THE CREATION AND APPLICATION OF HARMONIZED PRE-CRASH SCENARIOS FROM GLOBAL TRAFFIC ACCIDENT DATA

Henrik Liers
Marcus Petzold
Institute for Traffic Accident Research at Dresden University of Technology (VUFO)
Germany

Harald Feifel
Germany

Jörg Bakker
Asymptotic AI
Sweden

Ernst Tomasch
Graz University of Technology
Austria

Paper Number 23-0266

ABSTRACT

The development and test of future Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD) AD functions requires sophisticated data from pre-crash scenarios. As real-world traffic provides an infinite variety of scenarios and vehicles are usually sold in many markets, valuable simulation datasets from several countries seem indispensable. The paper describes how we combined the format of the Pre-Crash Matrix (PCM) with global accident data from IGLAD. The goal was to create harmonized pre-crash simulation files from real accidents coming from several countries/continents and to use them exemplarily within a field-of-view analysis for future ADAS.

The basic data source is the IGLAD database. Within the “Initiative for the Global Harmonization of Accident Data” (IGLAD) traffic safety researchers from Europe, North America, South America, Asia, and Australia bring together road accident data in a harmonized dataset. Each single accident is reconstructed and contains relevant information like vehicle data, injury severities, anthropometric data, and scaled sketches.

The PCM format describes the vehicle dynamics (trajectories) in a defined time before the collision. It is similar to the OpenX formats and contains relevant information about the road layout, markings, view obstacles, etc.

The paper describes the process of creating IGLAD-PCM data, including the establishment of requirements, the harmonization of country-specific characteristics, and the definition of quality features.

In 2022, IGLAD-PCM was released for the first time providing 200 pre-crash simulations from real accidents coming from seven countries on three continents. The paper presents descriptive statistics (e.g. accident characteristics, accident configurations, injury severities) from these cases and a comparison to the current IGLAD dataset (with approximately 9,400 accidents from 10 different countries). We provide an overview of relevant accident situations and country-specific characteristics for different regions of the world, e.g. US, India, China, Germany, France, Italy, etc.

The paper also highlights the benefit of PCM data as one essential source for data-driven system development. During the concept definition of safety systems, pre-crash trajectory data is used to derive the required functional behavior. First, the relative positions and orientations of other traffic opponents are the basis for defining the necessary sensor field-of-view in given accident scenarios. Second, the speed distributions of ego and opponent serve as key performance indicators for the vehicle actuation system. Here, a relevant accident scenario is discussed, and relevant regional differences analyzed.

The IGLAD-PCM forms a unique global dataset of pre-crash simulations based on reconstructed traffic accidents. Of course, case numbers are quite low at this early stage, but will increase annually by more than 200. Using the data can enhance the development of ADAS and AD functions and help to adjust systems towards country-specific characteristics.

We have demonstrated that the PCM allows to harmonize pre-crash data from different countries and still can cover regional specifics. As the PCM is an open data format, various scenario descriptions can easily be generated, and existing development tool chains can be supported. Thus, we believe that the PCM can serve as a standard format for data-driven system development and simulation.

RESEARCH QUESTION / OBJECTIVE

More than 10 years after its initiation, the IGLAD project database has firmly established itself as an in-depth data source for accident research and vehicle safety applications. More than 9,400 accidents from five continents have now found their way into the database. They all share a uniform coding, harmonized between the data providers, as well as a quality standard. With the current data, it is already possible to perform descriptive analyses of accidents and injuries in various countries.

Current developments in the field of vehicle and road safety are focusing on the topics of connected and automated driving. This requires complex development, testing, and validation processes, which in turn require suitable input data. For ADAS and AD functions, accident and traffic scenarios play a decisive role. These scenarios are usually stored in open formats and are mostly created generically. Rarely do they originate from real accidents.

In this paper we have addressed the question whether it is possible to transfer the originally heterogeneous accident data of different data providers into a uniform, open and usable scenario format. For this purpose, different actors from the IGLAD consortium have joined forces to establish respective tool chains, processes, and data exports to transfer the pre-crash phases of IGLAD accidents into the PCM format. The main goal is to continuously provide pre-crash / scenario data from IGLAD in future project phases. Initially, this will be in PCM format, with OpenX formats also being considered later.

METHODS AND DATA SOURCES

The IGLAD project

Since the last report at the ESV conference 2017 (1) there have been numerous changes in the IGLAD project and improvements to the data set. This paper describes one of the major changes, the introduction of the IGLAD-PCM data. But before diving into this new feature we want to give a short general overview and update on the status of the IGLAD project.

IGLAD is an international in-depth accident data project and consortium that was initiated by Daimler AG, ACEA and different research institutes and announced as a working group at the FIA Mobility Group in October 2010 (2) (3) (4). The goal of the group was to define a common in-depth accident data standard and provide a yearly data set that is created from international partners that are part of the consortium. The non-profit project is completely self-funded and offers its data to interested parties to be used for research purposes charging a moderate membership fee. The data includes information that is common to most in-depth accident studies, like:

- accident time, description, type, influencing factors, …
- participant type, vehicle data, reconstruction data, …
- occupant age, weight, gender, injury severity, …
- safety system types, activation, …

The codebook of the common in-depth accident data scheme is freely available on the web page of the project (5). Besides the data tables, the IGLAD database also includes a scaled sketch for each accident in a common vector format.

The IGLAD consortium currently consists of 24 members complemented by the chair and administrator from Chalmers University and the SAFER Vehicle and Traffic Safety Centre at Chalmers (6) in Sweden. The project will celebrate its 10th yearly data release in June 2023. The currently released 9th data set includes 9,425 cases in total contributed by 14 data providers located in North America, South America, Europe, Asia, and Australia. The data originates from 12 different countries: Australia (AU), Austria (AT), Brazil (BR), China (CN), Czech Republic (CZ), France (FR), Germany (DE), India (IN), Italy (IT), USA (US), Spain (SP), and Sweden (SE). New countries, data providers, and members are expected to join the project in the near future. The goal is to cover as many regions/countries of the world as possible.

IGLAD dataset

In principle, the compiled IGLAD data set contains only accidents with personal injuries. Only in the pilot phase (Phase I) the data providers exceptionally were allowed to include accidents with property damage. These accidents now account for only 0.5% of the total IGLAD dataset. By contrast, the much larger proportion of accidents was initially biased toward serious and fatal accidents. This results from the fact that some data providers obtain their original data from accident reports, which are preferably commissioned only for severe and fatal accidents. Other data collections are primarily focused on fatality accidents. Over the various IGLAD phases, intensive work was therefore carried out to detect and successively eliminate possible system-inherent or random biases in the data sets (for further explanations, see chapter Representativeness).

Figure 1 shows the maximum injury severities of the accidents in the current IGLAD data set (Phase IV / 2021) per country. Also included are the numbers of cases, which differ between 50 and 1,800. The background of these strong differences is that some countries or data providers have only provided data once, while other countries have provided data continuously each year since the beginning.

