KU LEUVEN CiTiP

CENTRE FOR IT & IP LAW

imec

Data Protection Impact Assessment and Data Protection by Design

Two elephants in the GDPR room

Pierre Dewitte – KU Leuven CiTiP-imec

✉ pierre.dewitte@kuleuven.be
🐦 @PiDewitte

DPbD, DPIA: A tale of two methodologies

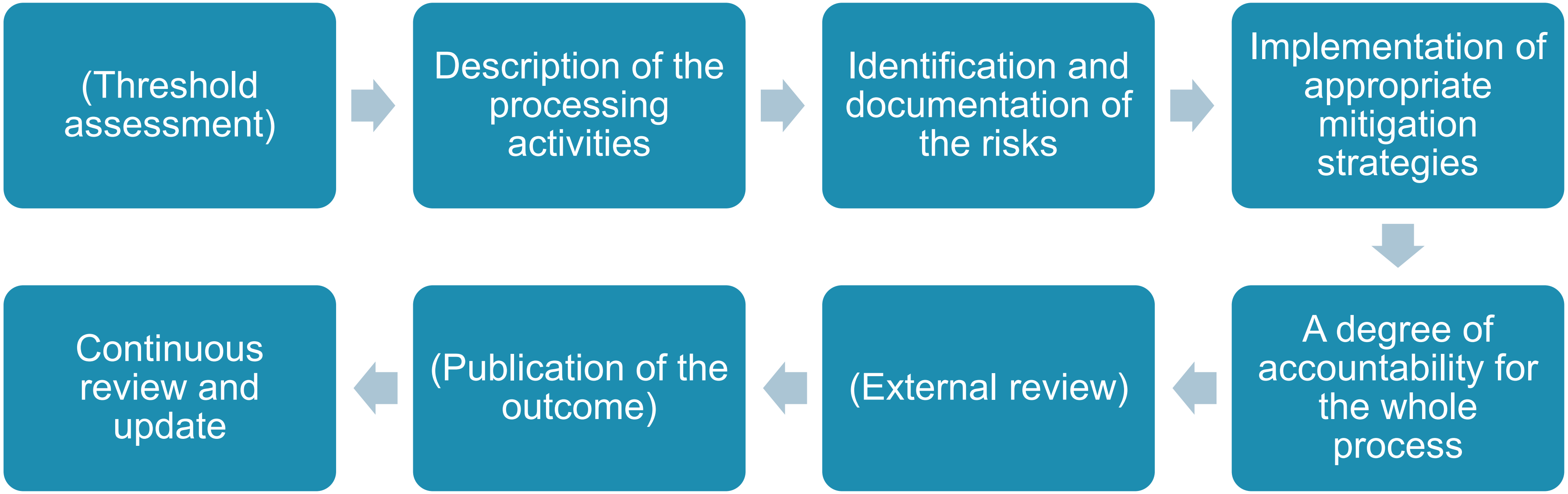
Data Protection by Design

According to Articles 24(1) and 25(1) GDPR, NSA and legal literature, controllers are under the obligation to:

1. Adopt a **risk-based approach**
   - Requires taking into account elements such as the state of the art, the cost of implementation, the nature, scope, context and purposes of the processing, the risks for the rights and freedoms of data subjects
2. When implementing appropriate **technical** and **organisational** measures
   - Suggests a close collaboration between experts in various disciplines
3. To **ensure** and **demonstrate** compliance with the Regulation
   - Requires the implementation of appropriate mitigation strategies
   - Requires a layer of demonstrability (accountability)
4. Both at the time of the **determination of the means** for processing and at the time of the **processing** itself
   - As of the design stage, and throughout the entire data processing lifecycle

Data Protection Impact Assessment

Following the combined reading of Article 35(7), WP29, NSA and legal literature, a DPIA essentially consists of the following steps:



(Threshold assessment)

Description of the processing activities

Identification and documentation of the risks

Implementation of appropriate mitigation strategies

A degree of accountability for the whole process

(External review)

(Publication of the outcome)

