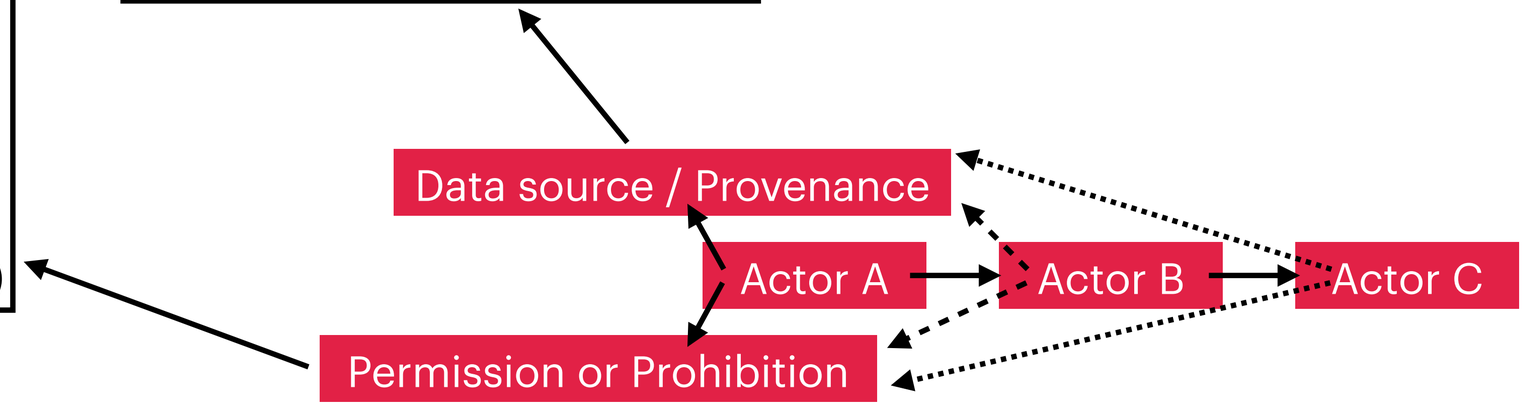
GDPR's principles providing a framework for ‘responsibility'

Principles (Article.5)
lawfulness, fairness and transparency
purpose limitation
data minimisation
accuracy
storage limitation
integrity and confidentiality
accountability

A12-A22 Rights
Transparency (A.12)
Notice (A.13, A.14) ;
Object to Processing
Rectification of Data
Erasure (Right to be Forgotten)
Restriction of Processing
Right of Access
Data Portability

A77 **Right to complaint**
Any Data Subject can complaint to their Supervisory Authority (DPA)
If DPA is in a different country than the company, then the DPA will ‘lease’ and ‘co-operate’ with the DPA of that country

Consent (Article.7)
Informed
Freely Given
Unambiguous
Balance of Power(s)
Right to Withdraw
Explicit Consent (e.g. for Article.9)



CC BY NC

Adding ‘legal metadata’ in FAIR datasets

GDPR

CC BY NC

FAIR datasets + legal metadata = Responsible Data Sharing

Taking lessons from CC-by as a 'globally agreeable license'