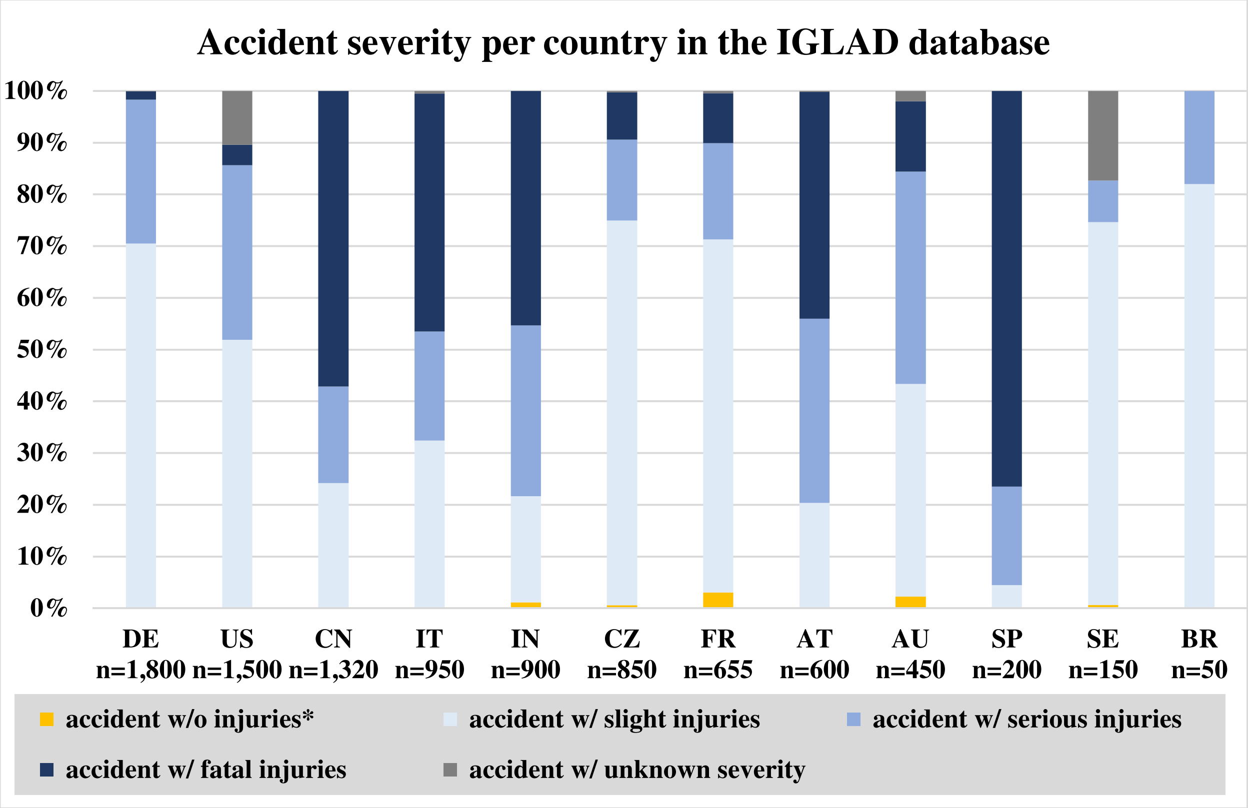


Figure 1. **Distribution of accident severity per country in the IGLAD database (Phase 4 / 2021).**

The diagram shows that among the regular data providers there are still some that deliver a comparatively high proportion of fatal accidents, e.g. China, Italy and India. On the other hand, many other data providers made some good progress and are now able to select accidents close to the national statistics to become representative gradually. A good example for this development is Austrian data, provided by TU Graz, see Figure 2.

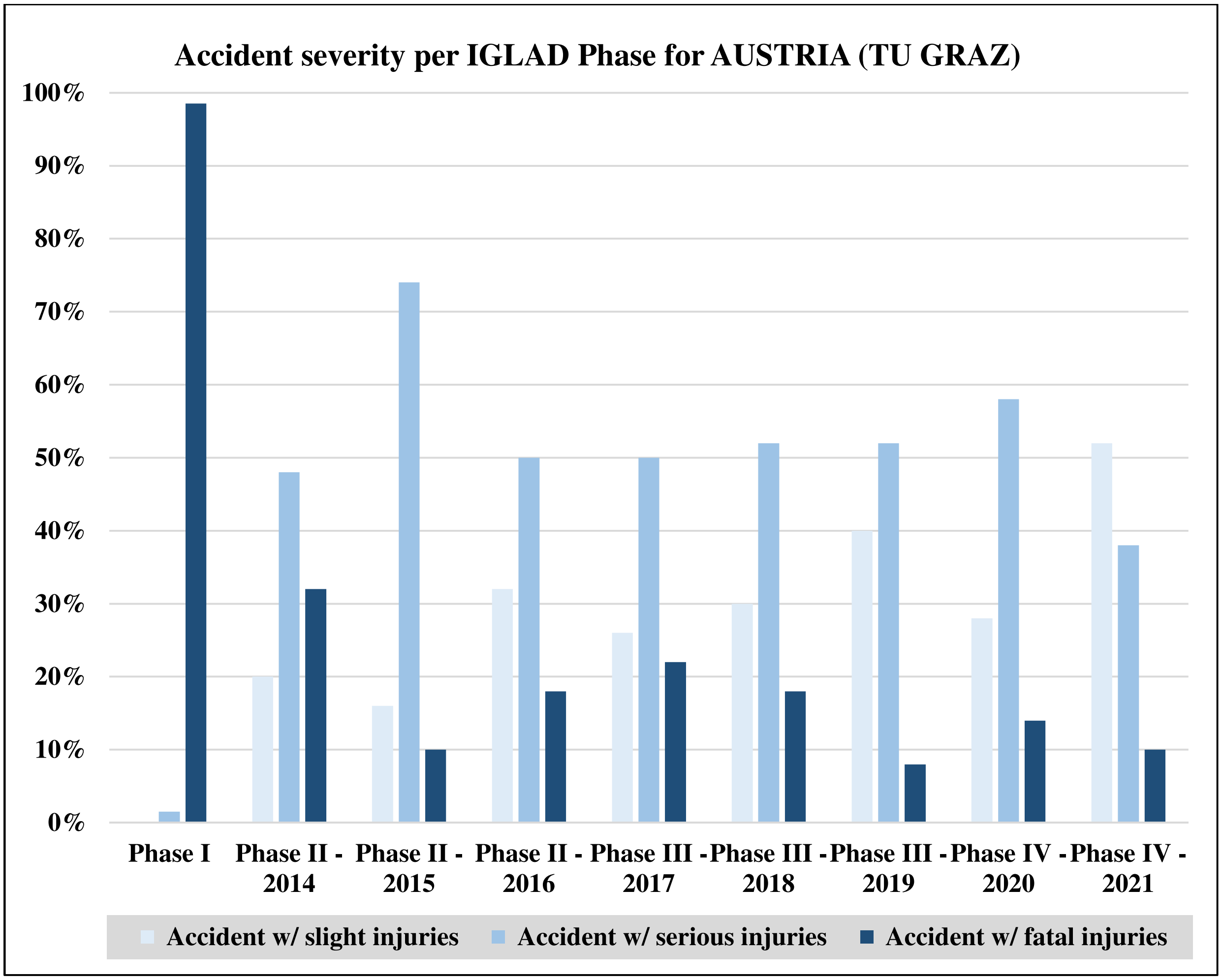


Figure 2. **Distribution of accident severity per IGLAD phase in Austrian cases (Data Provider: TU Graz).**

Although few data sources still have biases due to specific sampling criteria, the accidents in the IGLAD database cover all accident types and accident configurations. As an example, Figure 3 shows the distribution of main accident types in the IGLAD cases per country. The main accident type describes the conflict/critical situation which resulted in the accident.

