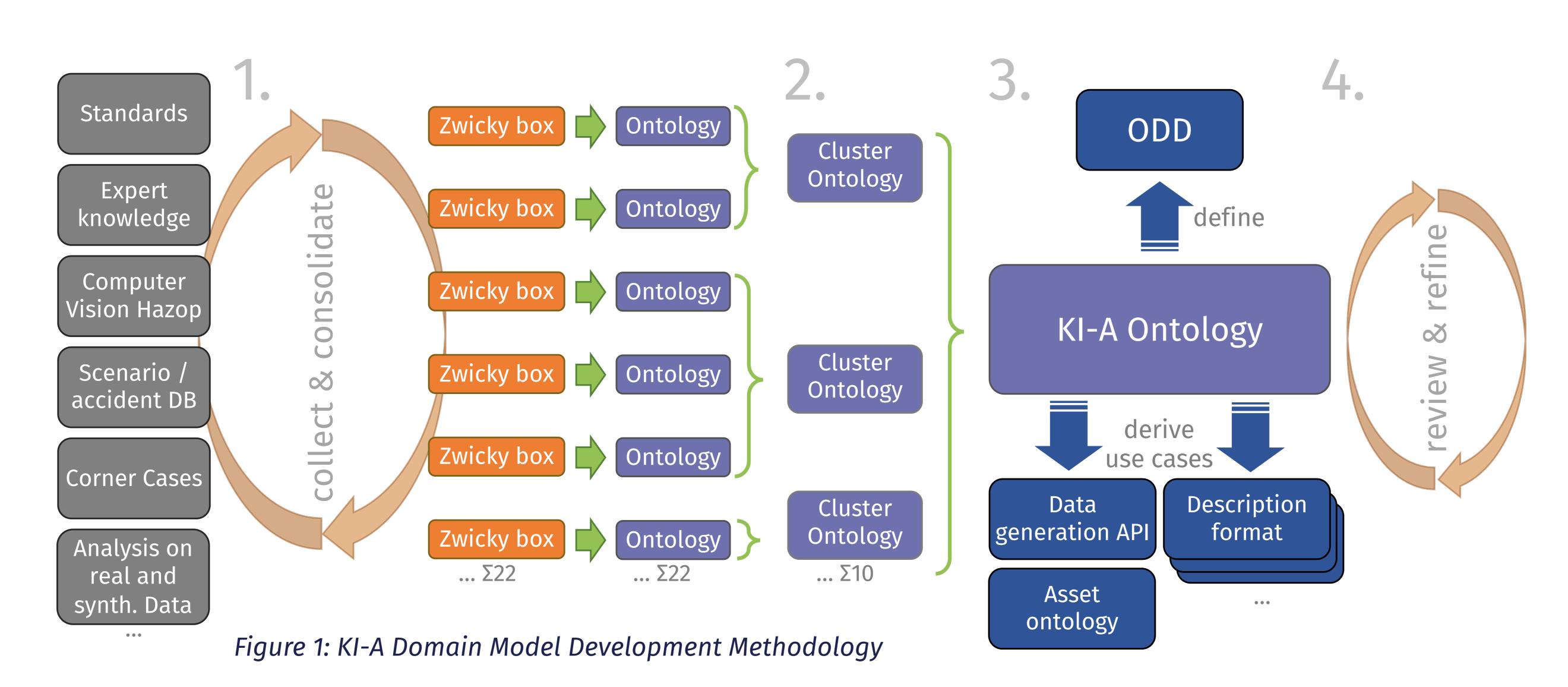


Methodology of Creating an Ontology for Dataset Engineering

Christian Witt¹, Martin Herrmann², Christian Heinzemann², Frédérik Blank²



Towards a domain description language

A base for a safety argumentation for an DNN function is a detailed description of the input domain. It forms a base for the design and analysis of datasets and safety related tests. Our strategy spans the entire process from analyzing the input domain until the derivation of the actual ontology.

1. Domain Analysis

We review public data sources like standards and law documents regarding specifications of traffic scenarios, traffic infrastructure, surface properties, hazard and accident analysis as well as corner cases of DNNs to classify base scenarios. In addition, we initiated multiple brainstorming workshops with experts from different fields of expertise (e.g. DNN, simulation, sensor, safety) to identify relevant context dimensions. Hereby, context dimensions are elements or properties that can be varied to achieve a specific effect on a DNN function like weather conditions or degree of occlusion.

2. Structuring of Clusters

After some intermediate representations we structured all context elements into several

ontologies that bundle multiple clusters.

A prime example for a shared context element is color, since color is used many elements like hair, object surfaces or clothes. For this reason, the description of color has been factored out into a separate cluster.