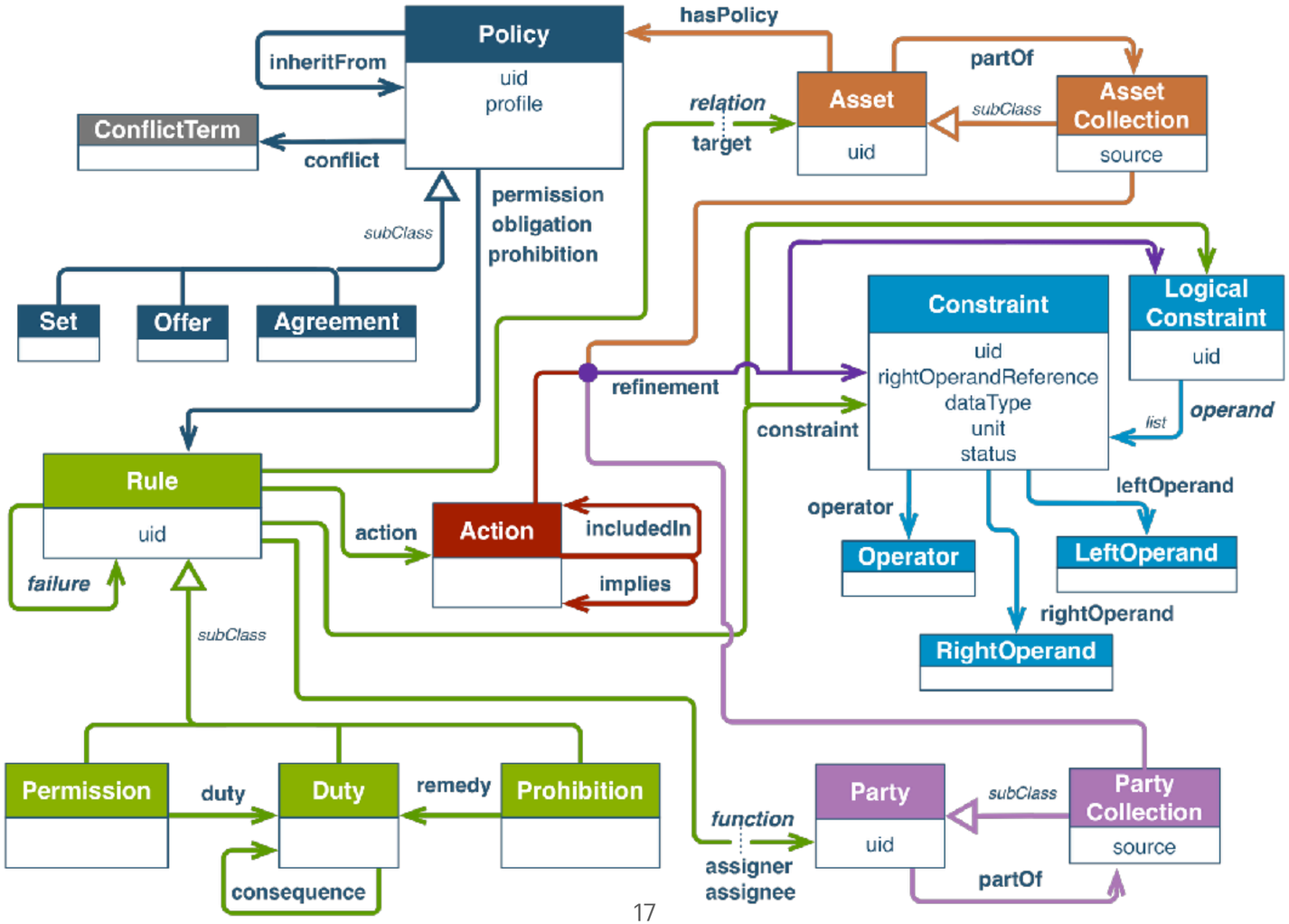
Q: Can we create a CC-by for FAIR datasets containing personal data?
Ans: Yes (with limitations)

- FAIR datasets are shared as 'catalogs' containing metadata (broadly used term)
- Metadata is context, domain, jurisdiction dependant
  e.g. data category or sensitivity, license to use or share, provenance
- CC-by is (nearly) globally recognised - its simple, legal, FAIR-able == popular == more adopted
- So what we want:
  A "commons" vocabulary that harmonises "terms and conditions" for FAIR data,
  and makes it possible to publish, share, and use it within those "policies"
- Use a machine-readable language to declare a policy, then
  '**stick**' the policy to the FAIR dataset and make it part of the license
- E.g. Policy: has sensitive data attributes, you can use it for scientific research OR you are a public institution
- E.g. Policy : data to be used only for X purposes, and shared with others (in EU) who may use it only for X

