Regulating Data

Harshvardhan J. Pandit | email: harshvardhan.pandit@dcu.ie
CA4025 | 06 March 2023 | Dublin City University
Slides available at: https://harshp.com/research/presentations

Harsh(vardhan J. Pandit)

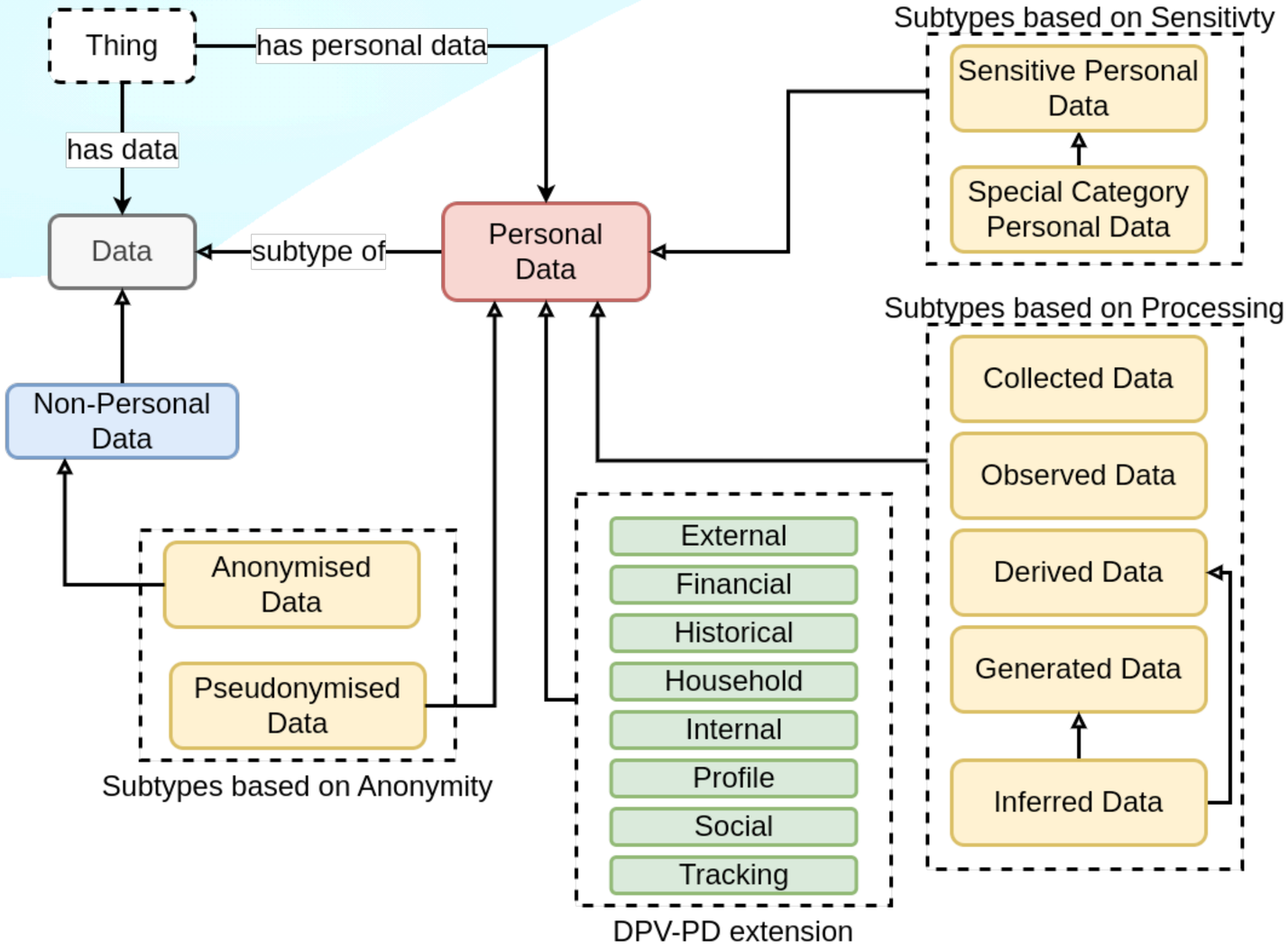
An Introduction

https://harshp.com/research

- Assistant Professor - School of Computing, DCU (2023+)
- Postdoctoral Researcher at Trinity College Dublin (2020-2022)
- 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)
- Member, ISO JTC SC27 **Cybersecurity** and **Privacy Protection**

Data

Personal Data



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)

GDPR[1]

World-Changing EU law that regulates 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

Personal Data

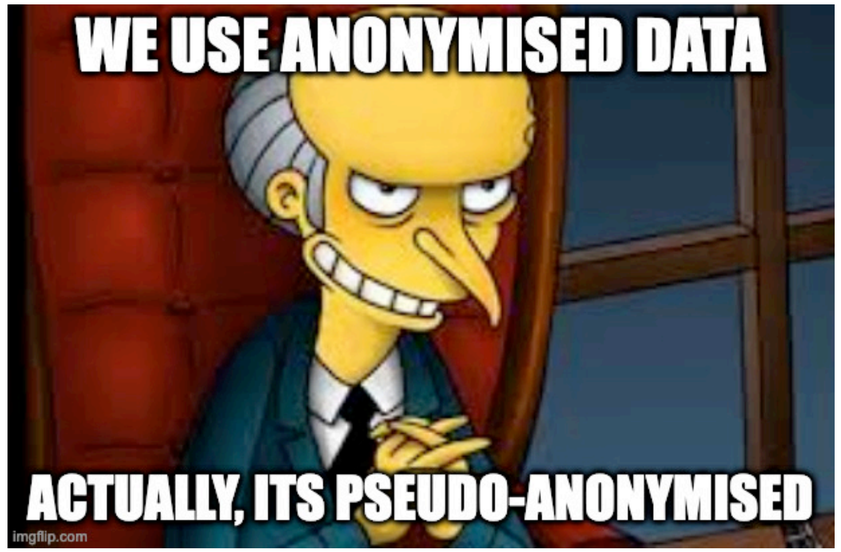
Identifiers, and Identifiability

1. Identifiers: Harsh (name), pandith@tcd.ie (email)
2. Non-identifiers: Black (hair), Brown (eyes), 1.66m (height), etc.
3. For a room full of people, combine non-identifier to uniquely identify a person (me) — thus creating an identifier !!!
4. Useful technique for **fingerprinting**, **profiling**, **tracking**

Q: When is Personal Data not 'Personal' anymore?