It can be seen that, although accident severities are quite different between many countries, the main accident types are rather similar, especially for countries in similar regions (e.g. Western Europe with AT, CZ, DE, FR, IT).

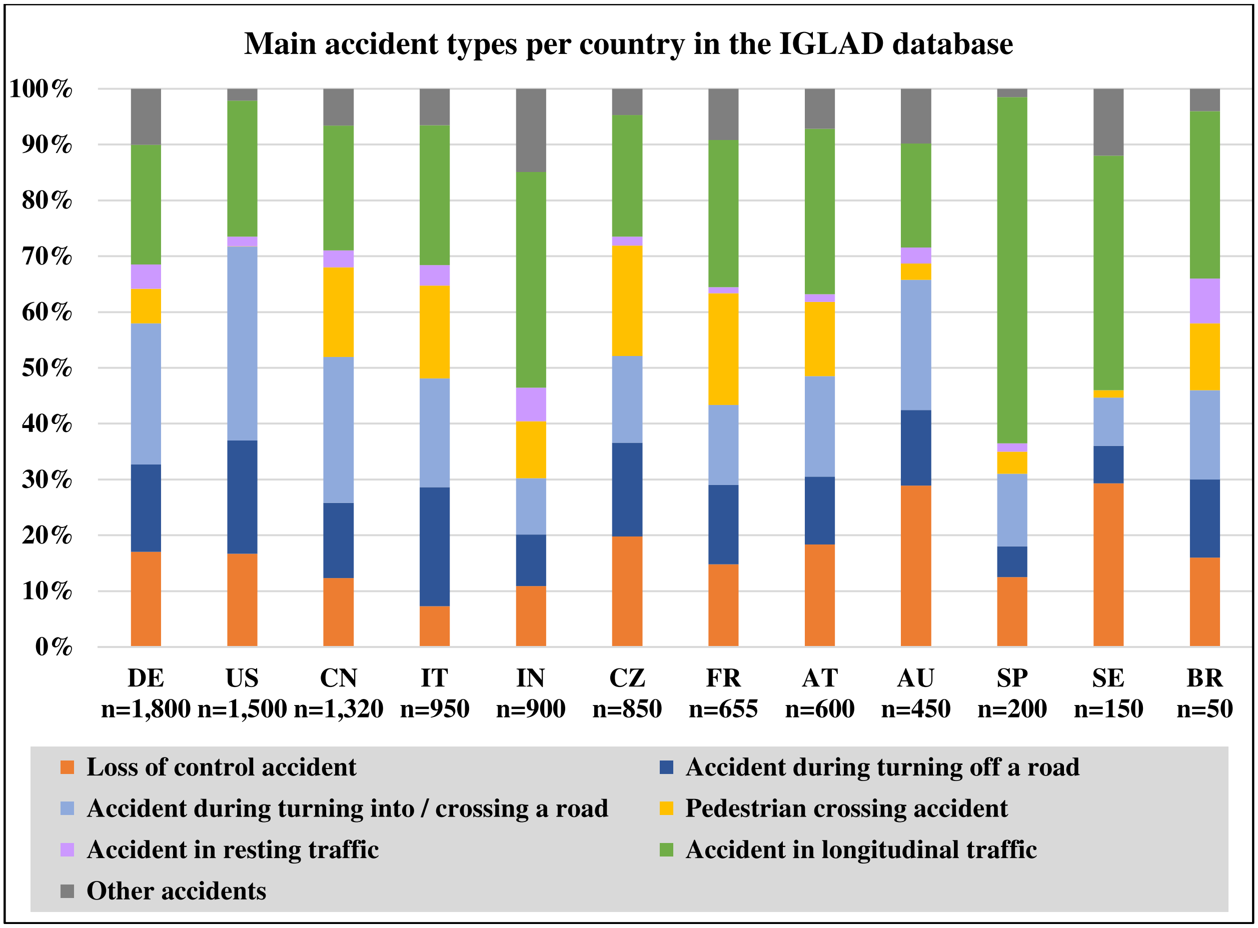


***Figure 3.* Distribution of main accident types per country in the IGLAD database (Phase 4 / 2021).**

Representativeness

Even though most countries provide national statistics on road fatalities, injured road users, etc. no detailed data is given to collision characteristics, collision speeds, etc. (1) Thus, the basic idea of IGLAD was to establish an international in-depth database comprising more specific data from many countries as possible (1) (4). Due to the different investigation priorities or sampling criteria (e.g. focus on car accidents, vulnerable road users, fatalities, etc.) of the data providers, the data are possibly lacking of representativeness. However, there are only a few ways to compare data for representativeness such as global status report on road safety (7), annual statistics of the EU (8), or national statistics. Although the mode of transport is available in more detail (8) only few details are available in other documents (7) and these documents comprise fatal accidents only. Additional documents from national statistics might comprise more detailed data but are not available for everybody or written in the countries' language.

Bakker et al. (1) made an attempt to compare IGLAD data with national statistics. Data from IRTAD (International Traffic Safety Data and Analysis Group) were used (9). Unfortunately access to the IRTAD road safety database is limited to members and only selected variables for comparison. Thus, the main objective of the IGLAD Representativeness group was to provide a document available to every IGLAD member. A standardized template with specific variables was developed (Table 1) for the calculation of weighting factors of the IGLAD data for each participating country. Within a survey the data providers were asked if they could provide aggregated accident data of their national statistics or region. Finally, data from seven countries (AT, AU, CZ, DE, FR, IT, US) out of twelve are available. For the other countries no data were provided due to several reasons (time effort, resources, availability of actual data).

Table 1: Weighting variables

ACCIDENT LEVEL	***CASUALTY LEVEL***
Accident severity	***Age***
Month	***Gender***
Weekday	***Injury severity***
Time	***Participant type***
Road condition	
Road type	
Accident type	
Collision type	
Number of participants	

Not every single attribute of the variables in IGLAD corresponds to the attributes of the variables in the national statistics. Thus, the data provider had to aggregate either their attributes to the IGLAD attributes or the IGLAD attributes need aggregation to be comparable with national statistics. Most data providers could submit full cross tables for all variables distinguishing between minor, severe and fatal injuries for their countries. Some only could separate their data into fatal and non-fatal accidents but were able to provide these data for most of the variables requested. In Figure 4 a comparison of different age groups in IGLAD and the corresponding countries is given as an example how the data in the national statistics are represented in IGLAD. If both marks in the figure are superimposed, the data in IGLAD fully represents the national statistics.