Continuous review and update

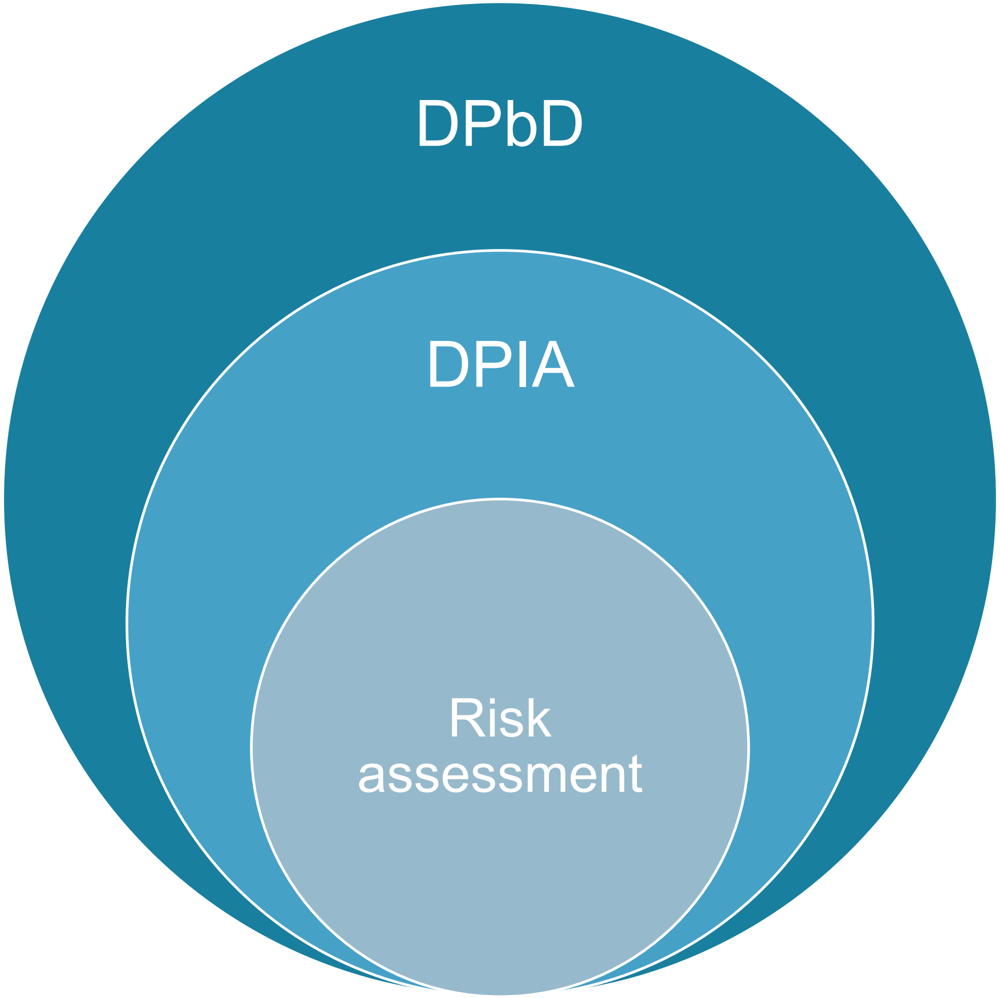
Not so different?

Description of the system

Risk assessment and implementation of appropriate mitigation strategies

Accountability

Not so different?



Yet, in practice…

- DPbD, DPIA and risk assessment are essentially interdisciplinary approaches, involving both software engineers and lawyers
- Clear dichotomy between the way software engineers and lawyers practically implement those requirements:

Software engineers		Lawyers
• Technical concepts • Technical view on the system, overinclusive • Risks understood from a narrower perspective • Agile, (partially) automated	≠	• Legal concepts • Legal view, often disconnected from reality • Risks to data subject's rights and freedoms • Rigid, manual
E.g. LINDDUN privacy threat modelling methodology		*E.g. CNIL PIA guidance and open source tool*