Ans: When it is (completely) anonymised

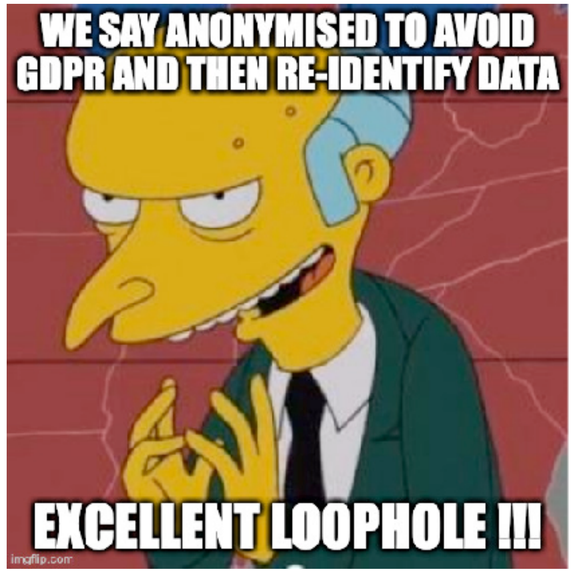
- Anonymisation is the removal of (some) 'identifying' attributes from data
- Merely using "**anonymisation**" does not produce anonymised data
- It produces '**pseudo-anonmised**' data, which is still personal data
- 'Completely anonymised' if it is **not identifiable**
- E.g.
  - Your exact location = personal data
  - approx. house = still personal data
  - approx. area = still personal data, but less
  - City = still personal data, but lesser
  - Country = anonymised, kind of

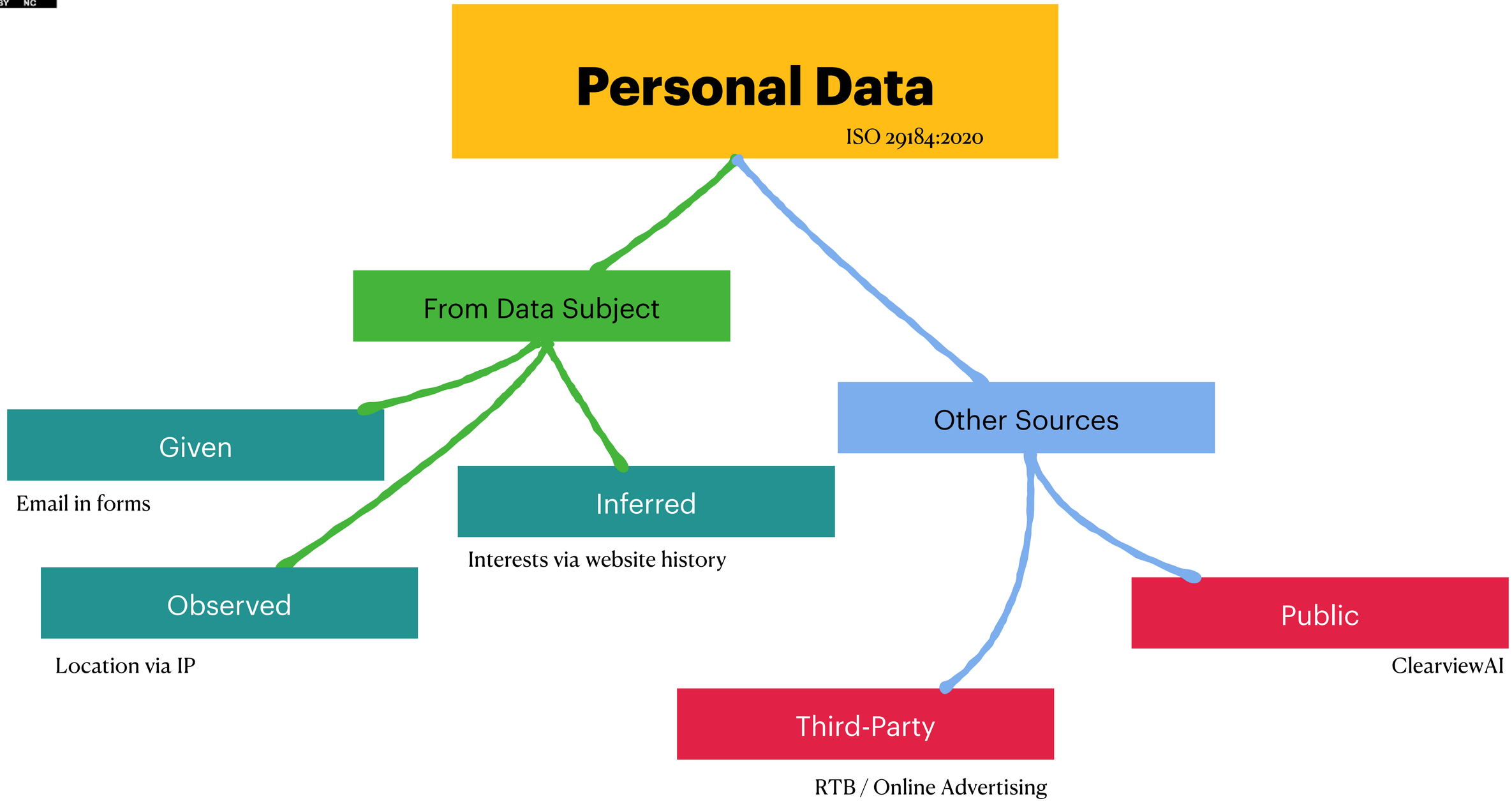


Q: When is Anonymised Data not Anonymised?

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'




Personal Data
ISO 29184:2020
From Data Subject
Other Sources
Given
Email in forms
Inferred
Interests via website history
Observed
Location via IP
Public
ClearviewAI
Third-Party
RTB / Online Advertising

Personal Data: Sensitive, and Special

Special category personal data is to GDPR what Ferrero Rocher is to chocolates

Sensitive:

- data that merits additional security

- older term used widely

Special:

- requires additional/specific legal permissions

- newer term introduced in GDPR

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

Processing Overview

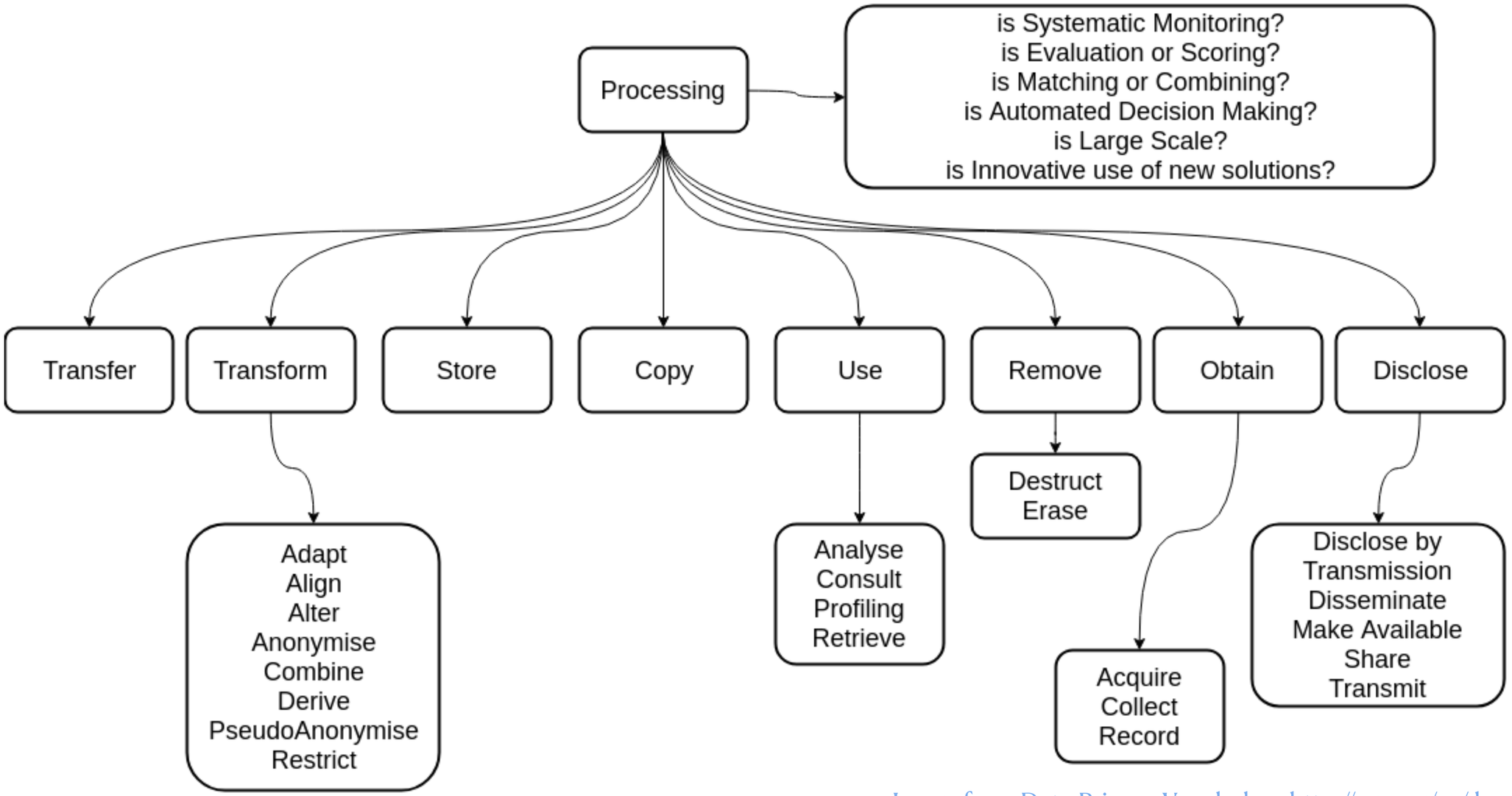


Image from Data Privacy Vocabulary http://w3.org/ns/dpv

GDPR applies before Processing starts

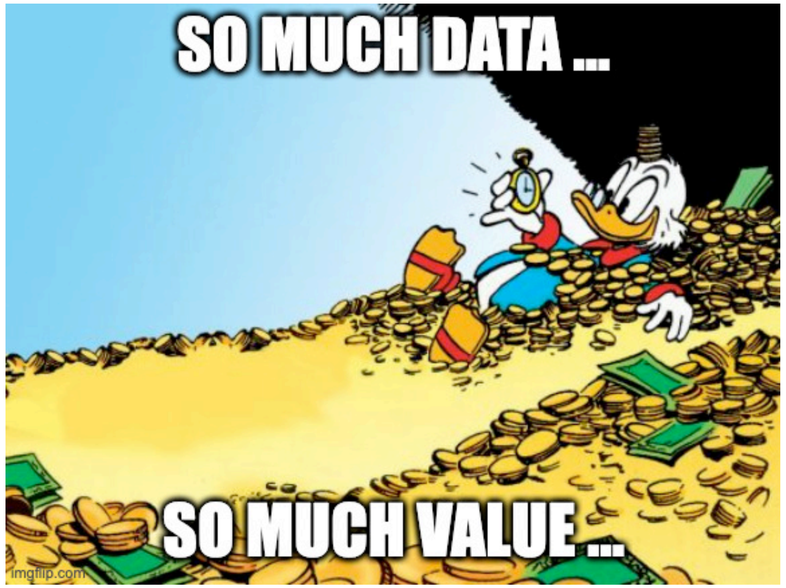
Common Misinterpretations

- Data collected but ‘anonymised’ is not subject to GDPR
- If data isn’t shared, nothing needs to be declared
- Collecting anonymised data and attaching an identifier to it
- Hiding things that require transparency and permission
  - Scale and scope of processing
  - Involvement of special categories
  - Involvement of any automated decision making
  - Creating, sharing, using - profiling

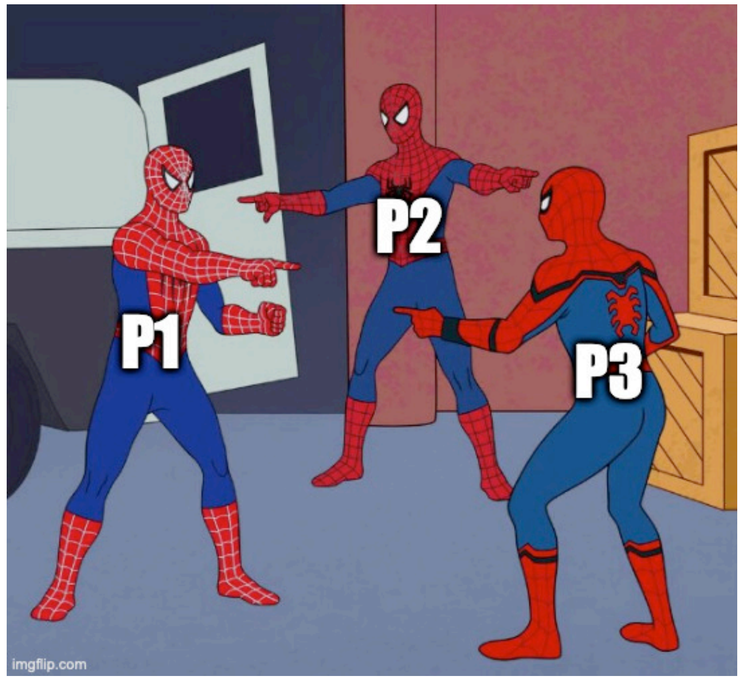
All Processing in GDPR *must* be towards a Goal