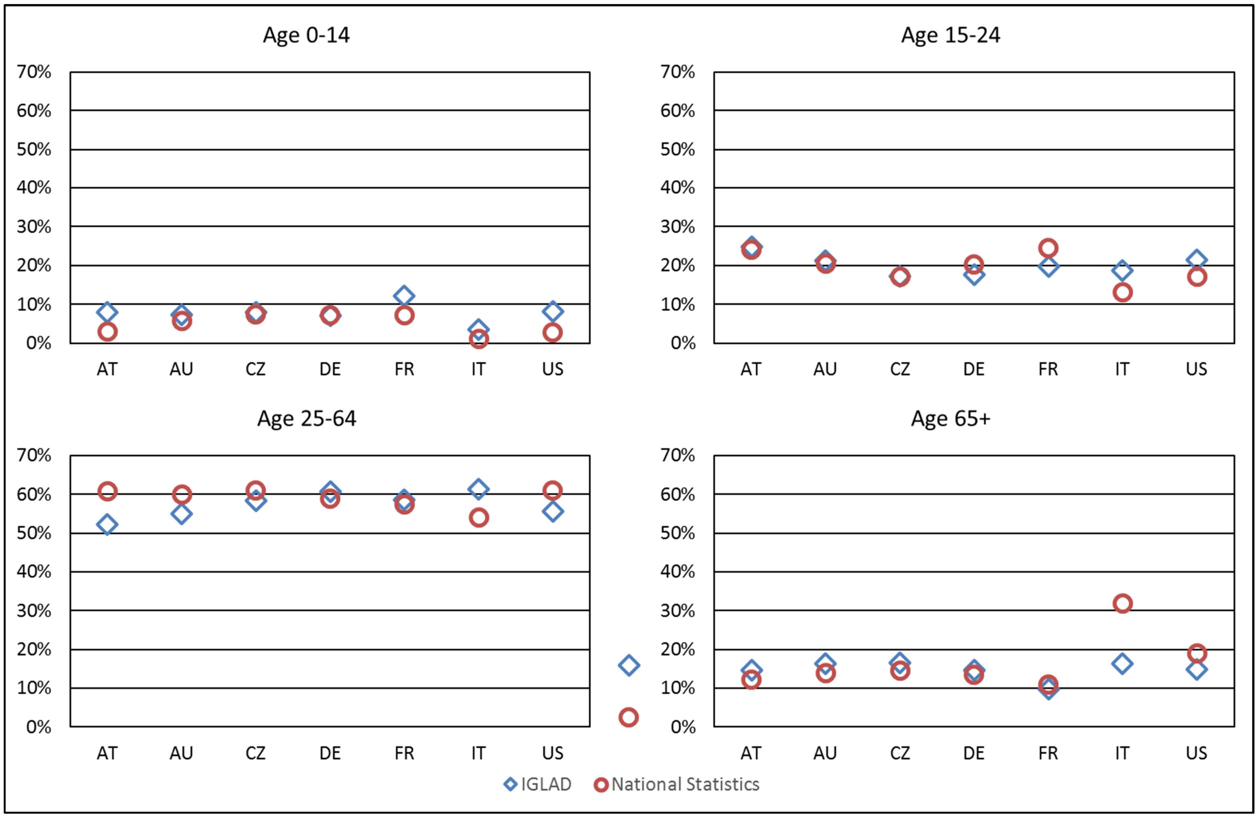


Figure 4. **Distribution of age groups in different countries in IGLAD compared to the national statistics.**

PCM (Pre-Crash-Matrix) Format

The sequence of accident events are often classified into three essential phases, the pre-crash phase, the in-crash phase and the post-crash phase. The vast majority of coded parameters in the IGLAD database is focusing on the in-crash phase (accident reconstruction) and passive safety aspects. However, for the evaluation of many safety systems or features the pre-crash phase is of particular interest. The assessment of the potential of sensor- or communication-based ADAS as well as AD functions can only be accomplished by a detailed analysis of the pre-crash phases of accidents (and incidents). Hence the necessity to analyze the early phase of accidents in detail arises.

Since its introduction in 2011, the Pre-Crash Matrix (PCM) format has offered the possibility to store and analyze information from the pre-crash phase of accidents (10). Until 2019, this format was not published and was used exclusively within the German In-Depth Accident Study (GIDAS) (11). Here, the time period of five seconds prior to the accident until the first collision (t0) was usually covered.

Since the revision and release of the format in 2019, there is also the possibility for any users to store and analyze data recordings of real driving situations (e.g. from Event Data Recorder (EDR), Naturalistic Driving Study (NDS), Field Operational Test (FOT), ...) in the format. The goal of the PCM format is to store accident and real driving data in a uniform, structured format in order to perform analyses and evaluations of time-dependent and driving dynamic variables. Furthermore, a detailed investigation of active safety systems has become possible.

The PCM format contains all relevant data to describe the pre-crash phase of an accident until the first collision. This includes the definition of the participants and their characteristics, the dynamic behavior of the participants as a time-series for at least five seconds prior to the accident, and the geometry of the traffic infrastructure. The following *Figure 5* shows the structure of the PCM format (12).