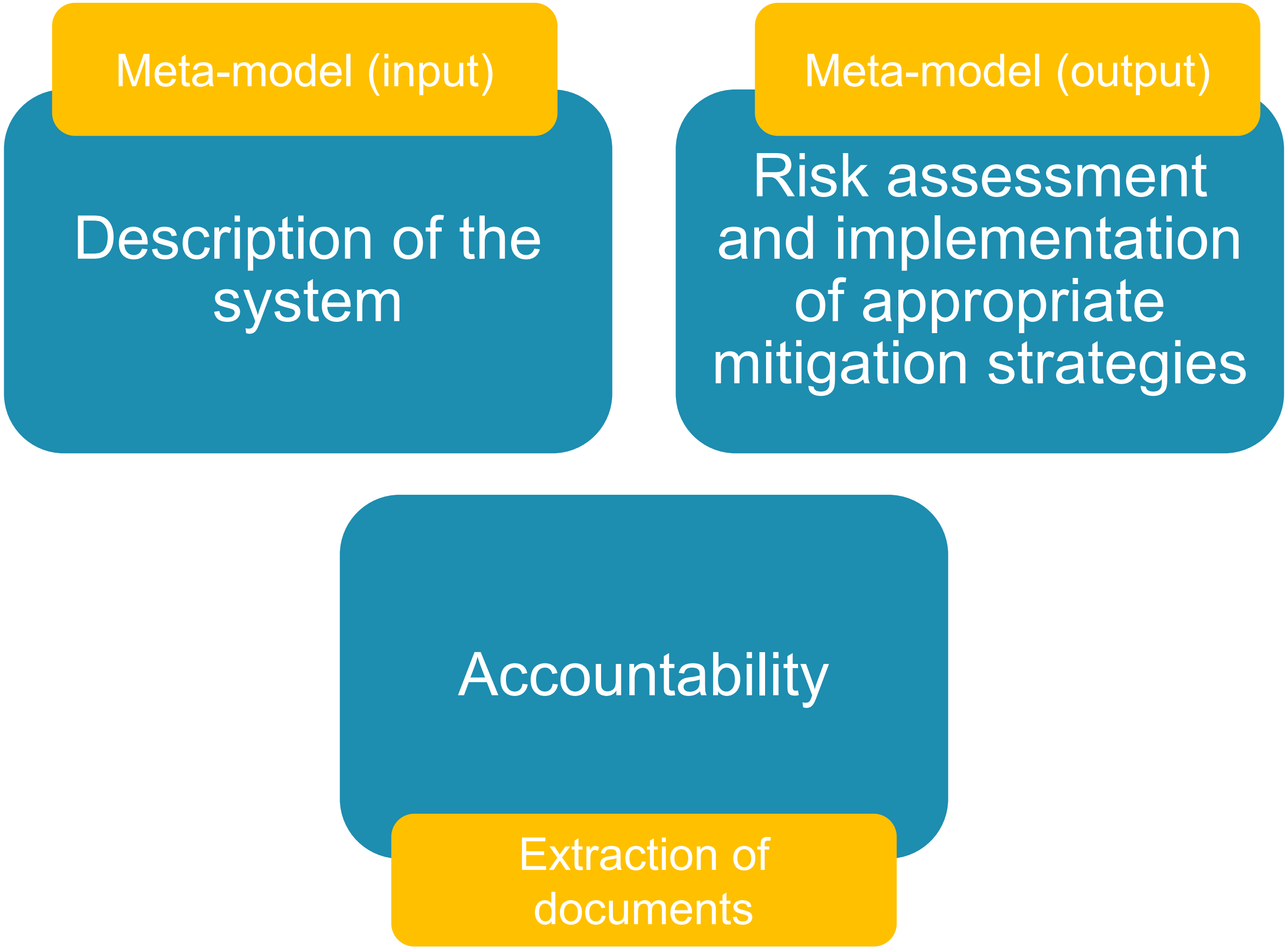
Impact on transparency

- Meaningful transparency, either *ex-ante* or *ex-post*, relies on the premise that one does know exactly what happens within a system
    - As underlined earlier, lawyers often lack the technical knowledge to understand **all the processing activities** happening within a system
    - This negatively impacts the drafting of comprehensive, adequate and up-to-date **privacy policies since** understanding is a prerequisite
    - As demonstrated during the empirical study on the right of access, this also hinder the exercise of **data subject's rights**

An efficient DPbD methodology **aligning the technical and legal views of a system** would significantly facilitate the compliance with transparency requirements, but not only

Aligning: Of software engineers and lawyers

The PRiSE project



Meta-model (input)

Description of the system

Meta-model (output)

Risk assessment and implementation of appropriate mitigation strategies

Accountability

Extraction of documents

Privacy-by-design **R**egulation **i**n **S**oftware **E**ngineering (PRiSE), KU Leuven CiTiP and DistriNet, <https://www.law.kuleuven.be/citip/en/research/projects/ongoing/prise>

The PRiSE meta-model (input)

- Software engineers rely on threat modelling techniques such as STRIDE or LINDDUN for the elicitation of security and privacy threats.
- Those methodologies are based on Data Flow Diagrams (DFD) which are mainly built using the following architectural modelling requirements: **events**, **processes**, **responses**, **data sources** and **recipients**.