Implied when a 'Purpose' is necessary as per Article.5

Every Processing *must* have a Purpose



Purposes must be separate from other matter, including other purposes



Purposes must be *specific* and *unambiguous*

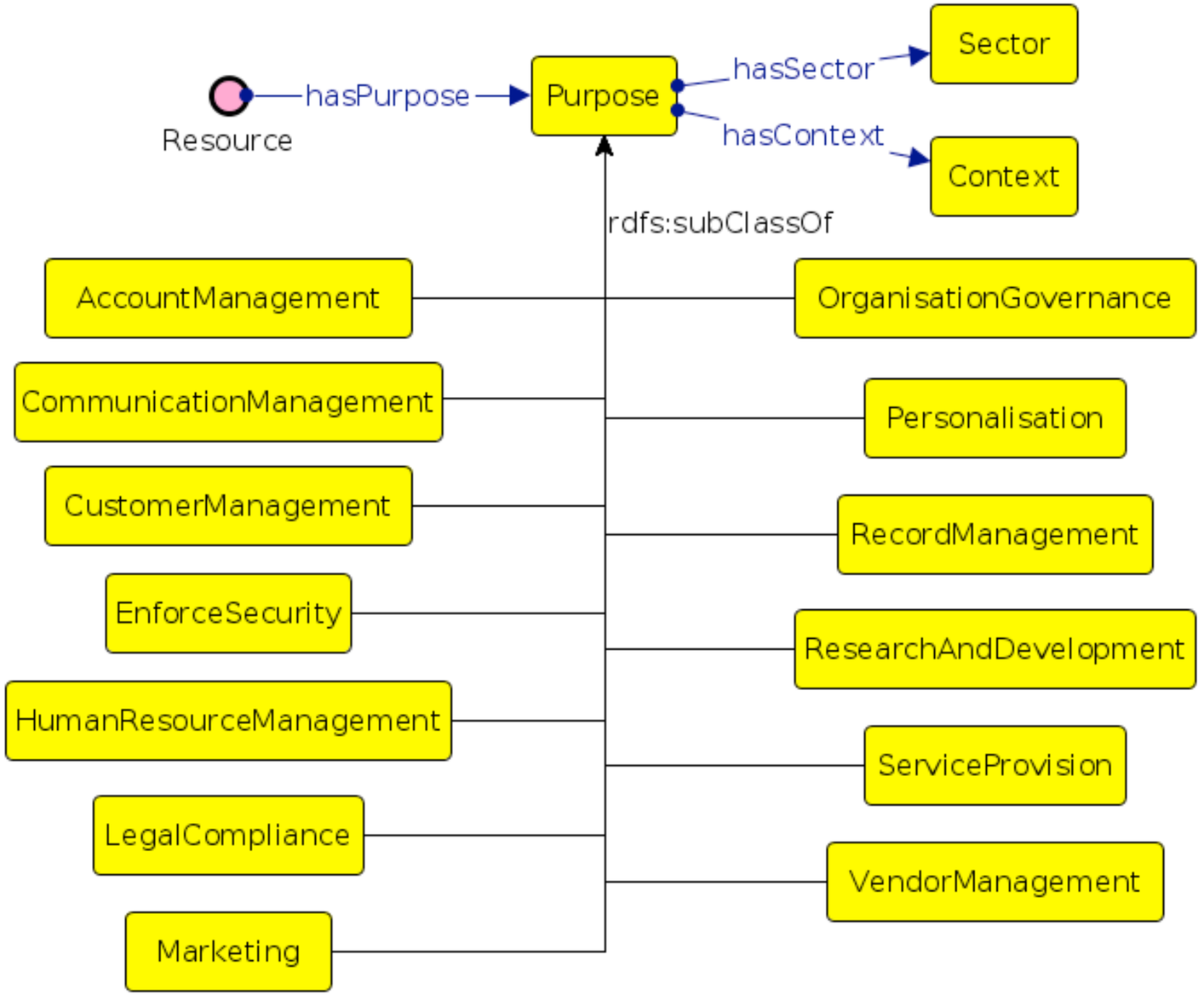
Purposes are intended to be human-readable and human-comprehensible

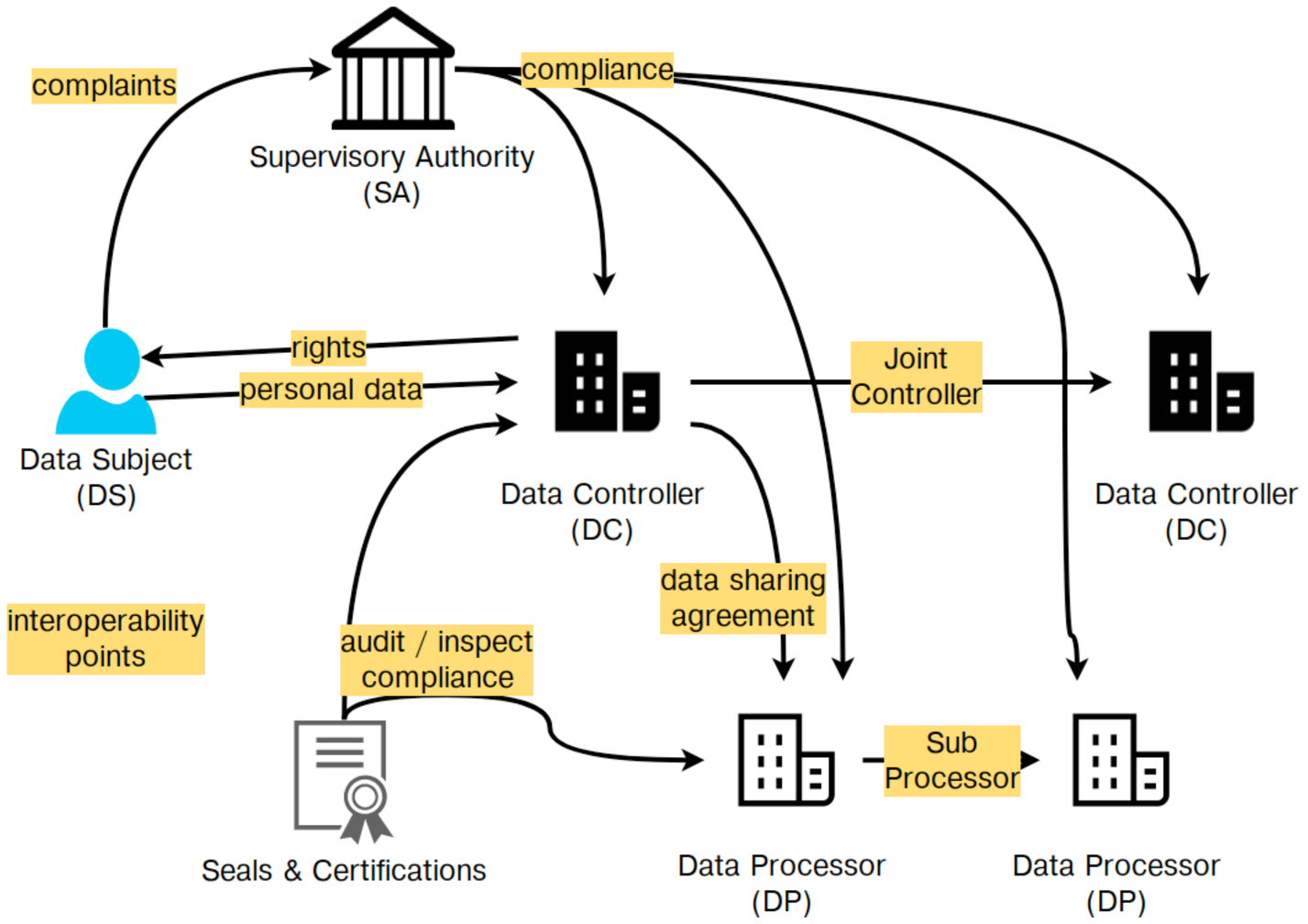
Purposes should not be broad and abstract

