

Privacy and Data Protection 4 Engineering

Model-driven Engineering Tool and Method for Privacy and Data Protection by Design

Gabriel Pedroza (CEA)

Victor Muntés-Mulero (Beawre)

Yod Samuel Martin (UPM)

Guillaume Mockly (Trialog)

- Introduction and objectives
 - ❑ Privacy and Data Protection by Design (PDPbD): context and challenges
- Proposed method for PDPbD
- Tool support for the method
 - ❑ Personal Data Detector Module
 - ❑ Module for Privacy Model-driven design
 - ❑ Module for Code Validation
- Summary of achievements

Privacy and Data Protection by Design

➤ Context

□ Design engineers' ecosystem:

- Several stakeholders and actors
- Variety of needs and objectives
- Solution for conflicting goals/reqs.

□ Designer's questions to address:

- Which privacy-aspects introduce during systems design?
- How identified concerns can be considered at early design steps?
- How privacy-by-design can be effectively realized?

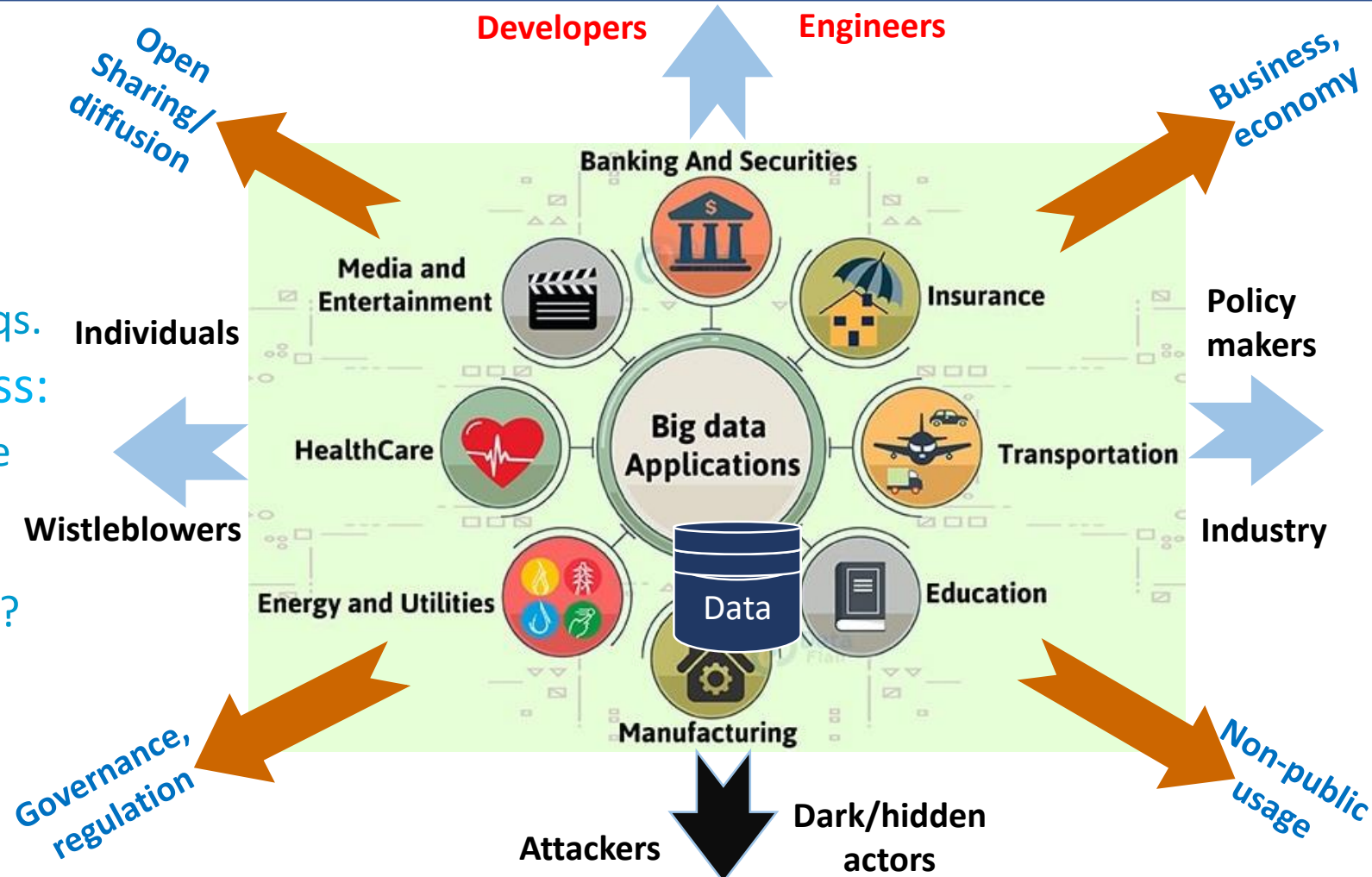
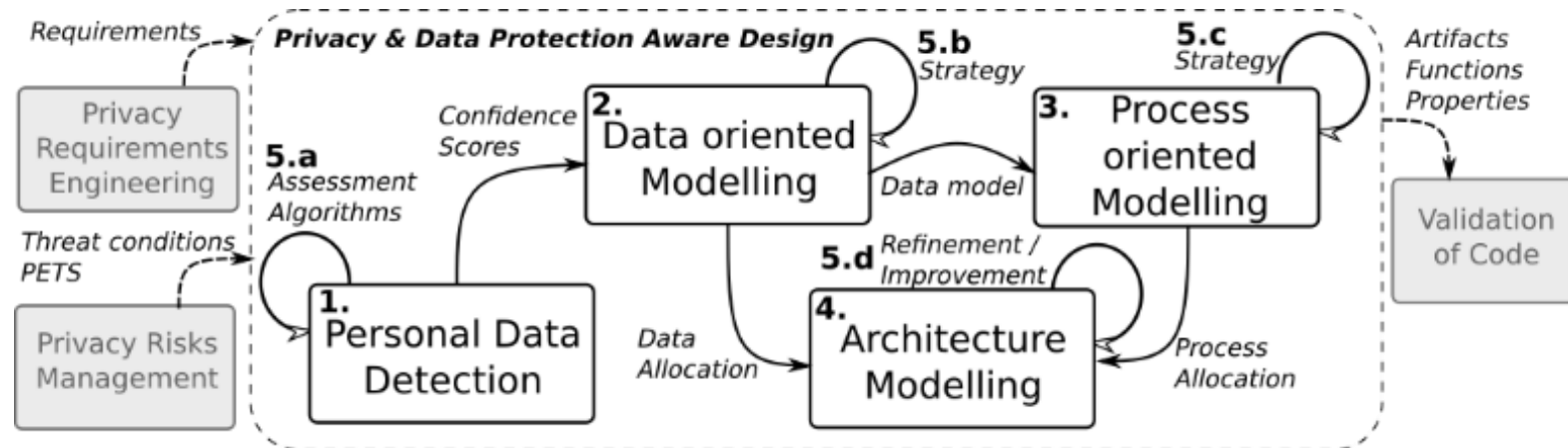


Image borrowed from <https://www.digitalvidya.com/>

PDP by Design Method

➤ Main characteristics:

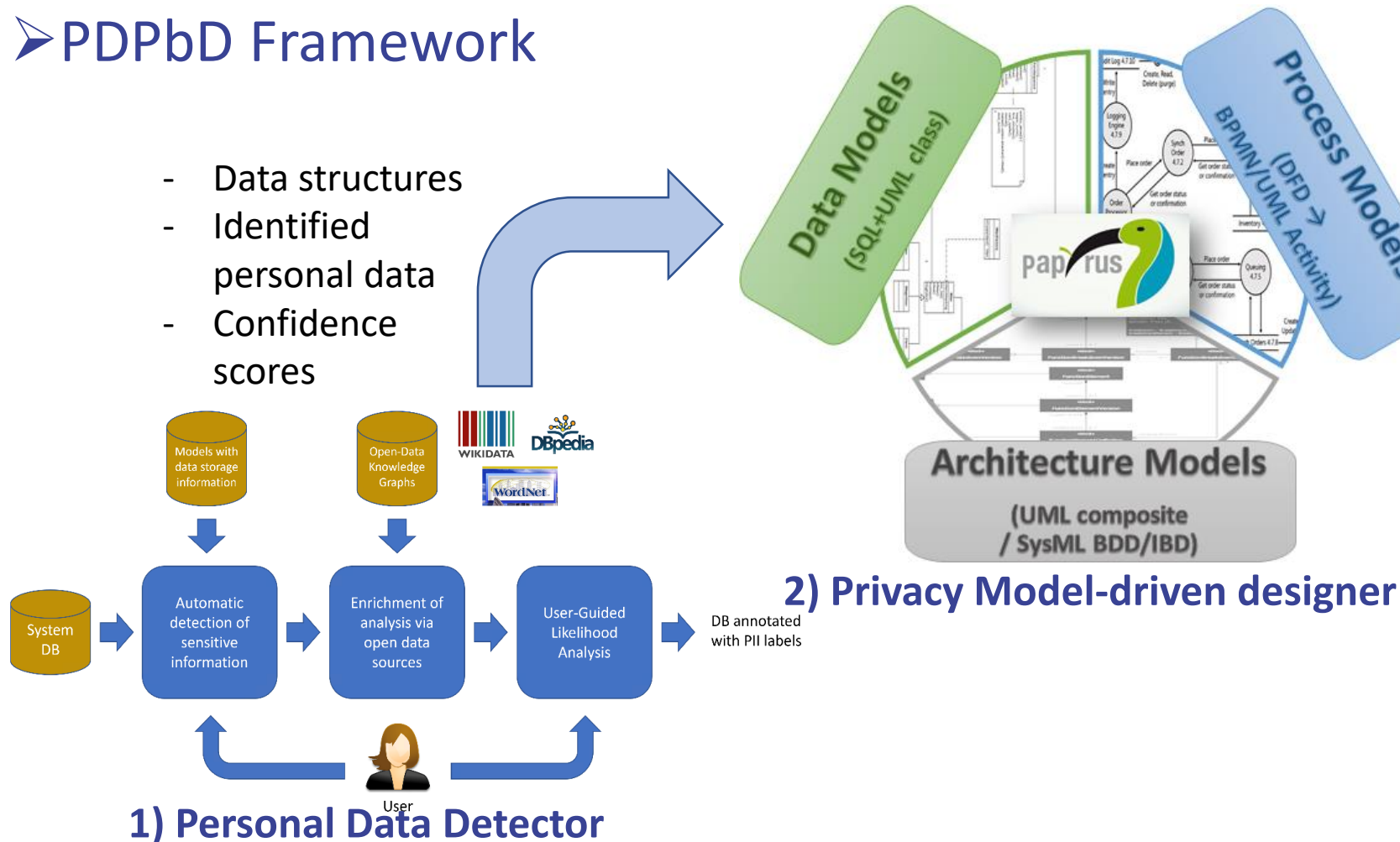
- ❑ Identification of personal data
- ❑ Combined bottom-up and top-down approaches:
 - From data structures to data and data-flow (process) models
 - Allocation over an architecture model
 - Architecture refinement towards code
- ❑ Models improved by Privacy-by-design strategies (ISO/IEC 27550)
- ❑ Validation of properties at code level