CC BY NC

Open Digital Rights Language (ODRL)

https://www.w3.org/TR/odrl-model/



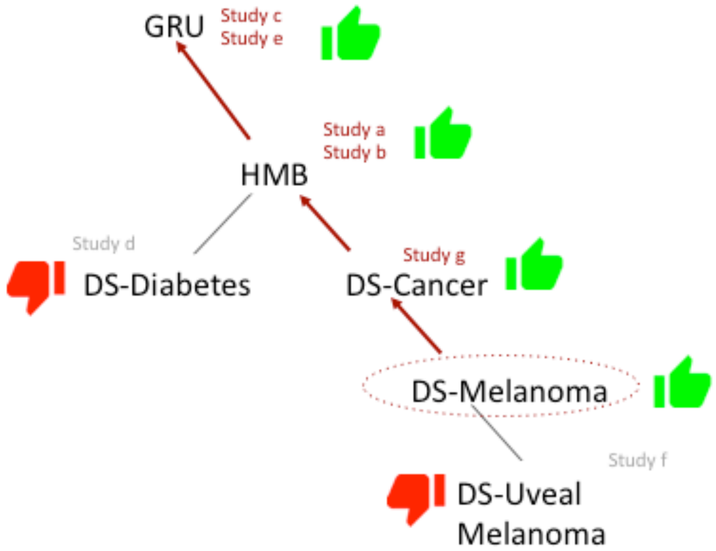
CC BY NC

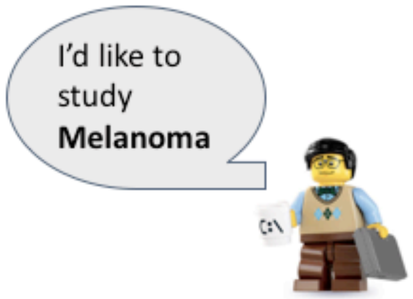
Data Use Ontology (DUO)

https://github.com/EBISPOT/DUO

Data Repository

Dataset	DU restriction
Study a	HMB
Study b	HMB
Study c	GRU
Study d	DS- Diabetes
Study e	GRU
Study f	DS- Uveal Melanoma
Study g	Cancer