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





GDPR Data Interoperability Model,
EURAS Annual Standardisation Conference (EURAS) 2018,
Harshvardhan J. Pandit* , Declan O'Sullivan , Dave Lewis
https://harshp.com/research/publications/010-gdpr-data-interoperability-model

- Data Controllers are responsible for deciding the 'purpose'
- Data Controllers may not even 'touch' the data they 'control'
- Data Controllers can 'team up' to become Joint (Data) Controllers
- Processors only act on 'orders' given (explicitly) by Controllers
- Processors can appoint other (sub-)Processors, still governed by instructions from Controllers
- Processors deciding/ processing on their own become Controllers
- Data Protection Authorities (DPA) are empowered by GDPR to enforce its obligations on all entities

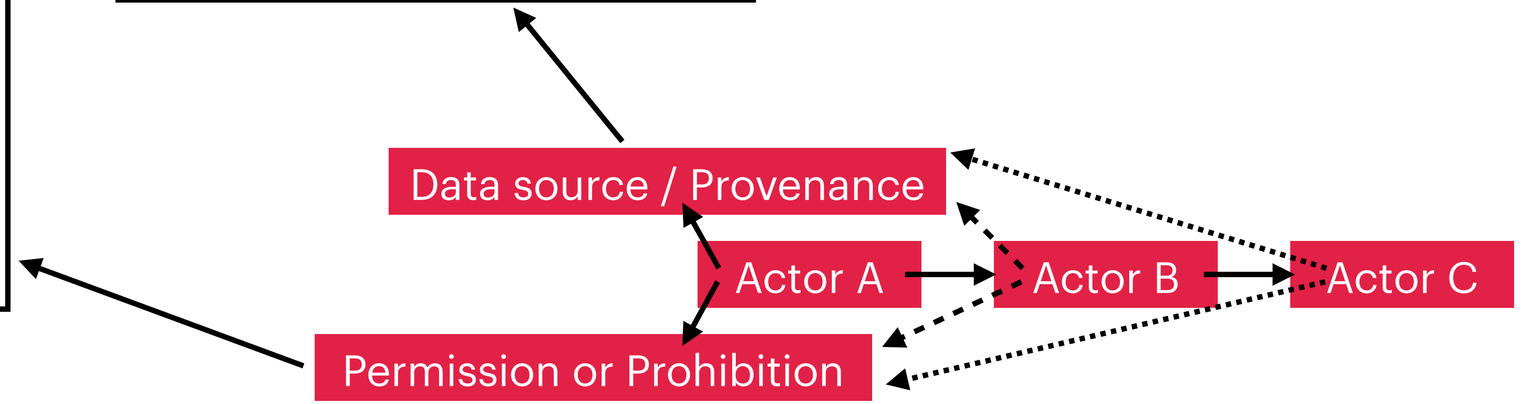
GDPR's principles providing a framework for ‘responsibility'

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)



Conduct a Data Protection Impact Assessment

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

Data Value

How to maximise the value of data?

How to minimise the "risks" associated with data?

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

Harmonising FAIR data sharing with Legal Compliance

Harmonising FAIR data sharing with Legal Compliance

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

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

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

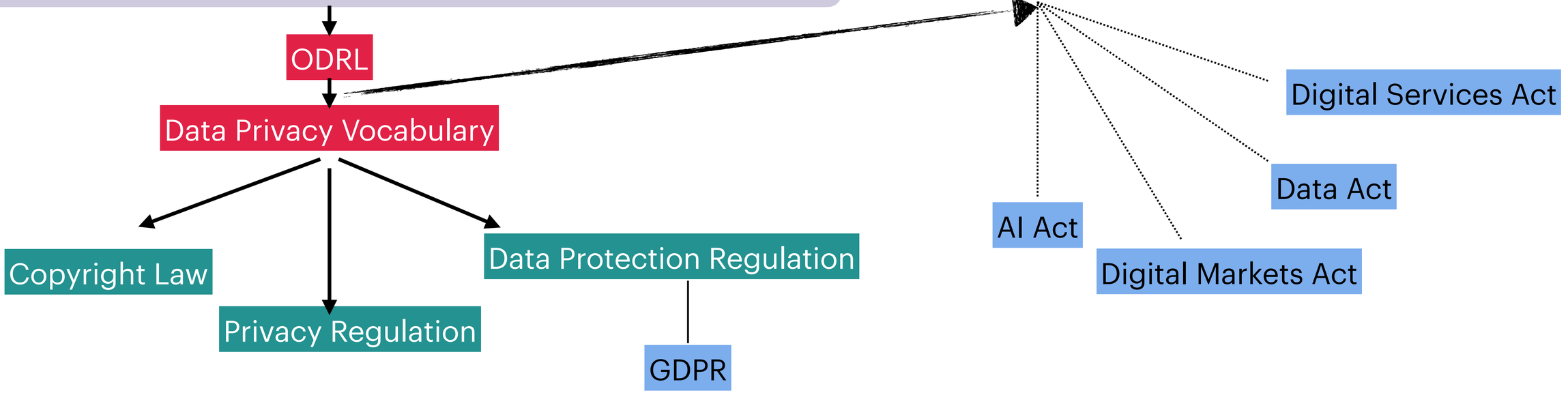
Harmonising FAIR data sharing with Legal Compliance