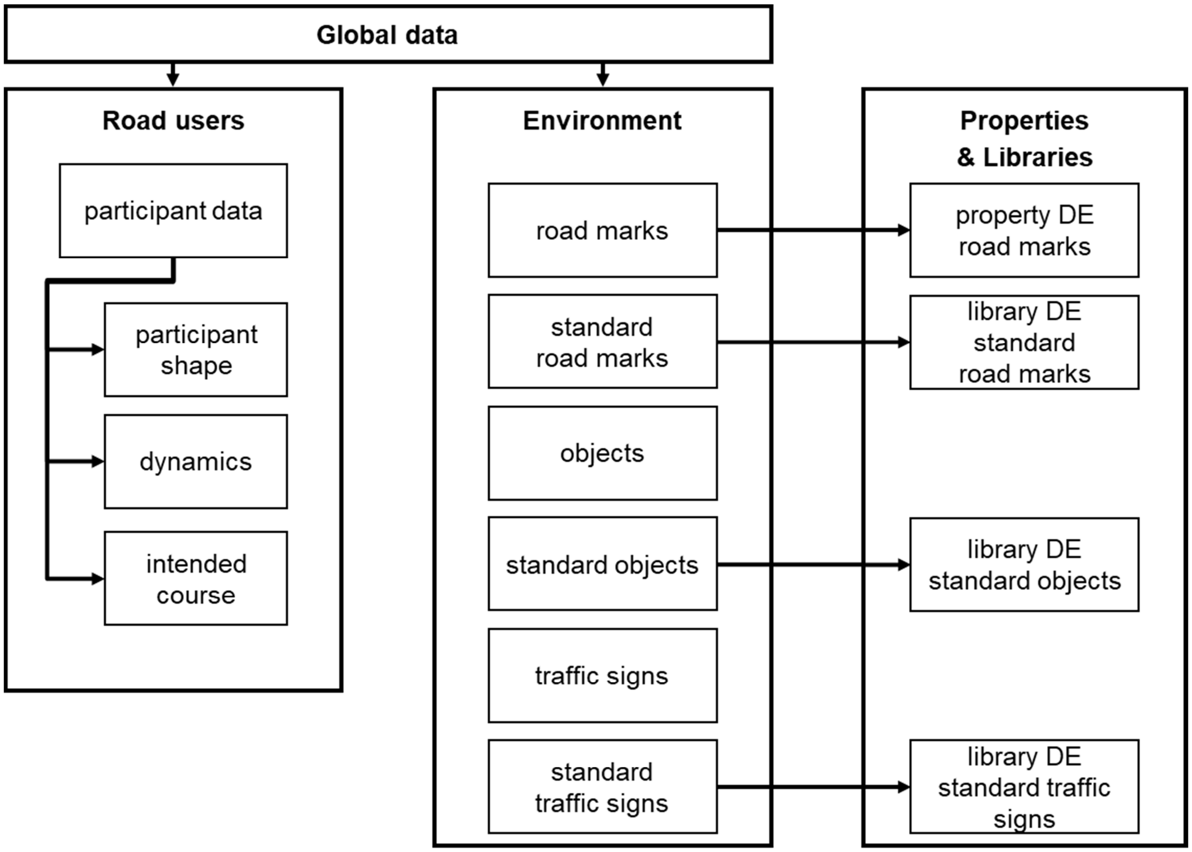


Figure 5. **Hierarchical structure of PCM (12).**

As can be seen in *Figure 5*, the PCM format consists of 15 different tables. These 15 tables can be roughly divided into the following categories: "Global data", "Road users", "Environment" and "Properties & Libraries". The table “Global data” provides general information of the accident / scenario, e.g. date and time and number of involved participants. The table "participant data" contains relevant variables to parameterize participants. This data can be used to model the geometry and further attributes. The table “participant shape” defines the geometrical shape for each participant by surfaces. Each surface contour is defined by points. Usually, simplified 2D/3D vehicle shapes are used within the PCM. Reducing the complexity of vehicle shapes increases the speed of simulation runs. However, it is also possible to realistically reproduce the vehicle contours in the PCM.

The table “dynamics” defines the global position of participants according to the global coordinate system as well as velocity and acceleration of the participants center of gravity (COG) according to the local COS at each time step of the simulation. The table “intended course” defines the course the participant initially intended to follow.

In the category "Environment" there are two basic approaches for defining objects. On the one hand, there are tables where information can be individually formatted (road marks, objects, traffic signs) and on the other hand tables which refer to a library (standard road marks, standard objects, standard traffic signs). The advantage of standard tables with the corresponding libraries is that they are defined only once and can be used for multiple cases, which saves storage space in the database. The PCM specification, explanations, and an example case can be found on the website of VUFO. (11)

Figure 6 shows an example visualization of a scenario in PCM format with Matlab.

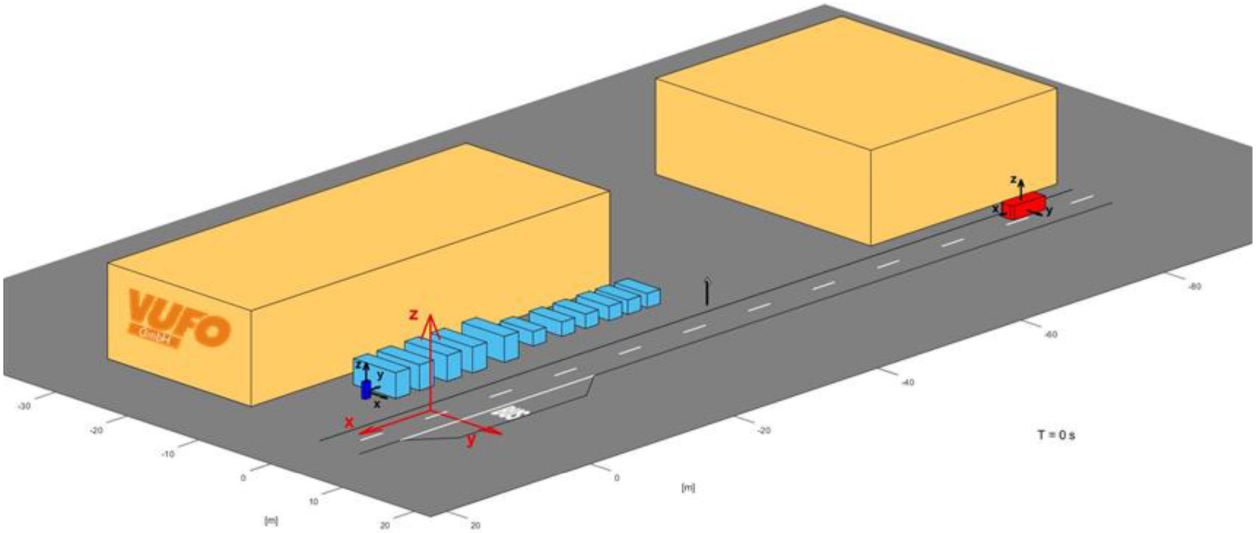


Figure 6. **Example visualization of a scenario in PCM format.**

With the PCM format, extensive user and application possibilities are thus available. In addition, it is freely available and independent of the data type. The simple structure allows easy access and understanding. However, two limitations of the PCM format specification should also be noted. Firstly, the environment has no logical information or metadata. This means that a participant has, for example, no direct information of the lane on which it is currently located. In addition, there is only one possibility of maneuver definition. In PCM format, all participants are defined by time-dependent trajectories of their center of gravity.

In order to visualize the data from the PCM format and also to perform a case by case analysis of quality as part of the creation of the IGLAD-PCM, VUFO has developed a PCM Viewer, which is shown in Figure 7.

Figure 7. **GUI of PCM Viewer.**

The PCM Viewer GUI is separated in two areas, the "tabs-area" with different settings on the left and the "visualization area" on the right. To visualize the PCM either a video can be played with "Play Simulation" or the participants can be shown in timesteps every one second and at the last timestep with "Show Timesteps". To visualize a specific timestep, the time can be chosen on the time slider of the visualization panel.