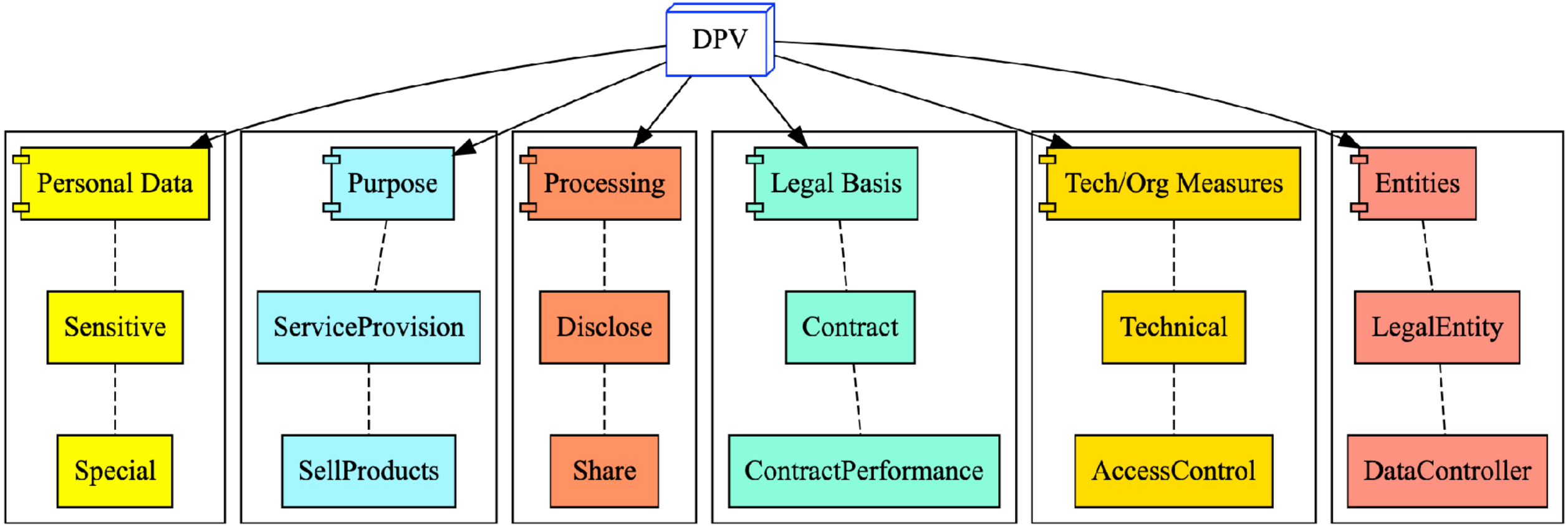




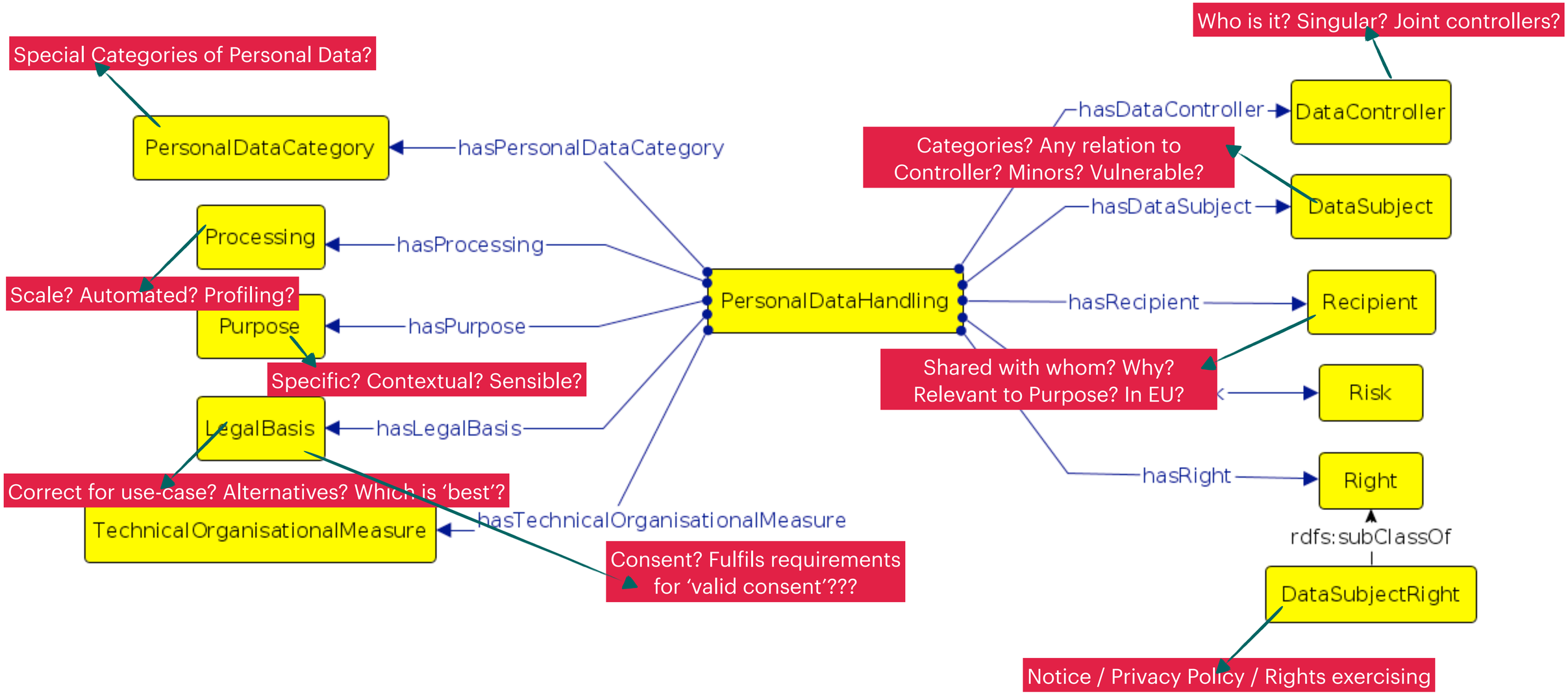
CC BY NC

Data Privacy Vocabulary (DPV)