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

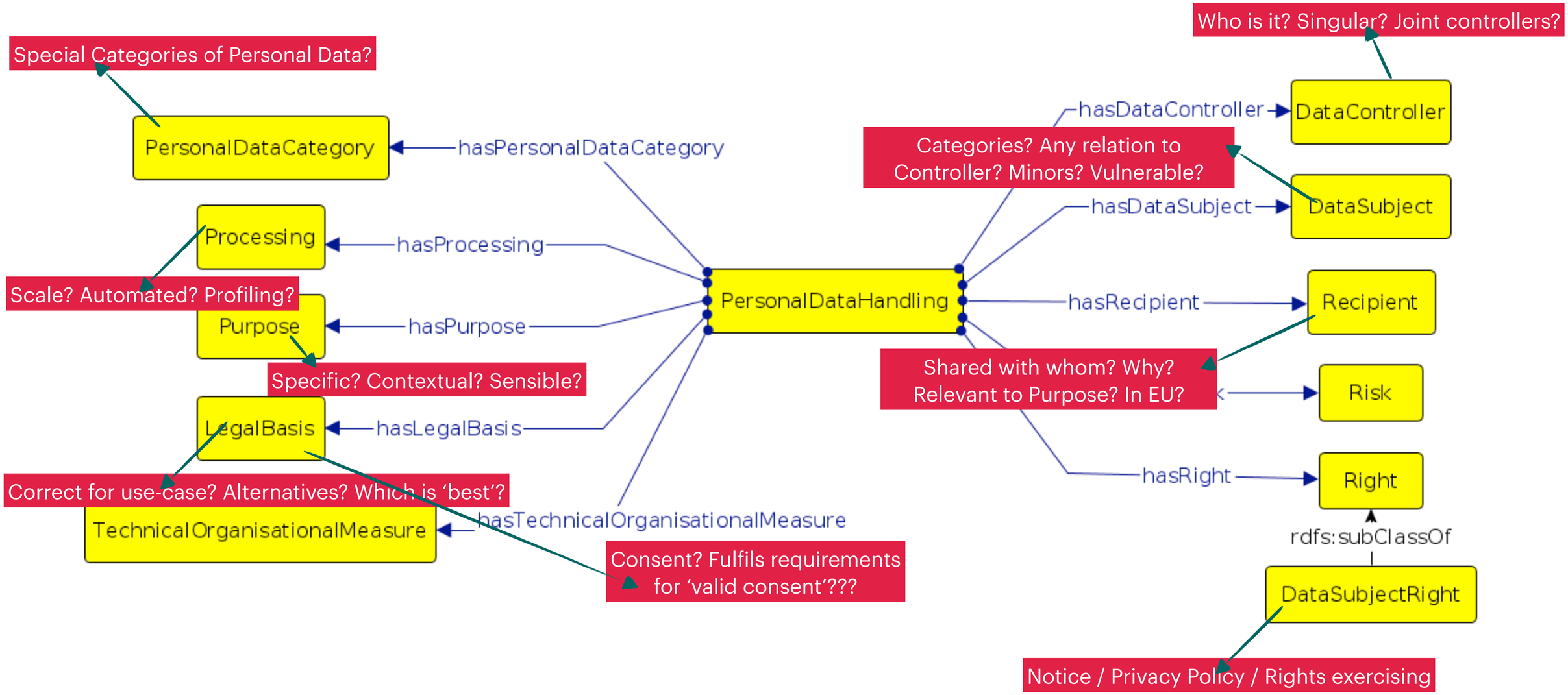
Legally Compliant FAIR =
Legally Compliant Data Sharing
= Legally Compliant Value

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

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



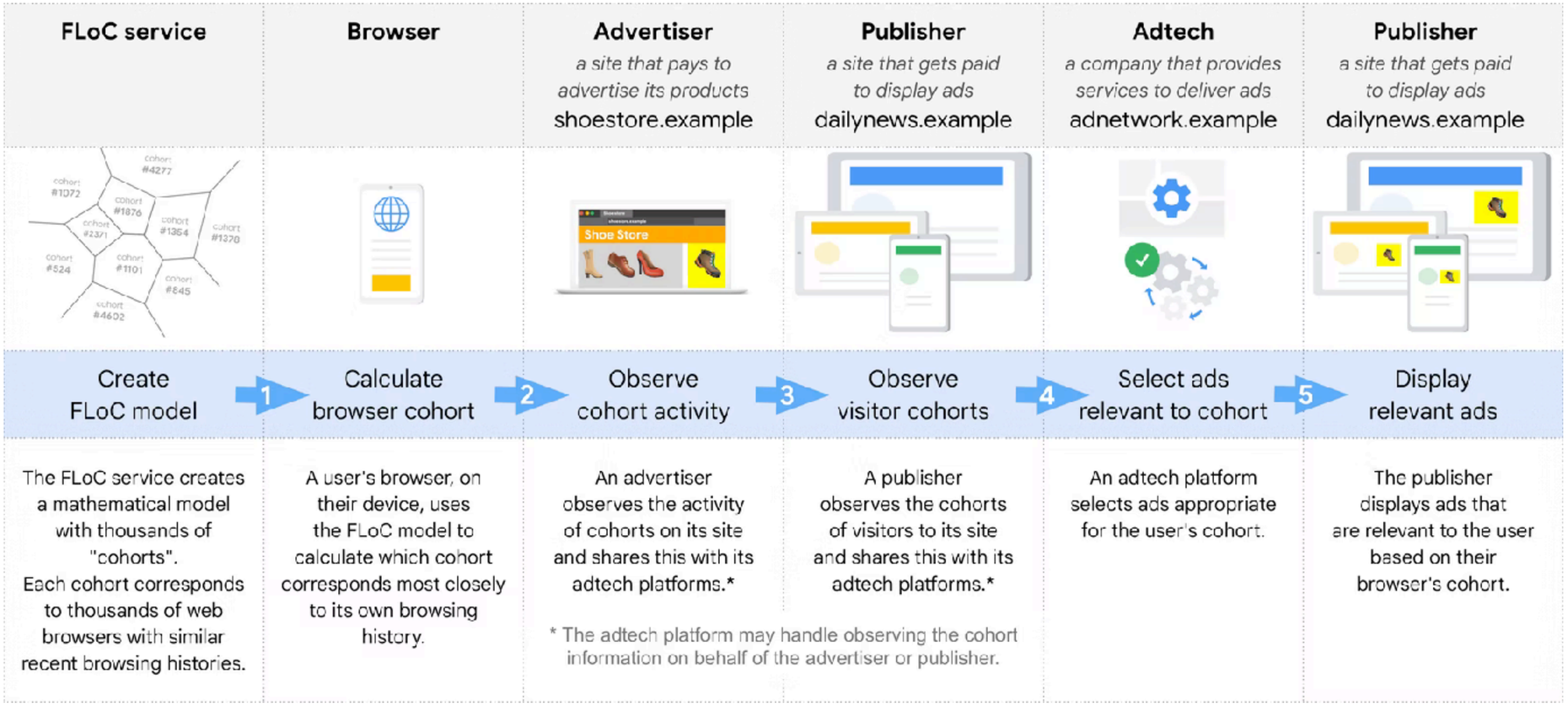
Labelling FAIR data for ‘sharing’ in legally compliant manner