Creation of IGLAD-PCM from Austrian (CEDATU) data

Graz University of Technology uses a fully retrospective accident data collection approach based on court cases (13). The court data includes police reports, accident sketches (not vectorized), pictures of the accident scene (road geometry, roadside information, road markings, etc.) and road users, witness reports, medical reports, etc. Graz University of Technology uses PC Crash for the accident reconstruction. A scaled bitmap of the accident site with the collision and rest positions, road layout is sufficient. However, a digital scaled sketch is not mandatory for this purpose. All functions included in PC Crash are sufficient to obtain an adequate overview of the accident situation including compiling videos.

For processing PCM, however, the accident sketches were digitized (vectorized) and layered based on the requirements of the PCM format. Digitization is done in PC Crash with the included drawing tools based on the accident sketch of the police. Appropriate layers were defined in a template to associate road marks, sight obstructions, vegetation, etc. with the PCM specifications.

For the translation of the dynamics the in-house tool X-RATE (Extended Effectiveness Rating of Advanced Driver Assistance Systems) was used. X-RATE was developed to assess the effectiveness of driver assistance systems or autonomous driven vehicles (14) (15) (16) (17). X-RATE interacts with PC Crash on a time-step basis using the OLE interface of PC Crash in a MATLAB environment. However, an additional function was necessary to include the processing of environmental data (roadside, road marks, etc.), participant data, and dynamics to create a complete PCM dataset.

Even though there is advise on how to position objects in the PCM manual, the orientation of objects in the PCM dataset was not plausible in some cases. In Figure 8 different orientations of road markings in the accident sketch and in the PCM data is given. The mistake could be either found in an incorrectly associated layer type, a wrong scaling factor or the rotation angle in an incorrect unit. Furthermore, artefacts are found in the PCM data which are not in the accident sketch (Figure 9). Whilst the position of road users is given in a global coordinate system, the velocities and accelerations are related to the local coordination system. This might lead to incorrect values. Specifically, if the road user negotiates a bend the velocity does not correspond to the acceleration. In Figure 10 on the left side the velocity in y-direction does not correspond to the acceleration. Approximately 0.8 s before the accident the acceleration increases but the velocity remains constant. In the x-direction no acceleration is observed and the velocity does not increase or decrease. The right picture shows correct values for both, the x-direction and the y-direction.

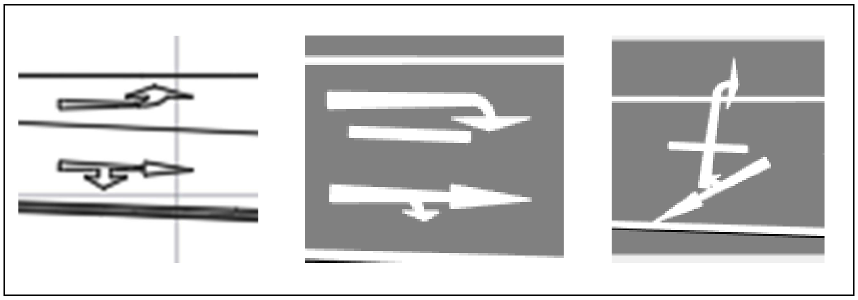


Figure 8. **Different orientation of standard road marks in the sketch (left) and in the PCM data (mid, right).**

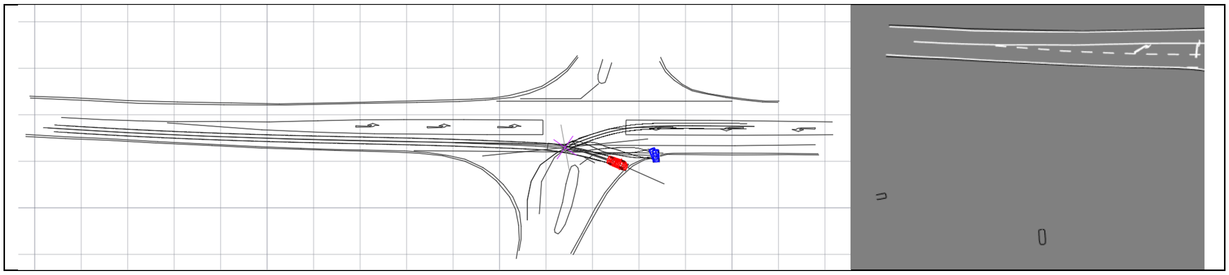


Figure 9. **Artefacts in the PCM data (right) which are not present in the accident sketch (left).**

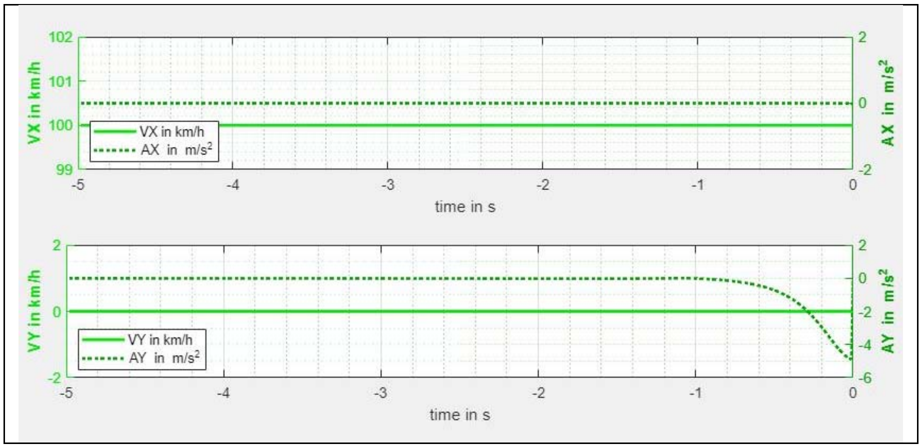


Figure 10. **Divergence between velocity and acceleration in y-direction**

The steps necessary to create a complete PCM dataset are either manually (draw vectorized sketches including the correct layers) or automatically (processing the sketches and dynamics). In the accident database CEDATU (Central Database for In-Depth Accident Study) of Graz University of Technology, the accident sketches are already created according to the PCM specifications on a regular basis. The effort to create a sketch according to PCM specifications compared to a simple vectorized accident sketch is not much higher. However, this manual related work leads to random errors e.g. incorrect object type and layer if the work is not carried out accurately.

Creation of IGLAD-PCM data from US (CISS) data

There are some differences that we want to outline in how the US IGLAD-PCM is generated compared to PCM data from other countries. The basis for creating the US IGLAD database and the PCM data set is the NHTSA CISS data (Crash Investigation Sampling System, see also (18). t is an in-depth data set and currently contains 3,000 – 4,000 accidents per year. Besides a complete reconstruction of each accident there are recordings of an EDR (Event Data Recorder; in the US usually named as *Crash Data Recorder / CDR*) available in most of the cases. They contain speeds, accelerations and other signals of a subset of the motorized participants at certain time intervals of the pre-crash phase. The PCM "dynamics" table that contains the kinematic information of each participant is generated using the data from the EDR of each involved vehicle. Additional checks are run to ensure that the resulting dynamics data is in line with the accident reconstruction in the IGLAD participant table which in turn is based on the CISS reconstruction data. Currently, the US PCM is the only IGLAD-PCM data set where the kinematic data is completely based on EDR so the process of creating it differs from that of the other IGLAD data providers which we will outline in more detail below.