Tool support for the PDPbD method

➤ PDPbD Framework

- Data structures
- Identified personal data
- Confidence scores



- Target of Validation
- Components
 - Pointers to code
 - Privacy properties

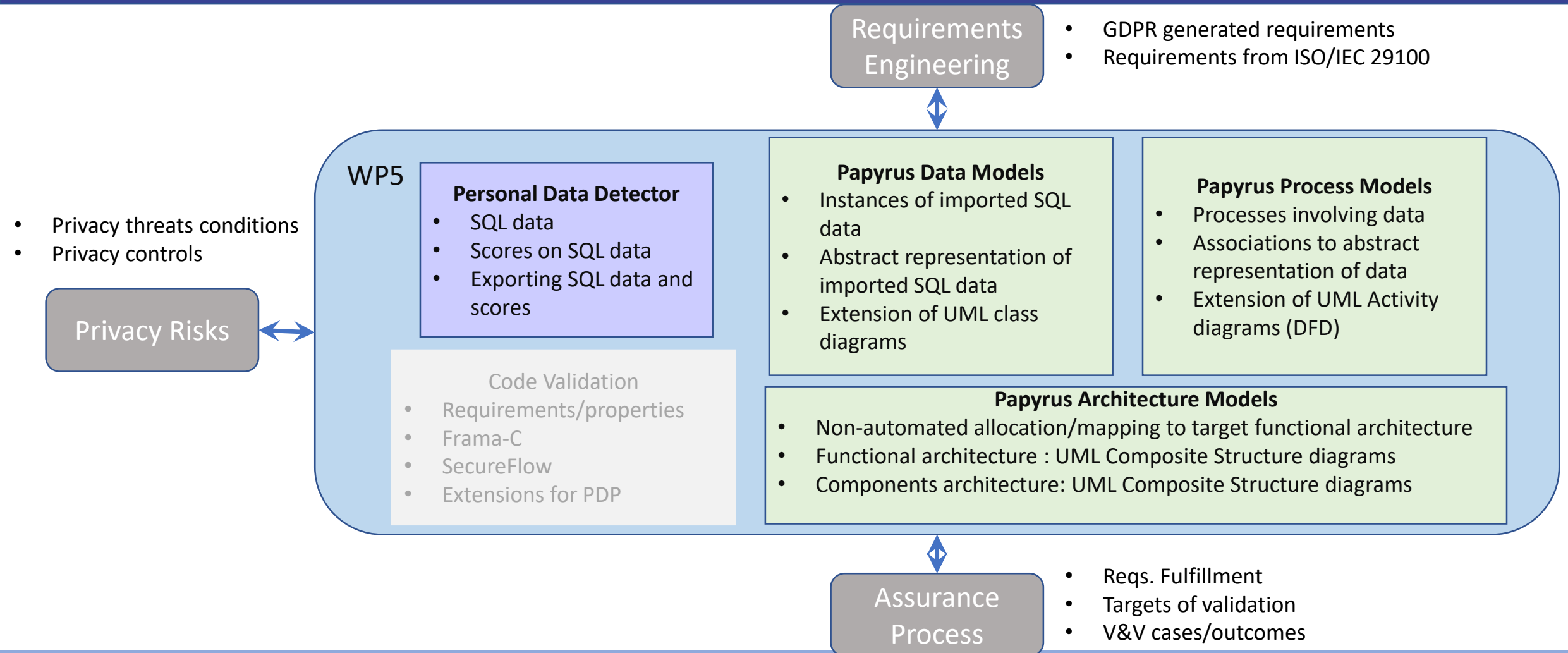
Code validation and verification



- Privacy flaws
- Code improvement

3) Module for Code Validation

Tool support for the PDPbD method

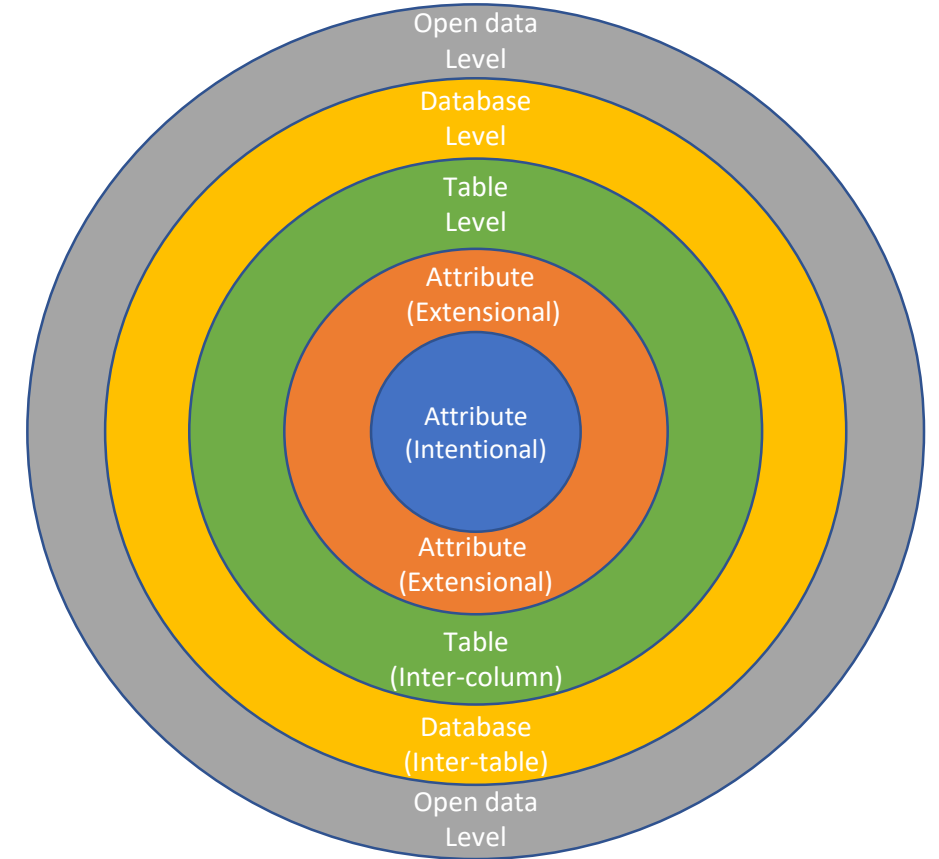
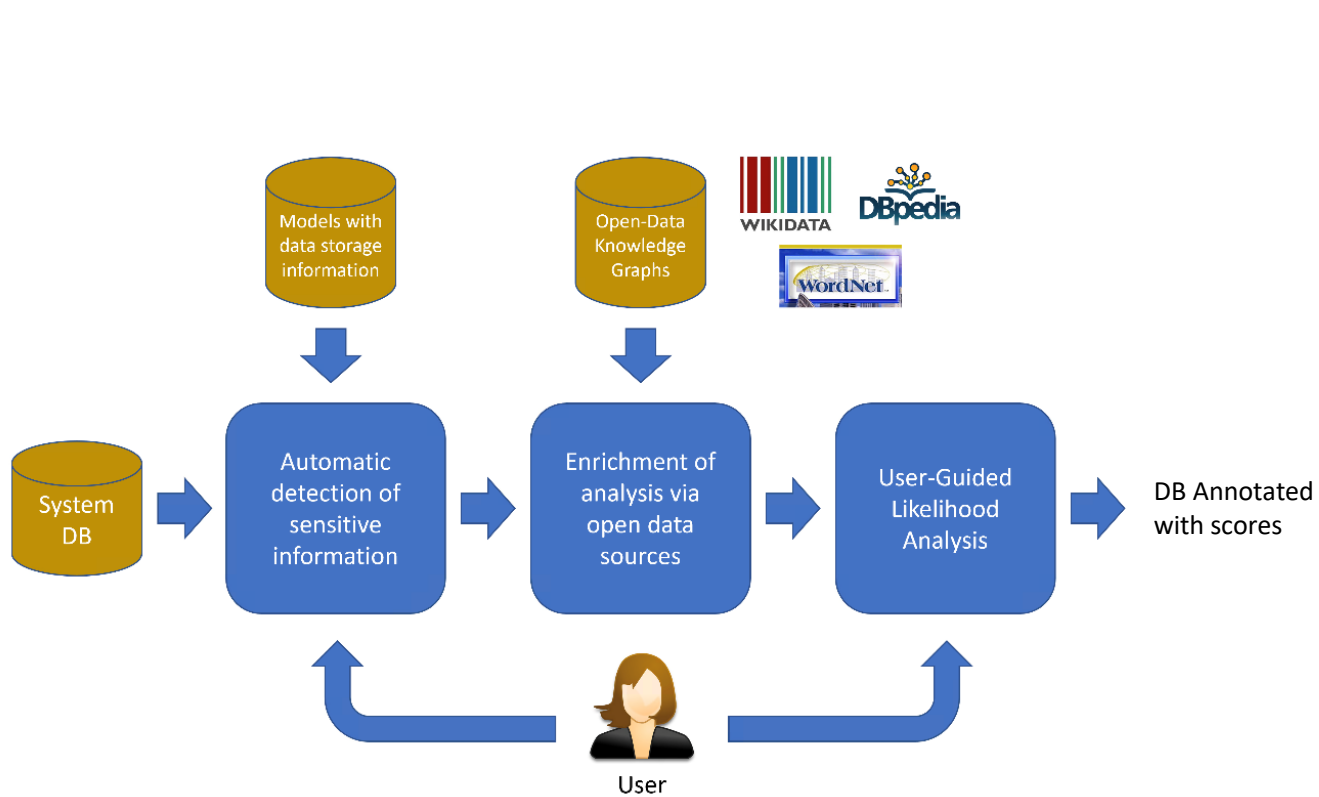