RISKY

Sometimes keeping data can be risky, so we have new mechanisms to process data ‘client-side’, or in a ‘federated’, or ‘decentralised’ governance structures

Google's FLoC Proposal
Federated Learning of Cohorts



https://developer.chrome.com/docs/privacy-sandbox/floc/

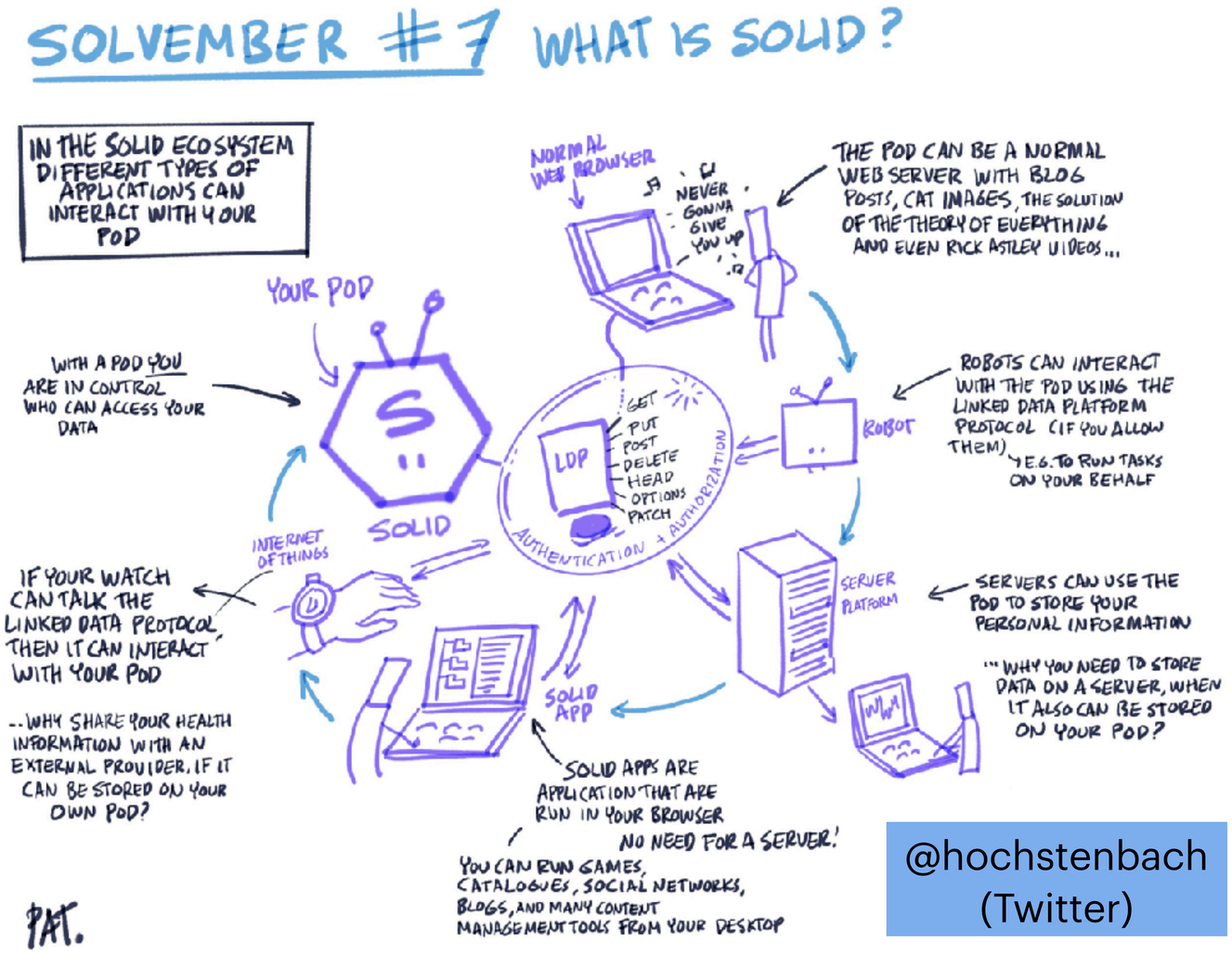
SOLID: A Decentralised Web

https://solidproject.org/

Centralised
- Companies decide how to collect, store data
- Companies decide how/where to use it
- Companies offer you choices and controls

Decentralised
- You “control” where your data is stored
- You “control” how it is used by apps/services
- You offer choices and controls

What will SOLID need to work?
- A new way to express privacy and preferences
- User-friendly UI/UX *without dark patterns*
- Legal enforcement to make companies respect negotiation of user preferences and settings



@hochstenbach (Twitter)