When creating the US IGLAD-PCM the following four data sources are leveraged:

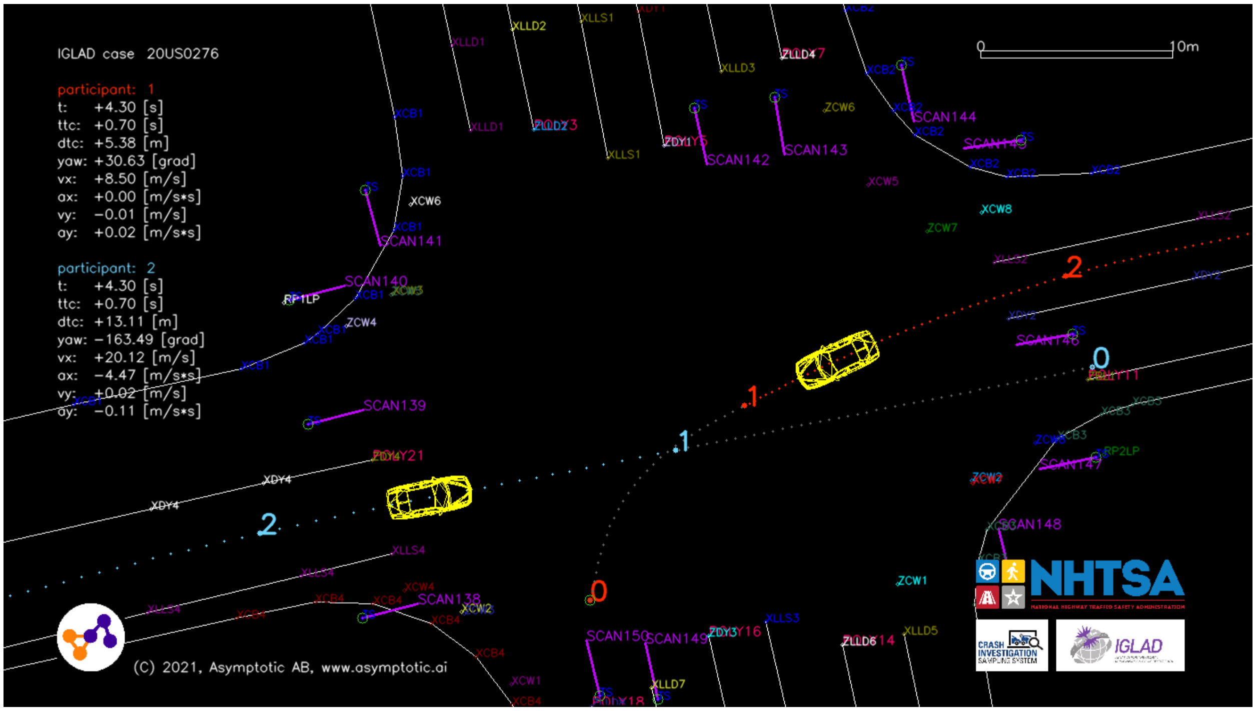
1. Accident and participant tables from IGLAD database: common accident and participant data which is imported from the core IGLAD and CISS database tables.
2. EDR data from the CISS database: measurements of different sensor inputs like velocity, acceleration and brake pedal activation typically starting five seconds before an event. Multiple events can be recorded.
3. Theodolite measurements stored in the CISS accident sketches: point coordinates of vehicle positions, road-, lane-markings, and object positions like signs and obstacles in 3D space, which are acquired by on-scene laser measurements using a theodolite and stored in FARO Blitz CAD sketch files.
4. Vehicle meshes stored in the CISS accident sketches: true to scale CAD meshes of the accident participant's exact vehicle model.

Based on these data sources a US IGLAD-PCM is created in five stages:

Stage 1. The general accident and participant information is re-coded to match the PCM "global data" and "participant data" tables.

Stage 2. The theodolite measurements are extracted from the CISS digital sketch of the accident scene. The CISS trajectory for each participant consists typically of a few points per participant that mark their positions during the pre- and post-collision phase including the collision point and the final rest. The PCM "dynamics" table is created by merging and enriching the trajectory points and the EDR data for each participant. First, the trajectory points need to be ordered and interpolated in a uniform manner. The trajectory curve of the US IGLAD-PCM is constructed using clothoid-splines as they tend to give the best fit for the vehicle's trajectory. Smooth tangent and curvature alignment between each clothoid-spline segment is achieved by using Hermite interpolation (19). With this enhanced representation of the participant's trajectory, the EDR data is complementing the geometric trajectory information and providing a speed and acceleration profile. This profile is mapped to the interpolated trajectory curve and the required values for the "dynamics" table are calculated for each simulation step along the trajectory of each vehicle up to the collision point. If more than one event is recorded before the primary collision in one of the participants EDR data a more sophisticated multi-segment mapping of the speed profile to the trajectory is applied.

Stage 3. The data for the "intended course" table of the PCM is calculated. Therefore, the trajectory needs to be extended beyond the collision point following an assessed intended course of the vehicle. This requires inserting new points into the CISS sketch by doing estimations based on the accident description and other information from the CISS database. These extra points are also interpolated and extend the already created clothoid-spline of the trajectory in a smooth manner without introducing unnatural accelerations. An example of the interpolated trajectory and intended course is shown in Figure 11.



```
IGLAD case  20US0276

participant:  1
t:    +4.30 [s]
ttc:  +0.70 [s]
dtc:  +5.38 [m]
yaw:  +30.63 [grad]
vx:   +8.50 [m/s]
ax:   +0.00 [m/s*s]
vy:   -0.01 [m/s]
ay:   +0.02 [m/s*s]

participant:  2
t:    +4.30 [s]
ttc:  +0.70 [s]
dtc:  +13.11 [m]
yaw:  -163.49 [grad]
vx:   +20.12 [m/s]
ax:   -4.47 [m/s*s]
vy:   +0.02 [m/s]
ay:   -0.11 [m/s*s]
```

***Figure 11*. US IGLAD-PCM with clothoid-spline interpolation of vehicle trajectory and intended course.**

Stage 4. Additionally, to the trajectory points all road-, lane-markings and object positions like signs and obstacles are extracted from the sketches and then grouped, rotated if necessary, and mapped to the coding conventions of the PCM tables “road_marks”, “standard_road_marks”, “standard_objects”, and “standard_traffic_signs”.

Stage 5. The PCM provides the “participant_shape” table with a geometric outline of each participant. The CISS sketches already include an accurate CAD mesh for the exact model of each participant. However, the PCM requires a different geometric representation which is created by extruding the 2D shape of the model’s mesh. This is accomplished by calculating the convex hull of the projection of the vehicle model to the ground surface and extruding the resulting shape to the height of the vehicle in vertical direction. An example of the shape extrusion is shown in Figure 12.