1. Personal Data Detector Module

Victor Muntés (Beawre)

Victor.Muntes@beawre.com

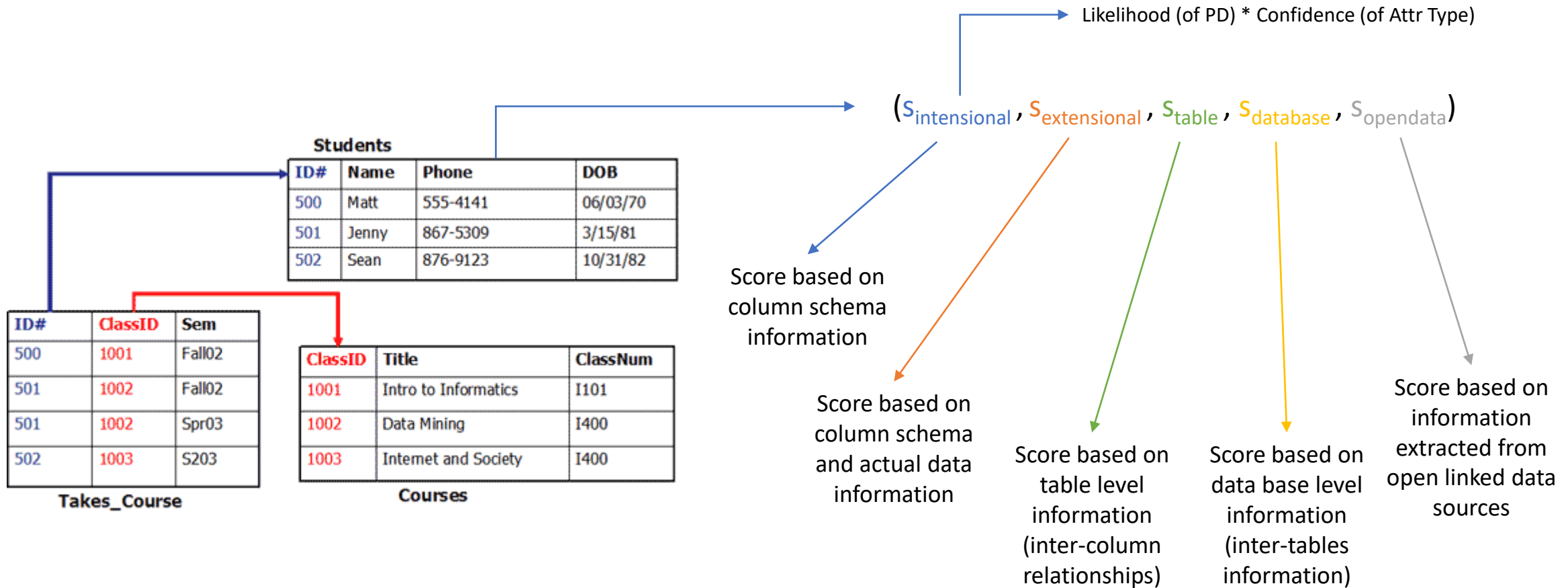
PDD Overview



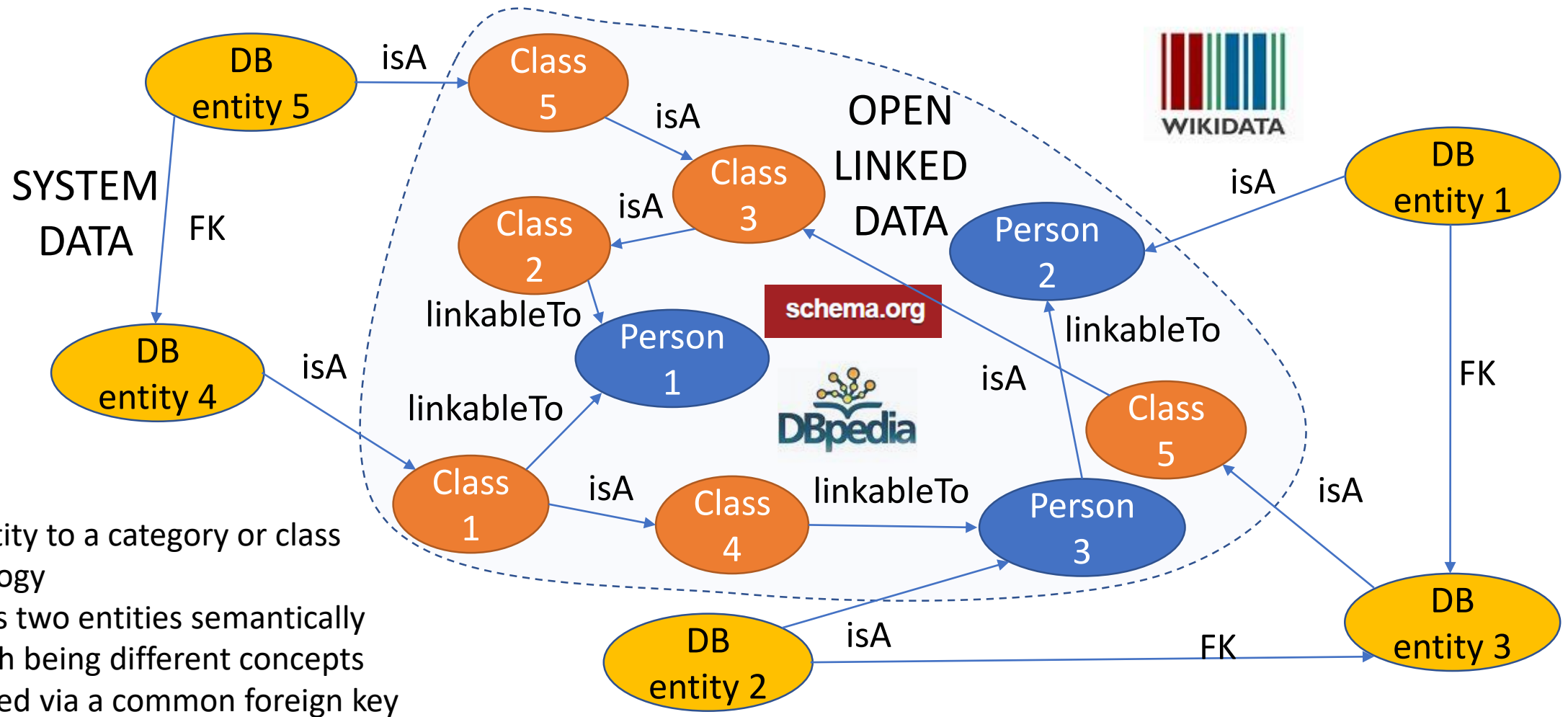
Onion model

PDD Output

➤ Every column is scored for each level:

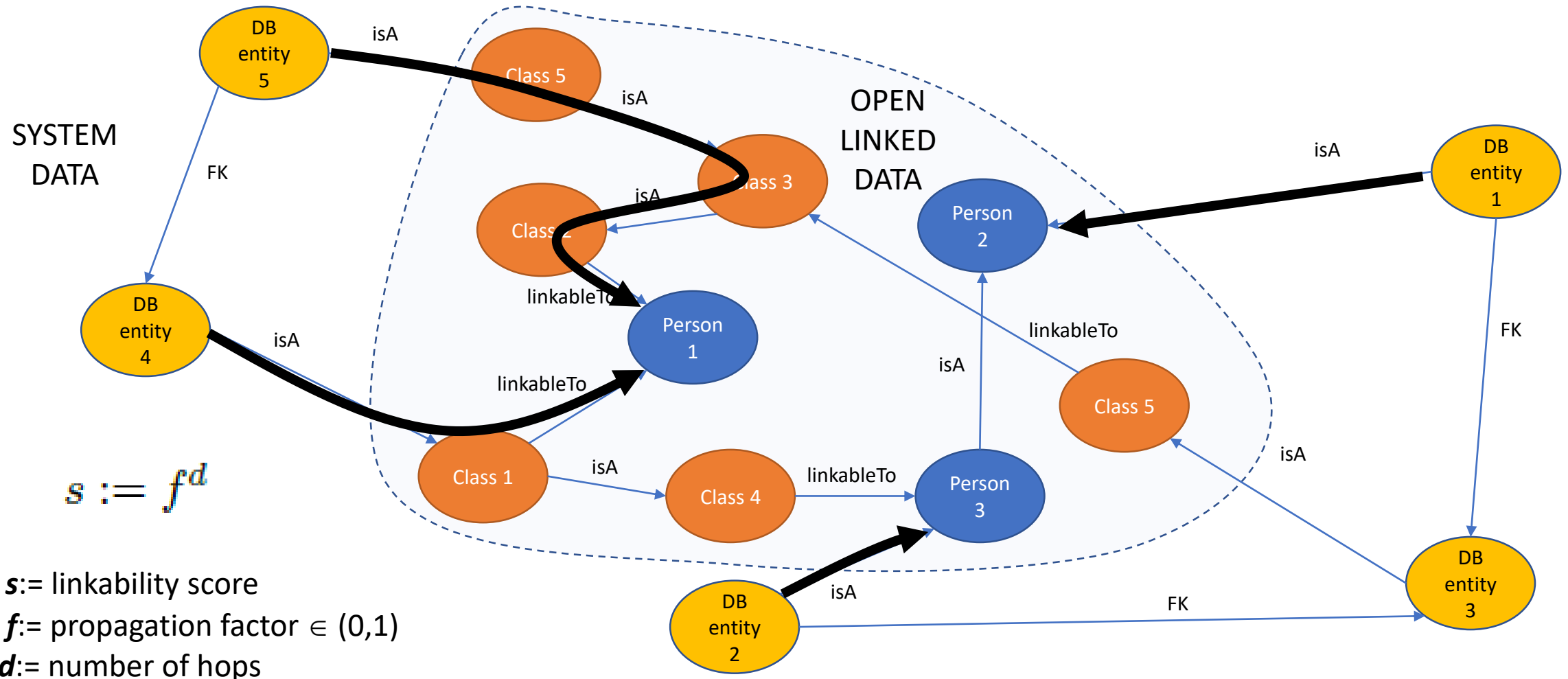


Leveraging Open Linked data: Graph Creation

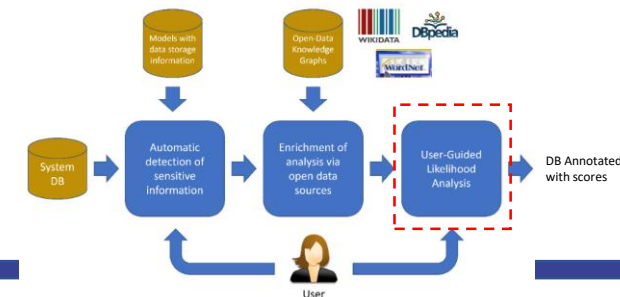


isA: links an entity to a category or class within an ontology
linkableTo: links two entities semantically related although being different concepts
FK: entities linked via a common foreign key