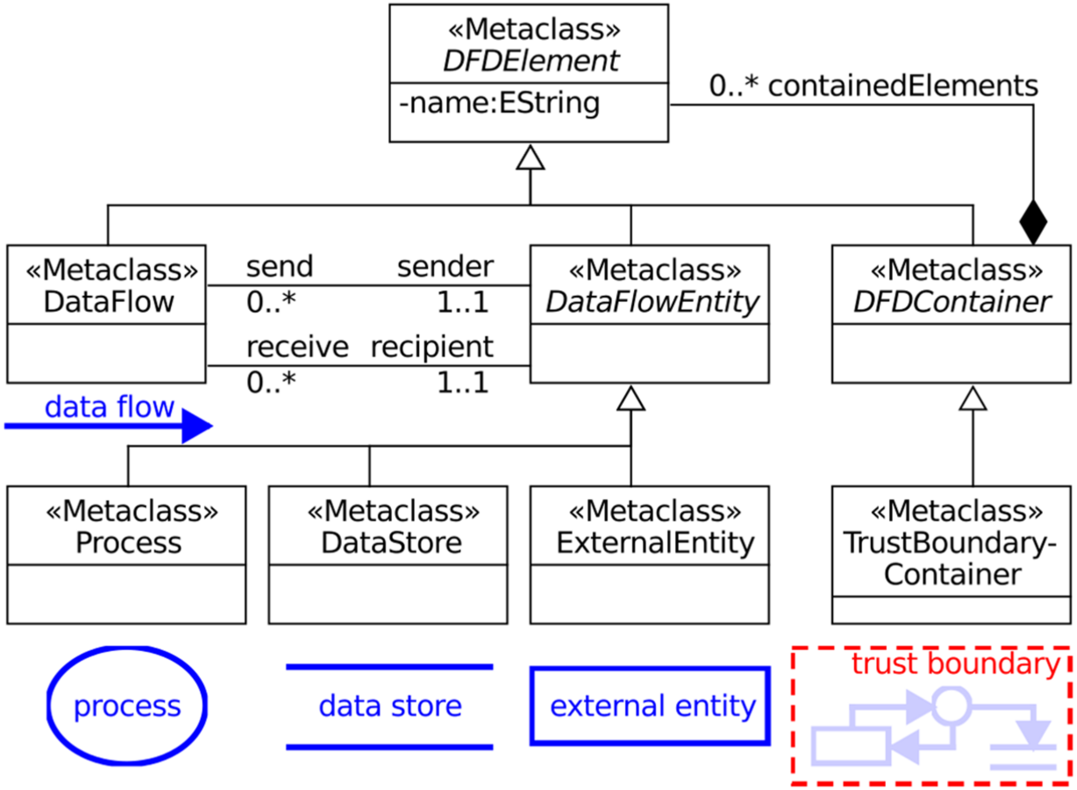
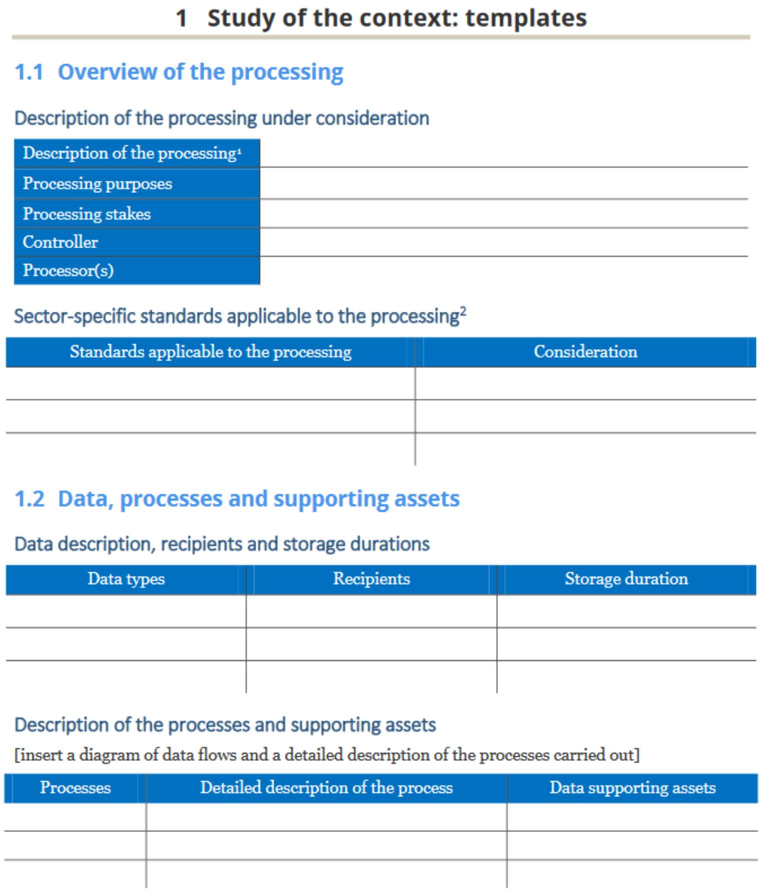


***Fig. 1**: Example of DFD using UML*

Dewitte et al., ‘A Comparison of System Description Models for Data Protection by Design’, *Proceedings of the 2019 Symposium on Applied Computing*, Limassol, 8 April 2019.