https://w3id.org/dpv/



Labelling FAIR data for ‘sharing’ in legally compliant manner



CC BY NC

Open Problems, Issues, and Challenges

Declarative Metadata vs Actionable Compliance

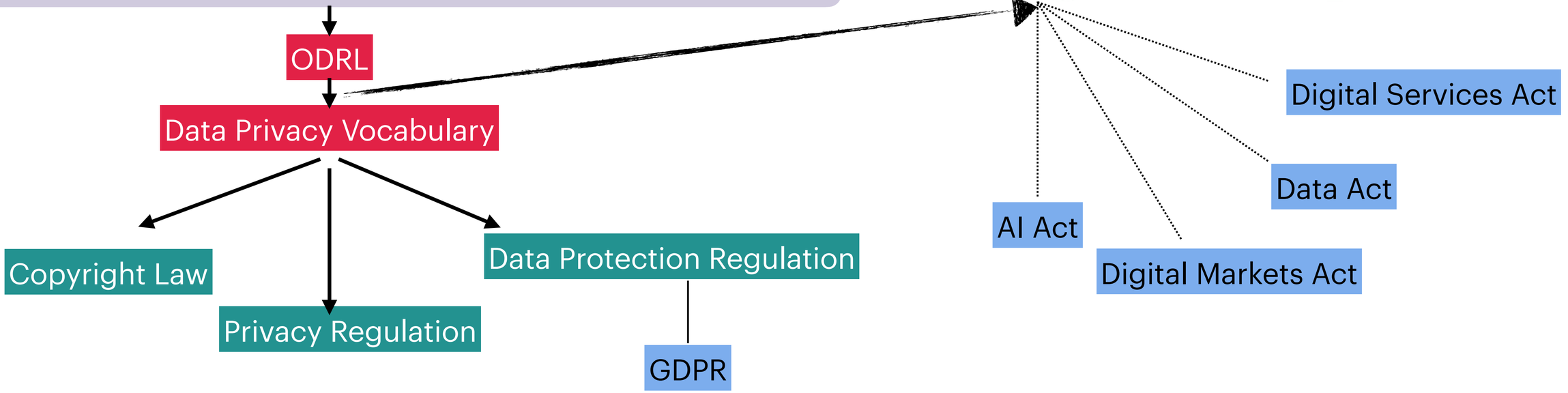
- The FAIR metadata is a ‘promise’, how to ensure the user or adopter respects the terms of that promise?
  e.g. sensitive data being shared for medical research only, but gets used by malicious actors who misuse it
- A possible solution: share only the ‘metadata’ or ‘schema’ or a ‘partial dataset’ that is not problematic on its own, but is enough for a user to determine whether such data is of use to them
  Then they contact the “data holder” to get access to data —> Hospital model

CC BY NC

Legally Compliant FAIR =
Legally Compliant Data Sharing
= Legally Compliant Value

F: Findable
A: Accessible
I: Interoperable
R: Reusable

Harmonising FAIR data sharing with Machine-Readable Metadata for Legal Compliance



Harmonising FAIR data sharing with Legal Compliance

Key take-away:
These are the 'open' challenges
(1) Quantifying Legal Requirements for data sharing for domain / use-case
(2) Aligning Legal Requirements with FAIR workflows :: which laws? Which jurisdictions?
(3) Creating Machine-readable vocabularies for policies :: generic —> specific
(4) Developing new FAIR workflows based on legal actors/policies

Harshvardhan J. Pandit
pandith@tcd.ie | @coolharsh55

FAIRPoints Events | 23 March 2022 | Online/Virtual
Slides available at: https://harshp.com/research/presentations

CC BY NC