Leveraging Open Linked data: Graph Creation



User-Guided Likelihood Analysis



IDENTIFY

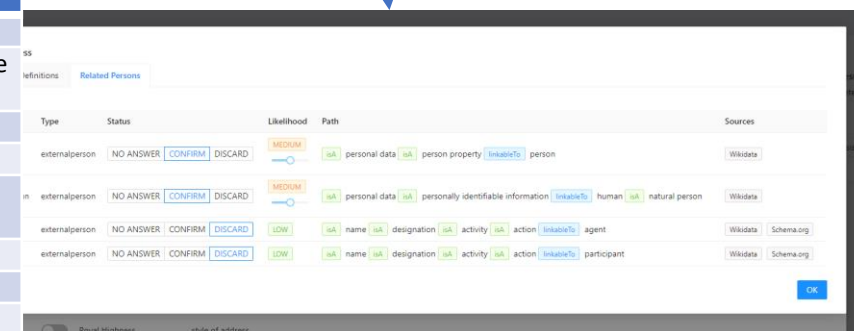
LINK

ASSESS

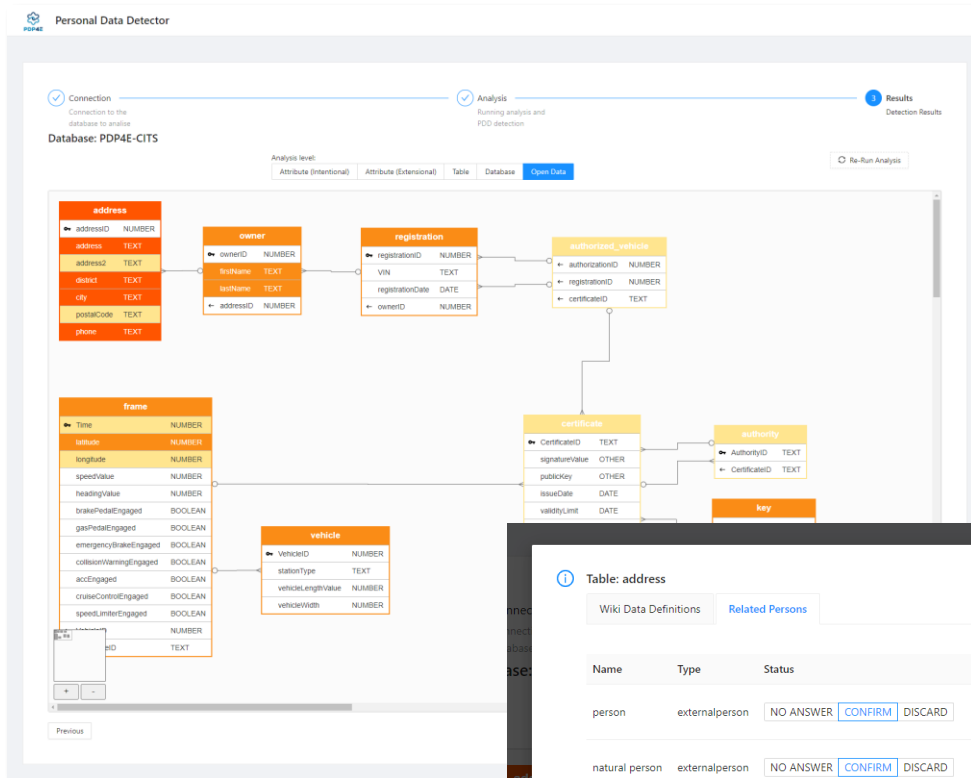
Original name	table	Originally proposed concept by Wikidata	Chosen by user	Original table name	Related persons
address		address (Q338075)	postal address (Q3196)	address	personal data -> person property -> person
owner		master (Q19357897)	Proprietor (Q1686912)	address	personal data -> personal identifiable information -> human -> natural person
registration		registration (Q2399307)	UNCHANGED	owner	legal person -> agent
authorized_vehicle		Emergency vehicle (Q1308737)	vehicle (Q42889)	registration	N/A
certificate		certificate (Q196756)	digital certificate (Q27)	authorized_vehicle	vehicle operator -> person
authority		authority (Q174834)	certificate authority (Q)	le	N/A
key		key (Q132041)	key (Q471771)	certificate	N/A
frame		frame (Q1324888)	UNCHANGED	authority	N/A
vehicle		vehicle (Q42889)	UNCHANGED	key	N/A
				frame	N/A
				vehicle	vehicle operator -> person

Automatically detected classes from open data are processed and refined by the user

Related persons are recalculated from new concepts refined by the user



User is allowed to decide the likelihood of an attacker to connect entities in the DB with external data subjects



Personal Data Detector

Table: vehicle

Wiki Data Definitions | Related Persons

Search Wiki Definitions

Wiki Data ID	Chosen	Name	Description
Q42889	☒	vehicle	mobile machine that transports people, animals or cargo
Q334166	☐	mode of transport	any form of vehicle or system used to transport people or goods from one place to another
Q13629441	☐	electric vehicle	vehicle propelled by one or more electric motors
Q29587	☐	traffic collision	collision of a vehicle with another vehicle, pedestrian, animal, or other object
Q752870	☐	motor vehicle	self-propelled wheeled vehicle
Q192152	☐	sport utility vehicle	type of automobile
Property:P1876	☐	vehicle	vessel involved in this mission, voyage or event
Q12876	☐	tank	heavy armored fighting vehicle
Q130368	☐	armored fighting vehicle	combat vehicle designed with both armament and armored protection
Q697175	☐	launch vehicle	rocket used to carry payload into outer space
Q1527901	☐	unmanned vehicle	vehicle without a person on board
Q22706	☐	vehicle registration plate	metal or plastic identification plate attached to a motor vehicle or trailer
Q484000	☐	unmanned aerial vehicle	aircraft without a human pilot aboard
Q43193	☐	truck	freight motor vehicle
Q193468	☐	van	covered transportation vehicle
Q1414135	☐	rolling stock	railway vehicles, both powered & unpowered
Q154015	☐	international vehicle registration code	code specifying the country in which a motor vehicle's vehicle registration plate was issued

OK

Table: address

Wiki Data Definitions | Related Persons

Name	Type	Status	Likelihood	Path	Sources
person	externalperson	NO ANSWER CONFIRM DISCARD	MEDIUM	isA personal data isA person property linkableTo person	Wikidata
natural person	externalperson	NO ANSWER CONFIRM DISCARD	MEDIUM	isA personal data isA personally identifiable information linkableTo human isA natural person	Wikidata
agent	externalperson	NO ANSWER CONFIRM DISCARD	LOW	isA name isA designation isA activity isA action linkableTo agent	Wikidata Schema.org
participant	externalperson	NO ANSWER CONFIRM DISCARD	LOW	isA name isA designation isA activity isA action linkableTo participant	Wikidata Schema.org

OK

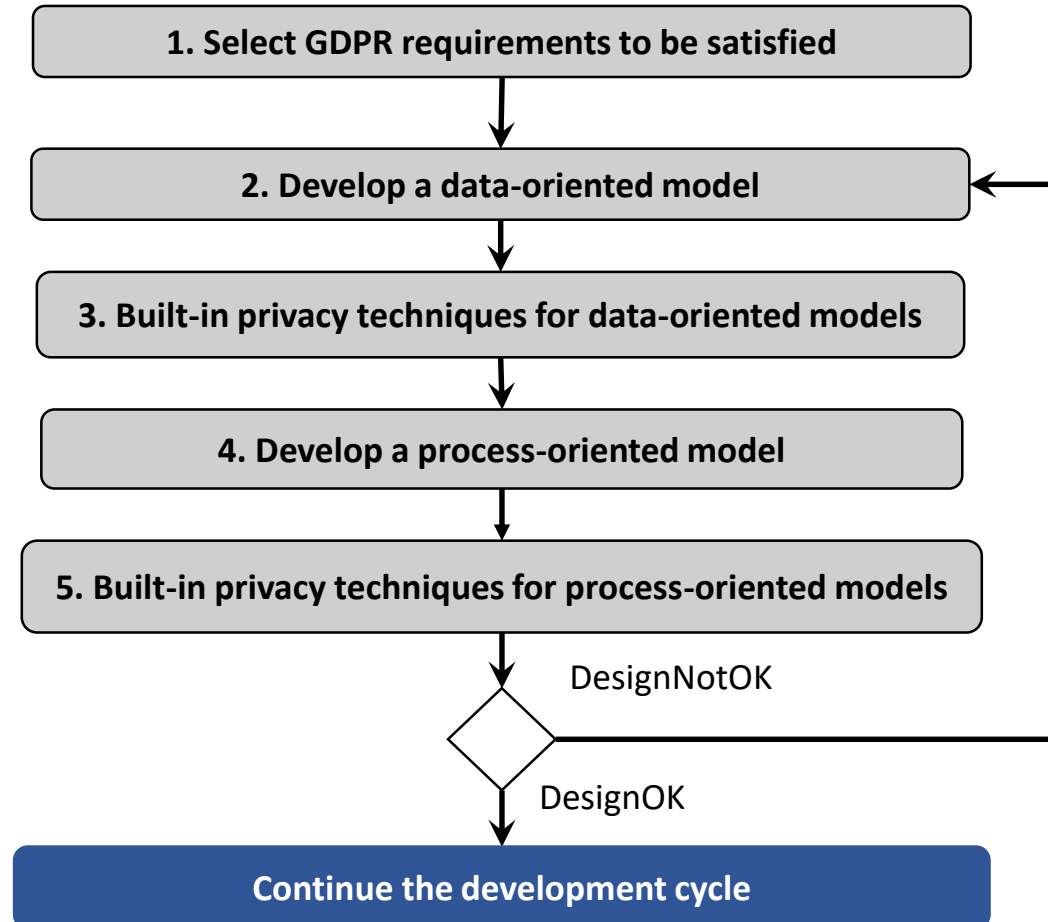
2. Privacy model-driven designer

Gabriel Pedroza (CEA)