The PRiSE meta-model (input)



1 Study of the context: templates

1.1 Overview of the processing

Description of the processing under consideration

Description of the processing[1]	
Processing purposes	
Processing stakes	
Controller	
Processor(s)	

Sector-specific standards applicable to the processing[2]

Standards applicable to the processing	Consideration

1.2 Data, processes and supporting assets

Data description, recipients and storage durations

Data types	Recipients	Storage duration

Description of the processes and supporting assets

[insert a diagram of data flows and a detailed description of the processes carried out]

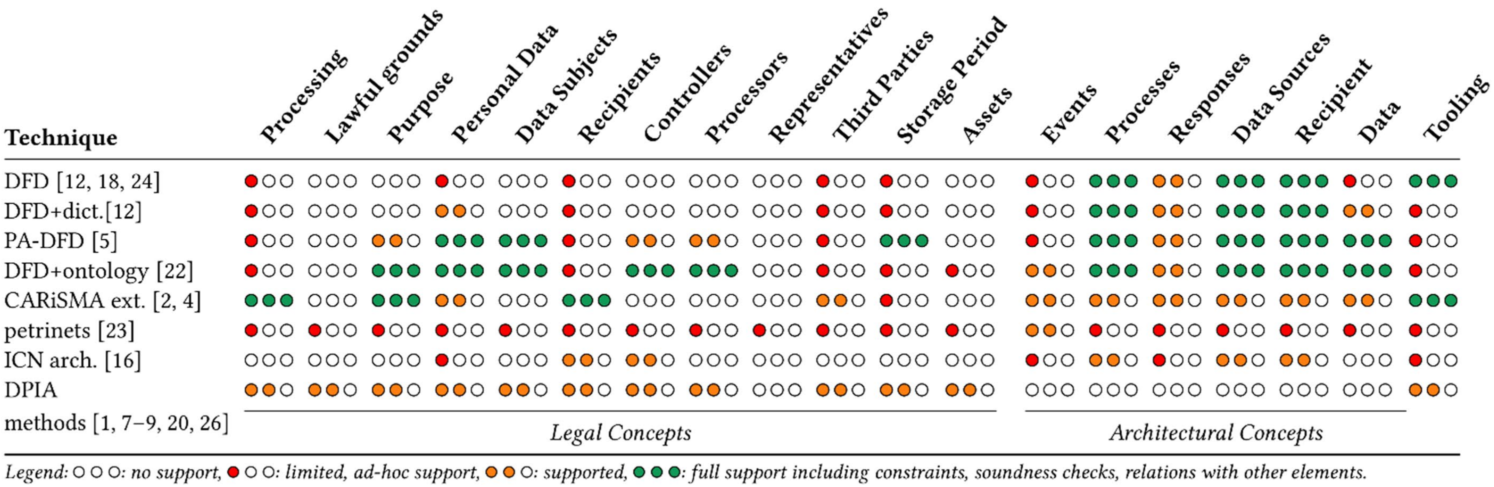
Processes	Detailed description of the process	Data supporting assets

***Fig.** 2: Example of system description using CNIL's template*

- A DPIA usually starts with the systematic description of the processing activities.
- Guidance has been issued by the Article 29 Working Party as to the elements that must be documented, which led to the extraction of the following legal modelling requirements: **personal data**, **data subject**, **processing**, **purpose**, **lawful ground**, **controller**, **processor**, **third party**, **recipient**, **representative**, **storage period** and **supporting assets**.

Dewitte et al., 'A Comparison of System Description Models for Data Protection by Design', *Proceedings of the 2019 Symposium on Applied Computing*, Limassol, 8 April 2019.

The PRiSE meta-model (input)