3. Consolidation of Context Dimensions

To improve the usage in ODD or coverage analysis, we tried to parameterize each dimension into discrete alternatives. This was based on physical quantities if possible or expert knowledge. The refined combination of dimensions and alternatives defined a vocabulary in human understandable terms that build a perfect bridge between different fields of expertise, but also between humans (e.g. data requester) and machines (e.g. simulation toolbox).

4. Refinement by Data Analysis

The final consolidation has two major goals: identifying shared context elements and grouping coherent information while factoring out shared information. Performing such a consolidation is key for obtaining a usable and well-structured input domain model.

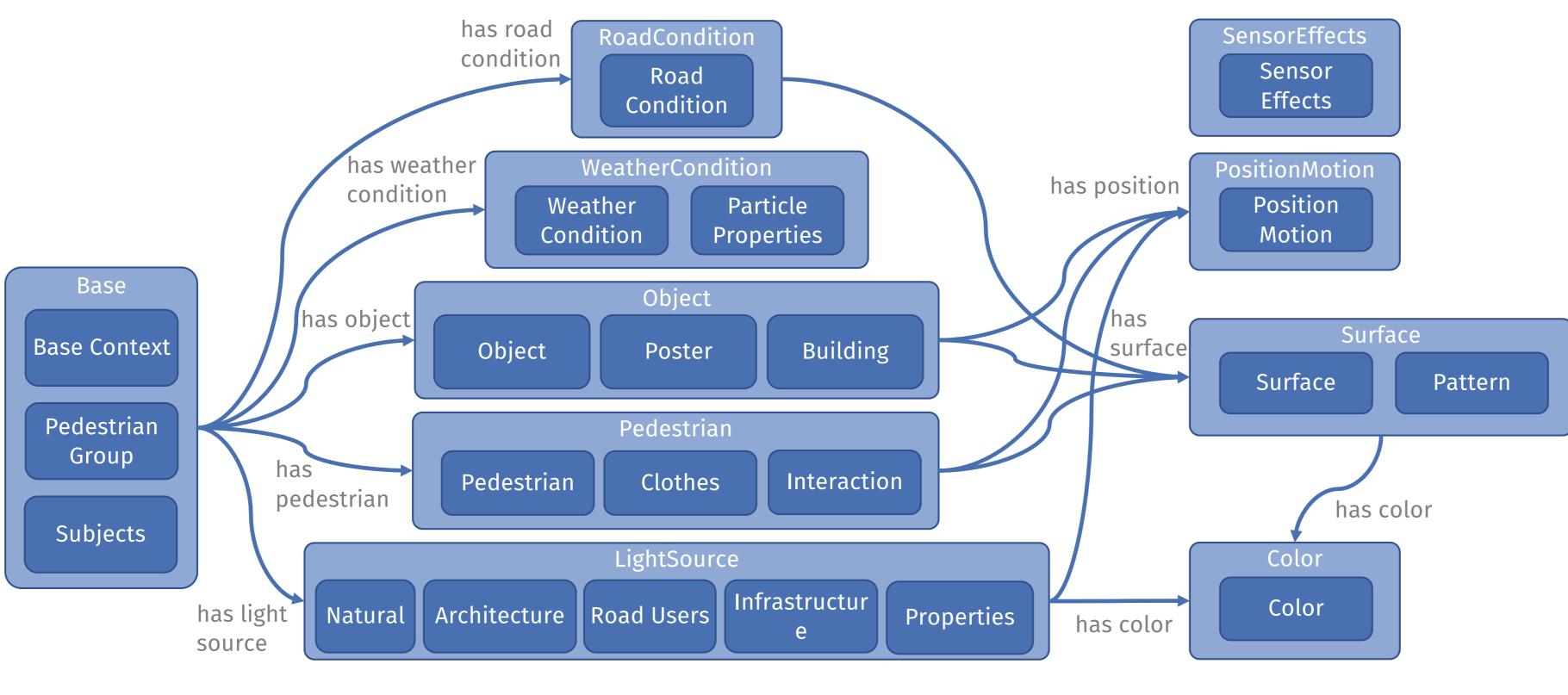


Figure 2: KI-A Ontology: 22 ontologies bundled into 10 clusters

References:

Herrmann, M. et al. "Using ontologies for dataset engineering in automotive Al applications", DATE, 2022.





ROSCH















For more information contact: christian.witt@valeo.com martin.herrmann@de.bosch.com christian.heinzemann@de.bosch.co frederik.blank@de.bosch.com

Related Posters: Using Ontologies in Automotive AI Applications /
Input Coverage Analysis using Domain Models and Combinatorial Testing

KI Absicherung is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.





Supported by:

on the basis of a decision by the German Bundestag