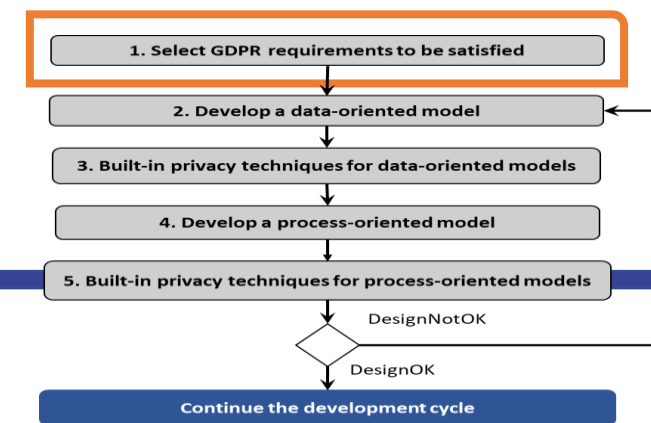
gabriel.pedroza@cea.fr

Privacy model-driven designer

➤ Implementation



1. Select GDPR requirements



- Goal: *select GDPR requirements to be fulfilled or analysed at the design phase*
- A model-driven interface amenable to:
 - ☐ Incorporate privacy and GDPR requirements
 - ☐ Keep traceability of requirements to be fulfilled (functional, GDPR)
- Model-driven tool support: *interoperable MDE interfaces requirements-design*
 - ☐ Feature 1: set links to allocate GDPR requirements to design (dependencies)
 - ☐ Feature 2: set links for satisfiability <<satisfy>>
 - ☐ Feature 3: set links for unitary test cases <<verify>>

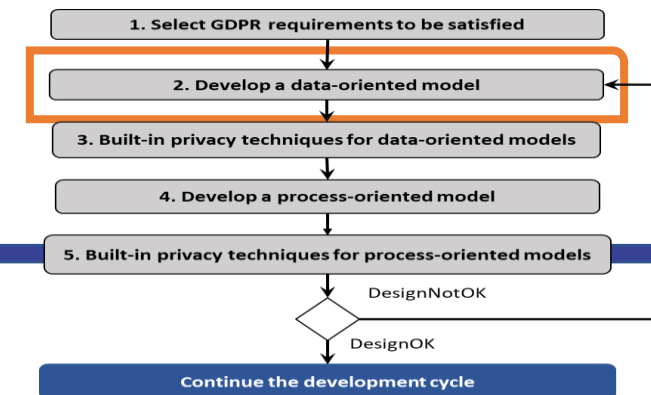


Notifications. This feature is meant to ensure the respect of the Data Subject rights, in particular, the right to be informed by the respective Controllers (or Processors) whenever a privacy breach impacting her/his Personal Data occurs.



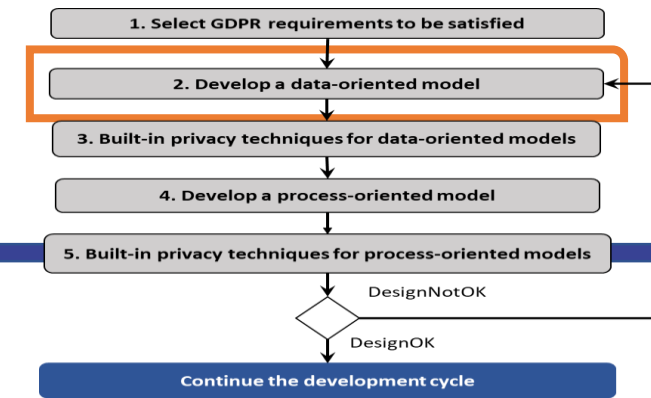
GDPRReq. When the <CITSFrame> breach is likely to result in a high risk to the rights and freedoms of <VehicleOwner>, the <RSUServiceProvider> shall communicate the <CITSFrame> breach to the <VehicleOwner> without undue delay.

2. Develop data-oriented model

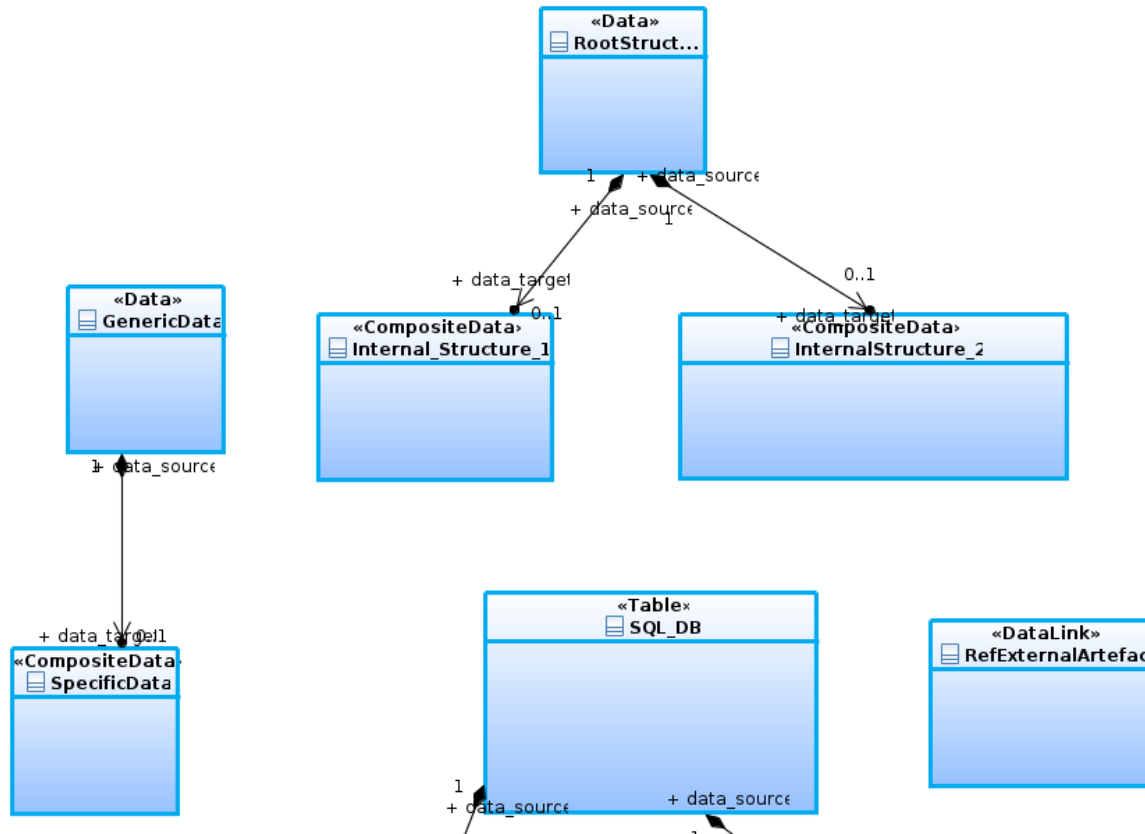


- Goal: *capture the data structures under study to analyse conformity w.r.t. privacy precepts*
- A modeling language amenable to:
 - ❑ Reuse outcomes from the PDD: scores for classifying personal (non-personal) data
 - ❑ Enrich, decompose, refine data structures
- Model-driven tool support: *a UML Class-like diagram to model data structures*
 - ❑ Feature 1: several built-in data types : *Generic, Composite, Table, Data links, Opaque data*
 - ❑ Feature 2: user defined data structures (suitable for framework customization)
 - ❑ Feature 3: full compatibility with SysML Requirement models
 - ❑ Feature 4: inherited traceability with GDPR requirements (PDP4E-Req tool)

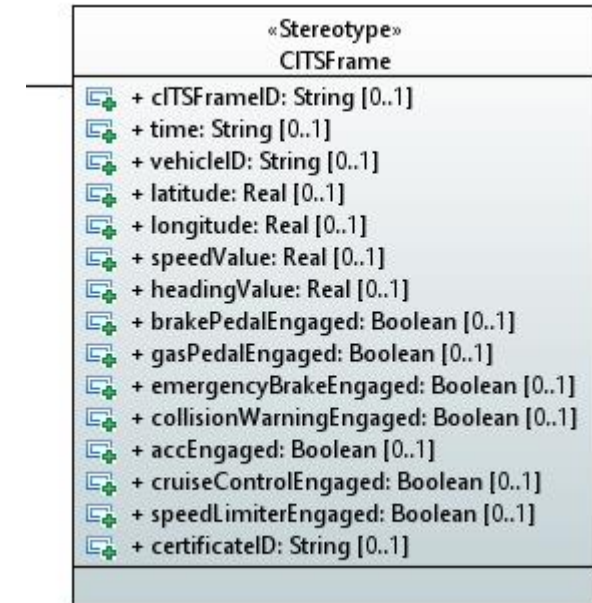
2. Data-oriented model overview



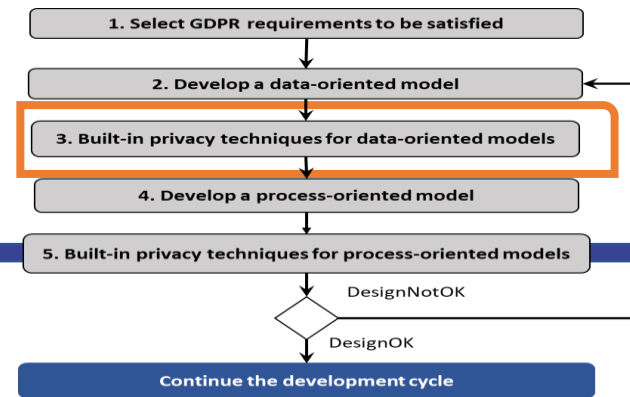
➤ Overview of a data-oriented model



User-defined data type ➔



3. Strategy for data-oriented model



- Goal: apply known strategies to ensure data protection
- Data-oriented strategies proposed by ENISA, ISO/IEC-27550