Technique	Processing	Lawful grounds	Purpose	Personal Data	Data Subjects	Recipients	Controllers	Processors	Representatives	Third Parties	Storage Period	Assets	Events	Processes	Responses	Data Sources	Recipient	Data	Tooling
DFD [12, 18, 24]	🔴○○	○○○	○○○	🔴○○	○○○	🔴○○	○○○	○○○	○○○	🔴○○	🔴○○	○○○	🔴○○	🟢🟢🟢	🟠🟠○	🟢🟢🟢	🟢🟢🟢	🔴○○	🟢🟢🟢
DFD+dict.[12]	🔴○○	○○○	○○○	🟠🟠○	○○○	🔴○○	○○○	○○○	○○○	🔴○○	🔴○○	○○○	🔴○○	🟢🟢🟢	🟠🟠○	🟢🟢🟢	🟢🟢🟢	🟠🟠○	🔴○○
PA-DFD [5]	🔴○○	○○○	🟠🟠○	🟢🟢🟢	🟢🟢🟢	🔴○○	🟠🟠○	🟠🟠○	○○○	🔴○○	🟢🟢🟢	○○○	🔴○○	🟢🟢🟢	🟠🟠○	🟢🟢🟢	🟢🟢🟢	🟢🟢🟢	🔴○○
DFD+ontology [22]	🔴○○	○○○	🟢🟢🟢	🟢🟢🟢	🟢🟢🟢	🔴○○	🟢🟢🟢	🟢🟢🟢	○○○	🔴○○	🔴○○	🔴○○	🟠🟠○	🟢🟢🟢	🟠🟠○	🟢🟢🟢	🟢🟢🟢	🟢🟢🟢	🔴○○
CARiSMA ext. [2, 4]	🟢🟢🟢	○○○	🟢🟢🟢	🟠🟠○	○○○	🟢🟢🟢	○○○	○○○	○○○	🟠🟠○	🔴○○	○○○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟢🟢🟢
petrinets [23]	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🟠🟠○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○	🔴○○
ICN arch. [16]	○○○	○○○	○○○	🔴○○	○○○	🟠🟠○	🟠🟠○	○○○	○○○	○○○	○○○	○○○	🔴○○	🟠🟠○	🔴○○	🟠🟠○	🟠🟠○	○○○	🔴○○
DPIA methods [1, 7–9, 20, 26]	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	🟠🟠○	○○○	🟠🟠○	🟠🟠○	🟠🟠○	○○○	○○○	○○○	○○○	○○○	○○○	🟠🟠○
	Legal Concepts												*Architectural Concepts*						

Legend: ○○○*: no support,* 🔴○○*: limited, ad-hoc support,* 🟠🟠○*: supported,* 🟢🟢🟢*: full support including constraints, soundness checks, relations with other elements.*

Tab. 1*: Evaluation of existing DPIA/threat modelling methodologies w.r.t. legal and architectural modelling requirements*

- No methodology supports both legal and architectural modelling requirements.
- Both assessments therefore rely on a different conception of the system.
  - **DPIAs** overlook technical specificities that could nonetheless prove relevant from a data protection perspective (e.g. non personal data flows).
  - **Threat modelling methodologies**, on the other hand, omit GDPR-specific concepts (e.g. lawful grounds (Art. 5(1)a and 6(1) and purposes (Art. 5(1)b) that are crucial to achieve in-depth, efficient compliance with the Regulation.

Dewitte et al., 'A Comparison of System Description Models for Data Protection by Design', *Proceedings of the 2019 Symposium on Applied Computing*, Limassol, 8 April 2019.

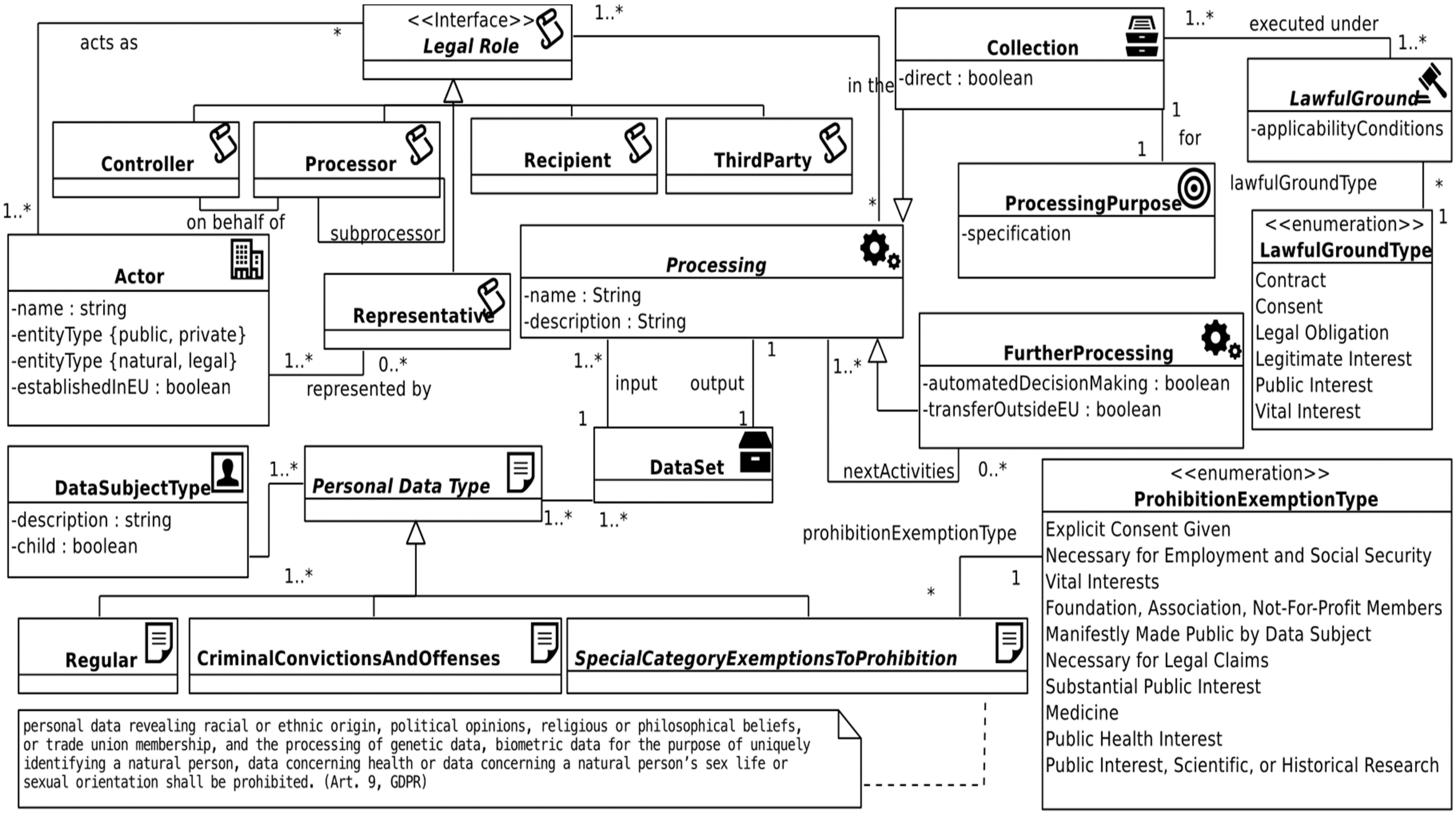
The PRiSE meta-model (input)