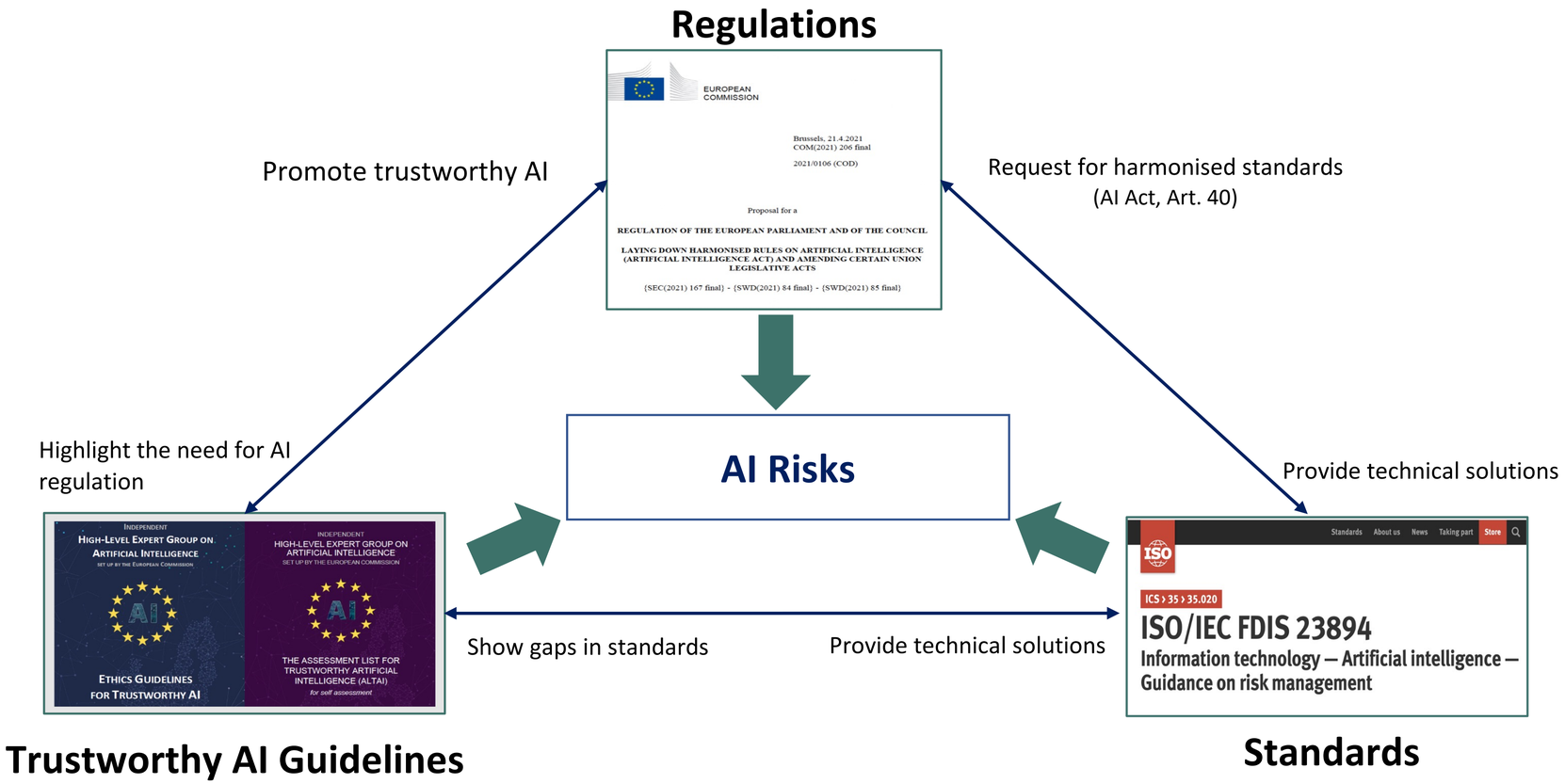
practicalities

Collecting and keeping EXCESSIVE data could be a problem
If it is personal data, especially if its “sensitive”

Data Value can depend on “data availability”. So the more interesting question these days is how to maximise data value with *minimum data being available,* or how to get data value without access to data itself. This requires new forms of algorithms and data governance e.g. decentralised.

All data has RISK, even if it is not personal data.
Here ‘risk’ can be security, IP, confidentiality, accuracy, performance, correctness, etc. At the same time, we want MORE MORE MORE data - so how to do both?

Efforts Addressing AI Risks



AIRO Requirements

Describing High-Risk AI Systems

Questions to identify whether an AI system is high-risk according to Annex III

Question	concept	Relation with AISystem
What techniques are utilised in the system?	**AI Technique**	usesAITechnique
What domain is the system intended to be used in?	**Domain**	isAppliedWithinDomain
What is the intended purpose of the system?	**Purpose**	hasPurpose
What is the application of the system?	**AI Application**	hasApplication
Who is the intended user of the system?	**AI User**	hasAIUser
Who is the subject of the system?	**AI Subject**	hasAISubject
In which environment is the system used?	**Environment Of Use**	isUsedInEnvironment

ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 1

(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods.

ANNEX III
HIGH-RISK AI SYSTEMS REFERRED TO IN ARTICLE 6(2)

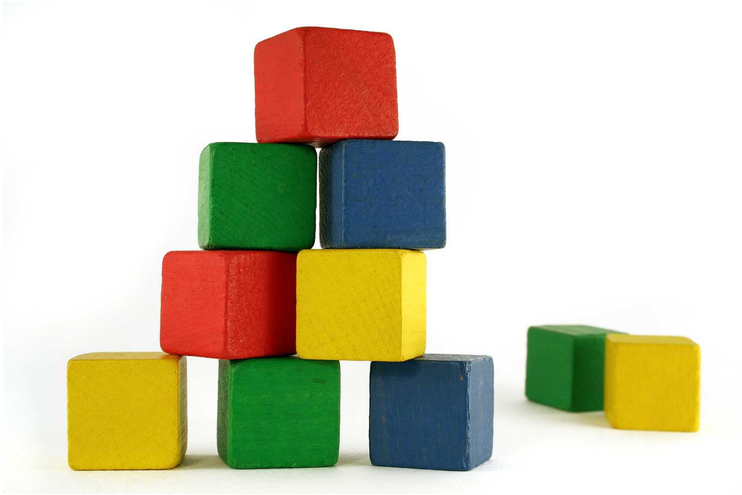
High-risk AI systems pursuant to Article 6(2) are the AI systems listed in any of the following areas:

1. Biometric identification and categorisation of natural persons:
   (a) AI systems intended to be used for the 'real-time' and 'post' remote biometric identification of natural persons;
2. Management and operation of critical infrastructure:
   (a) AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.
3. Education and vocational training:
   (a) AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions;
   (b) AI systems intended to be used for the purpose of assessing students in educational and vocational training institutions and for assessing participants in tests commonly required for admission to educational institutions.
4. Employment, workers management and access to self-employment:
   (a) AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;
   (b) AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.
5. Access to and enjoyment of essential private services and public services and benefits:
   (a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such

Future Solution

Modular + Simple Vocabularies focused on making practical tasks more effective rather than theoretical logical correctness

We **NEED** this for *fast & adaptive solutions* for the new data regulatory regime.



Law	Enforcement
GDPR	MAY-2018
DSA	NOV-2022
DMA	MAY-2023
DGA	SEP-2023
AI Act	draft
ePrivacy Reg	draft
Data Act	proposed
Health Data Space	proposed
Interop. Act	proposed

Regulating Data
CA4025 | DCU

interested? questions? contact at:

harshvardhan.pandit@dcu.ie
me@harshp.com
harsh@eupolicy.social (mastodon)