- ☐ Minimize
- ☐ Separate
- ☐ Abstract
- ☐ Hide

- Model-driven tool support: catalogue of strategies

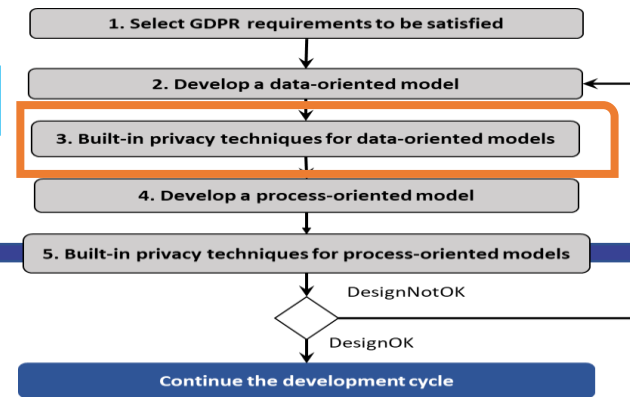
- ☐ Feature 1: strategies to Abstract data ; K-anonymity
- ☐ Feature 2: strategies to Minimize data ; α -anonymity
- ☐ Feature 3: import data structures, e.g., raw tables
- ☐ Feature 4: import data from schema, e.g., data base schema

Design strategy		Description	Privacy control examples
Data oriented strategies	Minimize	Limit as much as possible the processing of PII	Selection before collection Anonymization
	Separate	Distribute or isolate personal data as much as possible, to prevent correlation	Logical or physical separation Peer-to-peer arrangement Endpoint processing
	Abstract	Limit as much as possible the detail in which personal data is processed, while still being useful	Aggregation over time (used in smart grids) Dynamic location granularity (used in location based services) k-anonymity
	Hide	Prevent PII from becoming public or known.	Encryption Mixing Perturbation (e.g. differential privacy, statistical disclosure control) Unlinking (e.g. through pseudonymisation) Attribute based credentials

$$I(T_1 \cup T_2; \alpha * QI) = H((T_1 \cup T_2) \cap \alpha * QI) \quad (2)$$

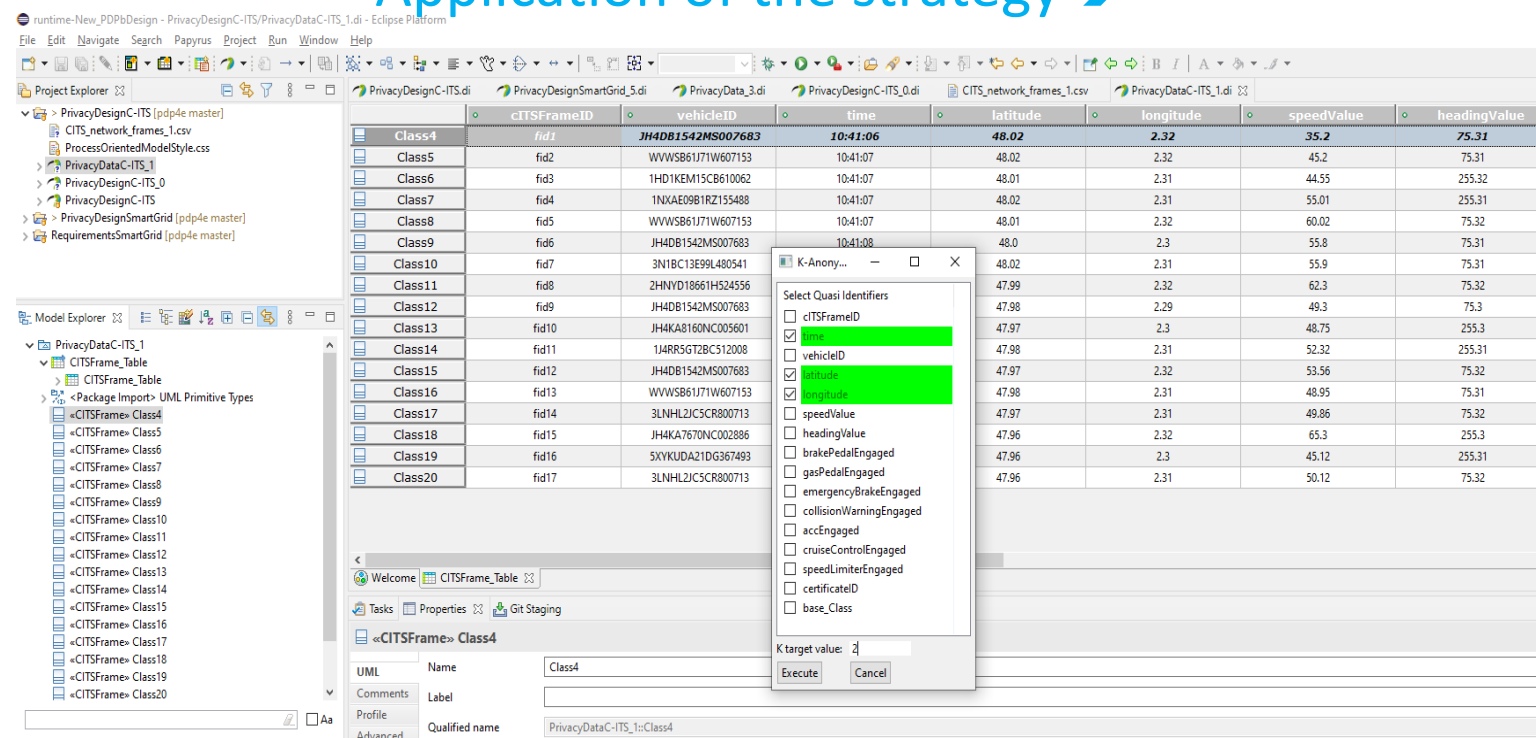
$$H(\alpha * QI) = \sum_{t \in QI} \sum_{t \in C_j} \alpha_i P(X = t) \log_b(P(X = t)) \quad (3)$$

3. Strategy for data-oriented model



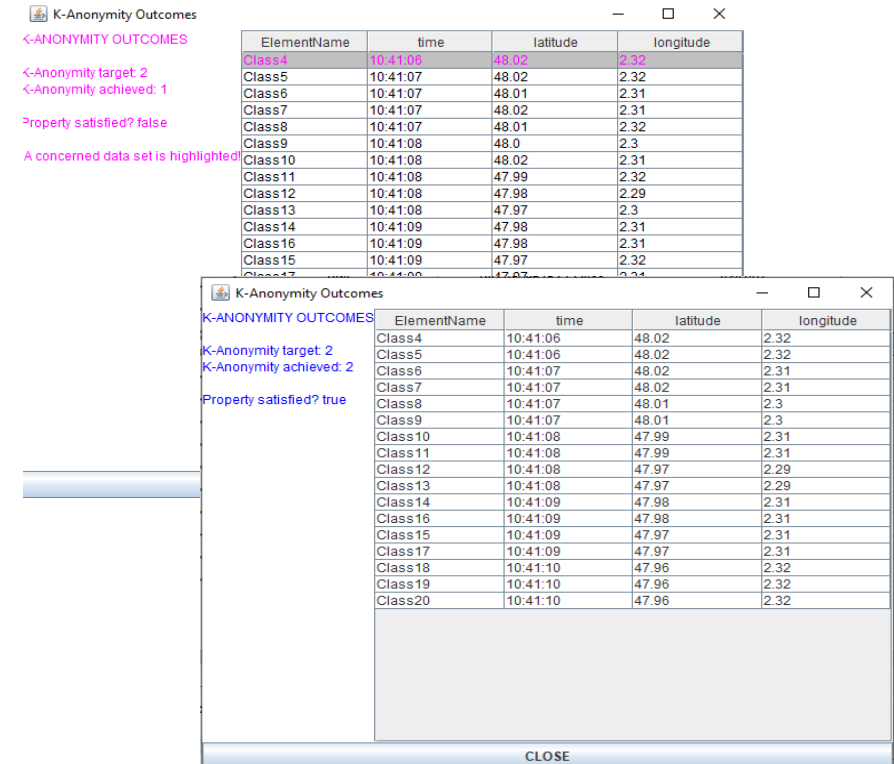
➤ Overview of data-oriented strategies Application of the strategy ➔

Strategy outcomes ➔



The screenshot shows the Eclipse IDE with the following components:

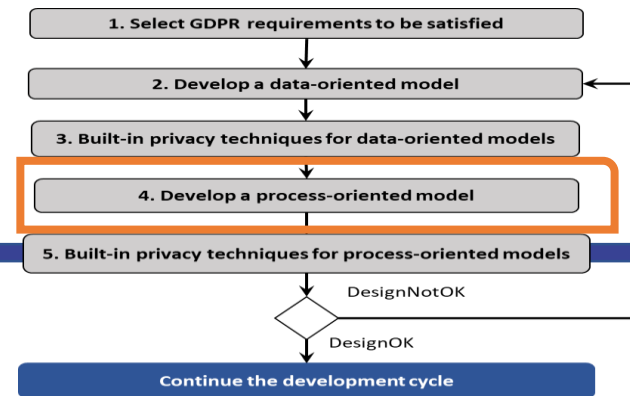
- Project Explorer:** Shows the project structure including `PrivacyDesignC-ITS` and its sub-projects.
- Model Explorer:** Displays the UML model for `Class4` within the `PrivacyDataC-ITS_1` package.
- Class4 Table:** A table with columns: `Class4`, `fid1`, `vehicleID`, `time`, `latitude`, `longitude`, `speedValue`, and `headingValue`. It lists 20 classes (Class4 to Class20) with their respective identifiers and values.
- K-Anonymity Dialog:** A dialog box titled "K-Anonymity..." with a "Select Quasi Identifiers" section. It lists attributes like `time`, `latitude`, `longitude`, `speedValue`, `headingValue`, `brakePedalEngaged`, `gasPedalEngaged`, `emergencyBrakeEngaged`, `collisionWarningEngaged`, `accEngaged`, `cruiseControlEngaged`, `speedLimiterEngaged`, `certificatedID`, and `baseClass`. The `time`, `latitude`, and `longitude` attributes are selected. The "K target value" is set to 2.