- When building the meta-model, the following choices were made:
  - Rely on **GDPR terminologies and concepts**, rather than on technical abstractions (~~entity~~, ~~process~~, ~~data store~~, ~~data flow~~, etc. but rather controller, processor, processing, further processing, etc.)
  - Rely on **software engineer's way to describe a system** rather than on the overly simplified descriptions often found in most DPIA methodologies (UML class diagram)
  - Rely on the **abstractions that are necessary to perform the various checks** that are required by the GDPR (avoids overinclusivity, allows superposition of the technical view if necessary)

A sound description of the system which reconciles the data protection and technical views of a system is the **essential first step** of a consistent, in-depth and agile DPbD methodology

Sion et al., 'An Architectural View for Data Protection by Design', *Proceedings of the 2019 International Conference on Software Architecture*, Hamburg, 29 March 2019

Work in progress, for demonstration purposes only



personal data revealing 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. (Art. 9, GDPR)

Sion et al., 'An Architectural View for Data Protection by Design', *Proceedings of the 2019 International Conference on Software Architecture*, Hamburg, 29 March 2019

The PRiSE meta-model (output)

- Modelling a system using the meta-model allows for:
  - **Fully automated checks** (constraints imposed by the meta-model itself)
  - **Semi automated checks** (‘Clippy’-like support providing relevant guidance, resources and indications)
  - **Manual checks** (checks that are not directly supported by the meta-model, but whose assessment is facilitated by the meta-model).

Example with Art. 5(1)b GDPR

- **Purpose specification** is fully automated, since the meta-model will requires one or more *ProcessingPurpose* for every *CollectionActivity*
- **Compatibility assessment** is semi automated, since the meta-model will pair every *FurherProcessingActivtiy* with the original *ProcessingPurpose*, the *LegalBasis*, the *DataSubjectType* and the *DataType* and raise the necessity to conduct such an assessment, together with guidance as to the relevant criteria.

Sion et al., ‘An Architectural View for Data Protection by Design’, *Proceedings of the 2019 International Conference on Software Architecture*, Hamburg, 29 March 2019

The PRiSE meta-model (agility and extraction of documents)

- Thanks to correspondence rules introduced between the technical and the data protection view, changes in the former are reflected in the latter, and *vice-versa*
  - Addresses **architectural erosion**
  - For instance, the introduction of a new data flow in the technical view will be considered as a new *FurherProcessingActivtiy*, which, in turn, will trigger all the **necessary checks** (*i.e.* compatibility assessment, new actor, rules on transfer, etc.) and pair them with the relevant information
- Tool support will be developed to allow the extraction of documents reflecting the data protection view built according to the meta-model
  - Streamlines the production of **accountability documents**
  - Format and content of the documents **flexible**

Sion et al., 'An Architectural View for Data Protection by Design', *Proceedings of the 2019 International Conference on Software Architecture*, Hamburg, 29 March 2019