The screenshot shows the "K-Anonymity Outcomes" window with the following information:

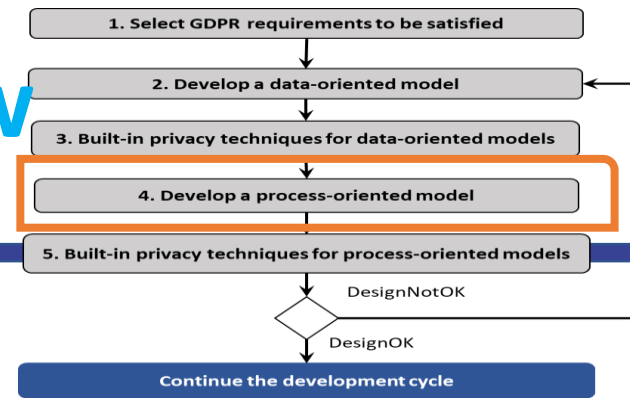
- K-ANONYMITY OUTCOMES:** A table with columns: `ElementName`, `time`, `latitude`, and `longitude`. It lists 20 classes (Class4 to Class20) with their respective values.
- K-Anonymity target: 2**
- K-Anonymity achieved: 1**
- Property satisfied? false**
- A concerned data set is highlighted**

4. Develop process-oriented model

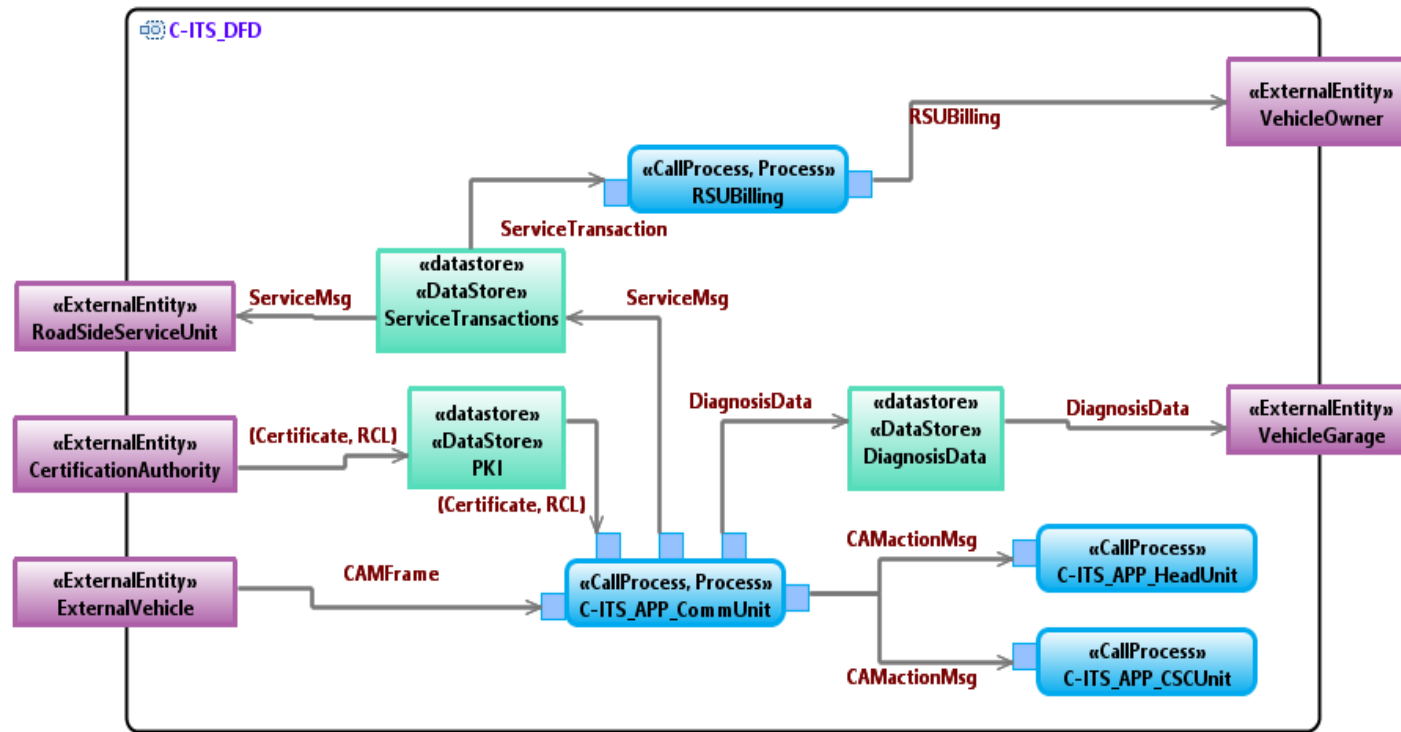


- Goal: *capture the data flows and processes under study to analyse conformity w.r.t. privacy precepts*
- A modeling language amenable to:
 - ❑ Support Data Flow Diagrams (DFD)
 - ❑ Incorporate aspects related to privacy and data protection by design
- Model-driven tool support: *UML Activity-like diagram to model processes & data*
 - ❑ Feature 1: DFD profile: External Entity, Process, Data flow, Data storage, Ports
 - ❑ Feature 2: Reusability of data-oriented structures (to type Ports)
 - ❑ Feature 3: Full compatibility with PDP4E-Req models (inherited traceability)
 - ❑ Feature 4: Leverage GDPR profile

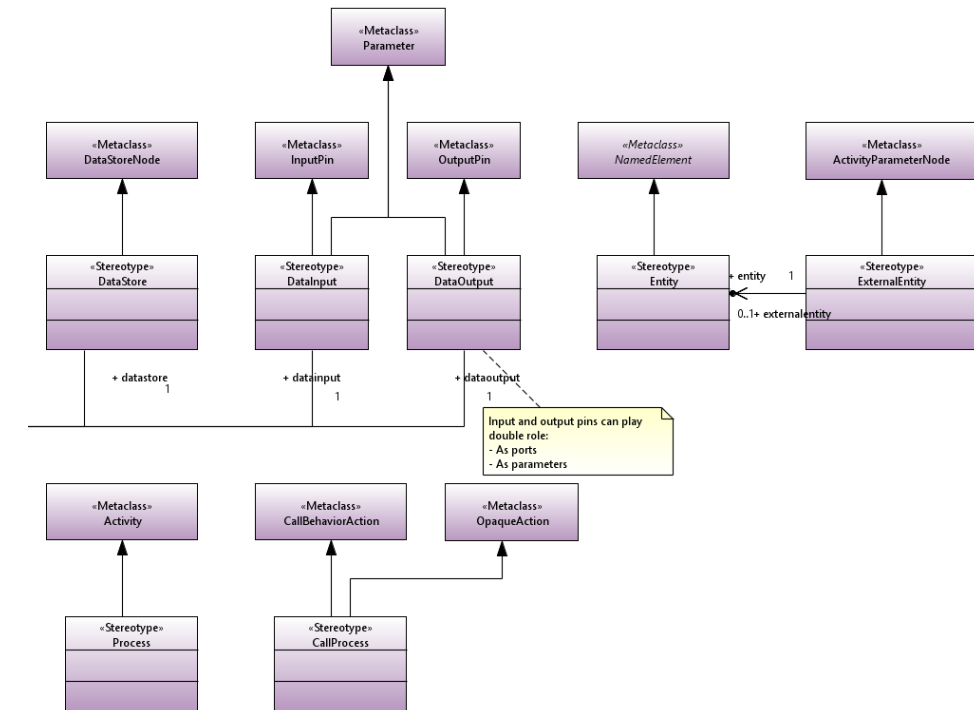
4. Process-oriented model overview



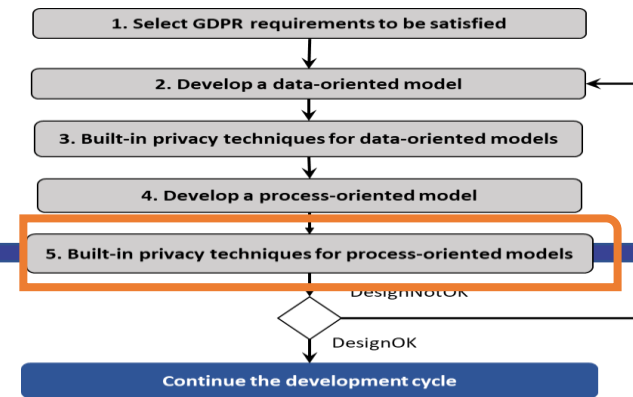
➤ Overview of a process-oriented model (DFD)



DFD profile ➔



5. Apply process-oriented strategy



- Goal: *apply known privacy strategies to improve DFD model*
- Process-oriented strategies proposed by ENISA, ISO/IEC-27550

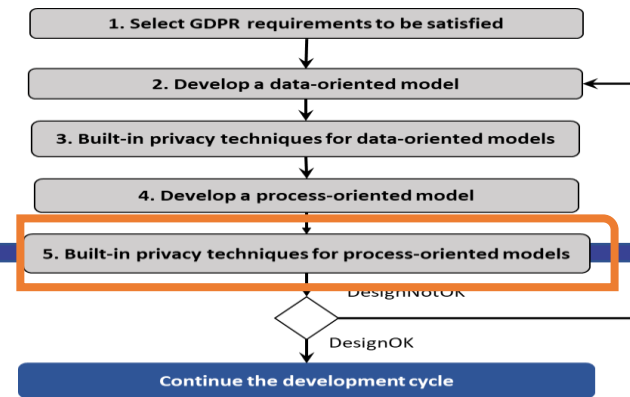
- ☐ Inform
- ☐ Control
- ☐ Enforce
- ☐ Demonstrate

- Model-driven tool support: catalogue of strategies

- ☐ Feature 1: strategies to Control ; Consent pattern
- ☐ Feature 2: strategies to Inform ; Data breach notification
- ☐ Feature 3: import/export DFD models from Privacy Risk Management (PRM) tool
- ☐ Feature 4: dedicated profile to support privacy threat conditions (PRM)