Of software engineers and lawyers

Communicating: Of design scientists and lawyers

Modalities surrounding transparency

- Once the controller is fully aware of what is exactly happening within its system, it must communicate that information to data subjects
- Article 12(1) GDPR pairs the obligation of transparency with several modalities, be it for information included in the privacy policy or provided to data subject following the exercise of their rights
    - See Article 29 Working Party guidelines on transparency (WP260)

=

- Concise + Transparent = data controllers should present the information/ communication efficiently and succinctly in order to avoid information fatigue + clearly differentiated from other non-privacy related information such as contractual provisions; online: layered privacy notice

Modalities surrounding transparency

- Intelligible = understood by an average member of the intended audience (= (1) identify intended audience + (2) ascertain level of understanding; + (3) regularly check)
- Easily accessible = data subject should not have to seek out the information (=immediately apparent); online: by way of contextual pop-ups which activate when a data subject fills in an online form, or in an interactive digital context through a chatbot interface
- Clear and plain language = best practices for clear writing: avoid complex sentences + concrete (≠ “may”, “might”, “some”, “often”, “possible”, “research”, “personalisation”) + active voice – legalese, technical vocabulary + translation when targeting data subjects speaking different languages

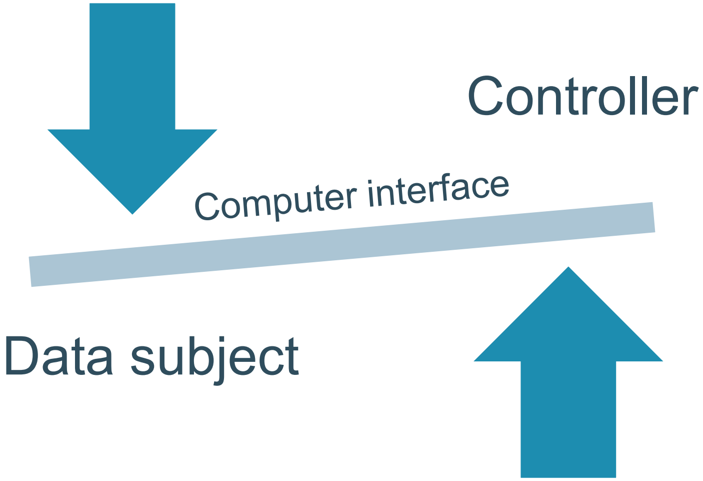
Modalities surrounding transparency

- In writing or by other electronic means = method chosen to provide information must be appropriate to the circumstance (*e.g.* written only for screenless devices such as IoT is wrong); means: layered privacy notice; "just-in-time" contextual pop-up notices, 3D touch or hover-over notices, privacy dashboards, cartoons, infographics or flowcharts
- Appropriate measures = taking into account the device used, the nature of the user interfaces/interactions with the controller and the relevant limitations; trial different modalities by way of user testing
- Other means = for specific environments: hard copy, oral explanation, icons, GR codes, videos, SMS, email, public signage, newspapers campaigns, notice in media, etc.

A job for lawyers only?

The role of design scientists

- Most interactions between data subjects and controllers happen through computerized interfaces, especially in the context of:
  - **Transparency requirements** (*e.g.* legibility/availability of a privacy policy; *c.f.* WP29 guidelines)
  - The exercise of **data subject's rights** (e.g. erasure, access, etc.)



GDPR and HCI, perfect match? A data protection lawyer's perspective on CHI 2018, CiTiP blog, <https://www.law.kuleuven.be/citip/blog/gdpr-and-hci-a-perfect-match-a-data-protection-lawyers-perspective-on-chi-2018/>

The role of design scientists

- Many initiatives in the field of HCI (Human Computer Interaction) address the need for user-friendly, efficient transparency:
  - **Methodology** taking both controllers' and data subjects' expectation into account at the interface design stage (Eiband et al.)
  - Exploration of **design spaces** for effective privacy notice (Schaub et al.)
  - Field-testing **user's expectations** of transparency (Lyngs et al.)
  - Exposing **nudges** for privacy and security (Acquisti et al.)
  - Development of **privacy languages** (Zhao et al.)
  - **Interactive display** visualizing data subject's exposure to third party tracking activities on smartphone app (Van Kleek et al.)

Design science is a fertile ground to **develop intrinsically interdisciplinary solutions** to address the many challenges raised by transparency

GDPR and HCI, perfect match? A data protection lawyer's perspective on CHI 2018, CiTiP blog, <https://www.law.kuleuven.be/citip/blog/gdpr-and-hci-a-perfect-match-a-data-protection-lawyers-perspective-on-chi-2018/>

Of design scientists and lawyers

Transparency – and GDPR compliance altogether – is a truly interdisciplinary endeavour

Thanks for your attention!

KU Leuven
Centre for IT & IP Law (CiTiP) –
imec

www.law.kuleuven.be/citip