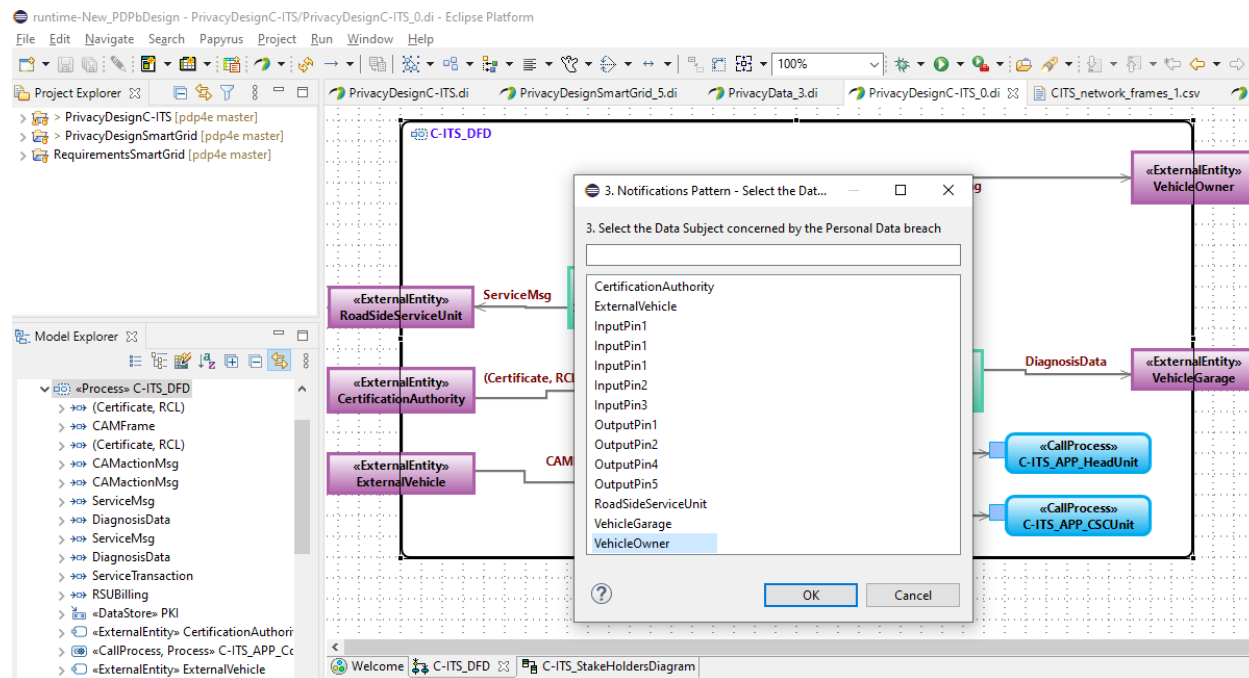
	Inform	Inform PII principals about the processing of PII	Privacy icons Layered privacy policies Data breach notification
Process oriented strategies	Control	Provide PII principals control over the processing of their PII.	Privacy dashboard Consent (including withdrawal)
	Enforce	Commit to PII processing in a privacy friendly way, and enforce this	Sticky policies and privacy rights management Privacy management system Commitment of resources Assignment of responsibilities
	Demonstrate	Demonstrate that PII is processed in a privacy friendly way.	Logging and auditing Privacy impact assessment Design decisions documentation

5. Apply process-oriented strategy

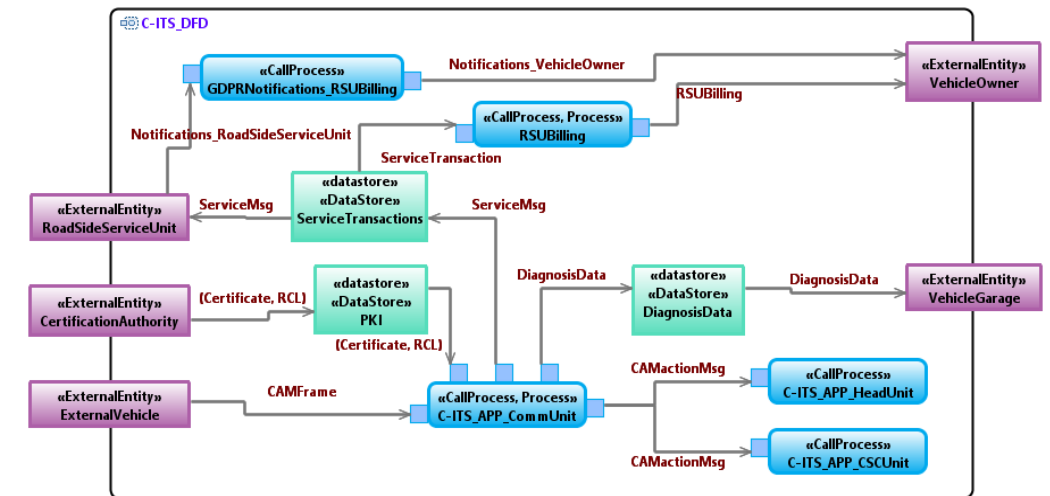


➤ Overview of process-oriented strategies

Application of the strategy ➔



Outcome of the strategy ➔





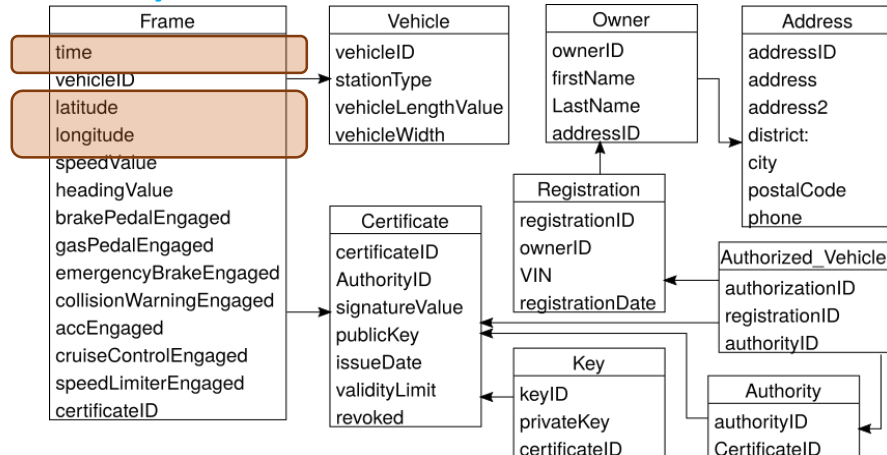
Automotive case study: C-ITS

Personal data detection outcomes →

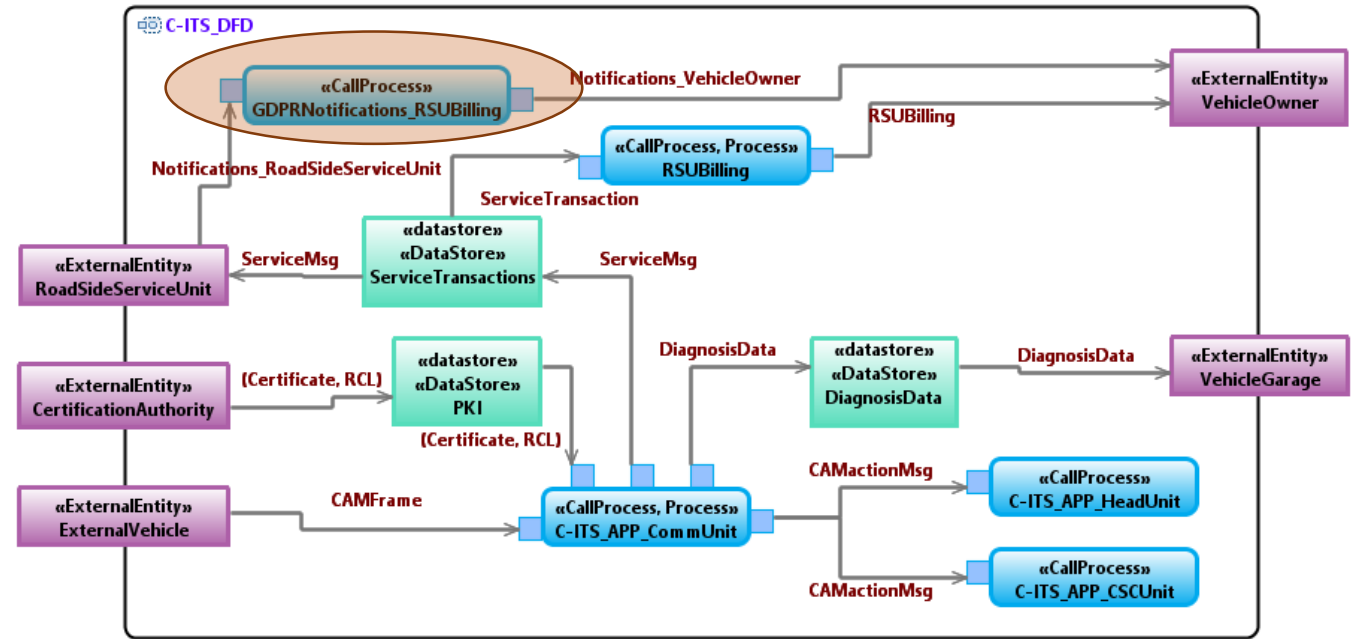
Location	Table	Score (s)	Open Data Score
PKI	Owner	High	High
	Address	High	High
	Registration	High	High
	Authorized Vehicle	Medium	High
Shared	Certificate	Medium	High
	Authority	Low	Medium
	Key	Low	Medium
Network	Vehicle	Low	High
	Frame	Low	Medium

TABLE 2. PERSONAL DATA IDENTIFICATION SCORES

Privacy assessment on data →



Privacy-aware modelling →



Privacy assessment on DFDs →

Summary of achievements

- PDPbD modules released as open-source:
 - ❑ Privacy designer:
 - <https://git.eclipse.org/c/papyrus/org.eclipse.papyrus-privacydesigner.git/>
- PDPbD framework implements the methodology to realize “privacy by design”:
 - ❑ Method to incorporate knowledge from three domains:
 - Systems engineering
 - Privacy and Data Protection
 - Regulations like GDPR and standards like ISO/IEC 27750
 - ❑ Tool support for the method via three modules:
 - Personal Data Detection
 - Privacy model-driven designer
 - Code validation
 - ❑ Proof of concept validated via an automotive case study

Acknowledgements



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 787034.

<https://www.pdp4e-project.eu/>

Privacy and Data Protection 4 Engineering

For more information, visit:
www.pdp4e-project.org

Thank you for your attention

Questions?

gabriel.pedroza@cea.fr
victor.muntes@beawre.com