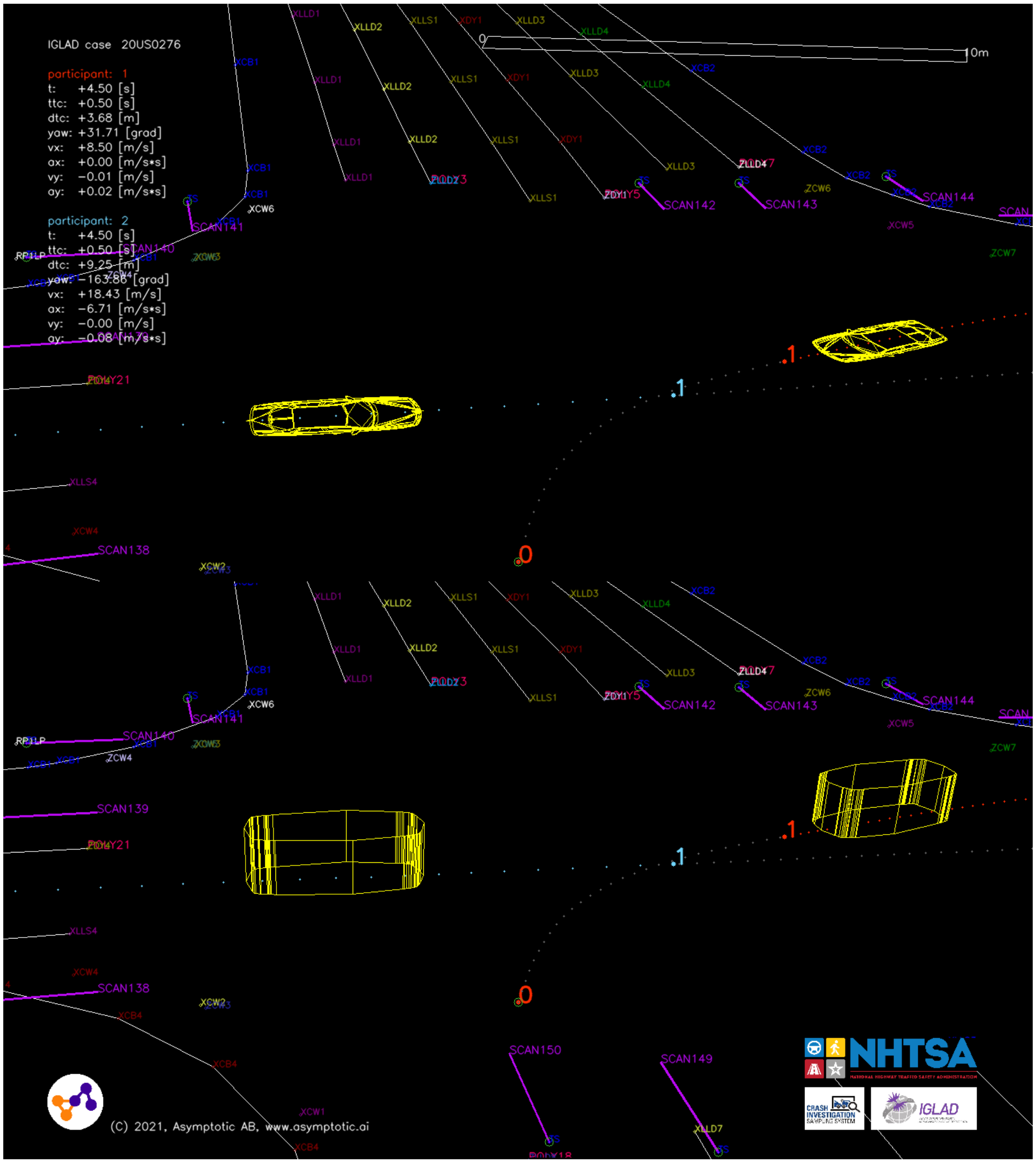


***Figure 12*. PCM vehicle shape extruded (right) from the convex hull of the CISS CAD mesh (left).**

The resulting PCM generated from the US CISS data offers a complete data set with a high-quality standard which conforms to the latest PCM specifications. One central piece of the PCM data is the "dynamics" table where the US IGLAD-PCM can excel by providing data which is based on real sensor measurements supplied by the EDR units in the participating vehicles. Opposed to reconstruction based PCMs there are less assumptions involved in the underlying data. Differences to reconstruction based PCMs are for example pre-crash maneuvers that are visible in the EDR measurements but would usually not appear in a reconstruction because there is no on-scene evidence present for these. Examples found in the data showing these differences were short braking maneuvers or accelerations of a participant at a crossing that attempts to avoid a collision when recognizing a car approaching from behind or a quick acceleration followed by sharp braking at traffic lights that most likely were switching from green to red.

The upcoming CISS releases are expected to include more detailed data from updated EDR units and also to extend to a greater variety of accident types. This will even further improve quality and richness of the US IGLAD-PCMs.

Quality assurance process

As the vast majority of data providers have never created PCM files before, the aspect of quality is quite important. From VUFO's more than 10 years of experience with PCM creation and format development, it is known that there are many pitfalls and many small details to consider. For this reason, each of the 200 cases was subjected to a quality check by VUFO on a case-by-case basis. The focus was on compliance with the PCM format specification, e.g. use of the correct units and completeness of the tables on the one hand, and on the other hand on the correct transfer of each individual scenario (case-by-case review). The case-by-case review was performed by using the VUFO PCM Viewer. The following items were analyzed within the process:

- Has all relevant information been transferred from the sketch?
- Do the trajectories match the sketch?
- Are there jumps or interruptions in line segments of the sketch?
- Are the objects and markings rotated correctly?
- Are the speeds and ac-/decelerations plausible?

Finally, all Data Providers managed to deliver PCM data in a good quality after few bilateral iterations between VUFO and the Data Providers.

RESULTS

Content of the IGLAD-PCM

The first ever published IGLAD-PCM contains 200 personal injury accidents from seven different countries, provided by eight data providers. They represent a subset of the regular IGLAD datasets, although the very first release did not have too strict requirements for representativeness in case selection. The primary goal was to prove the feasibility and to establish the corresponding process and tool chains.

However, quite promising results could be derived. Figure 13 shows the distribution of accident configurations in all 9,425 IGLAD accidents and the distribution for the IGLAD-PCM cases respectively. Nearly all configurations are covered in both the IGLAD database and the IGLAD-PCM. Generally, accidents involving a car and a second road user are overrepresented. The most frequent accident configuration in the IGLAD-PCM is the group of car-to-car accidents. Single car accidents are underrepresented as they are particularly challenging because they are usually characterized by unstable driving conditions (understeer or oversteer) in the pre-crash phase. An automatic transfer of the dynamic data into the IGLAD PCM is more difficult than accidents without unstable driving conditions.

Accidents involving two VRU are not included at all so far. The same applies for single bicycle accidents. However, they are usually not of big interest for the current members and stakeholders.

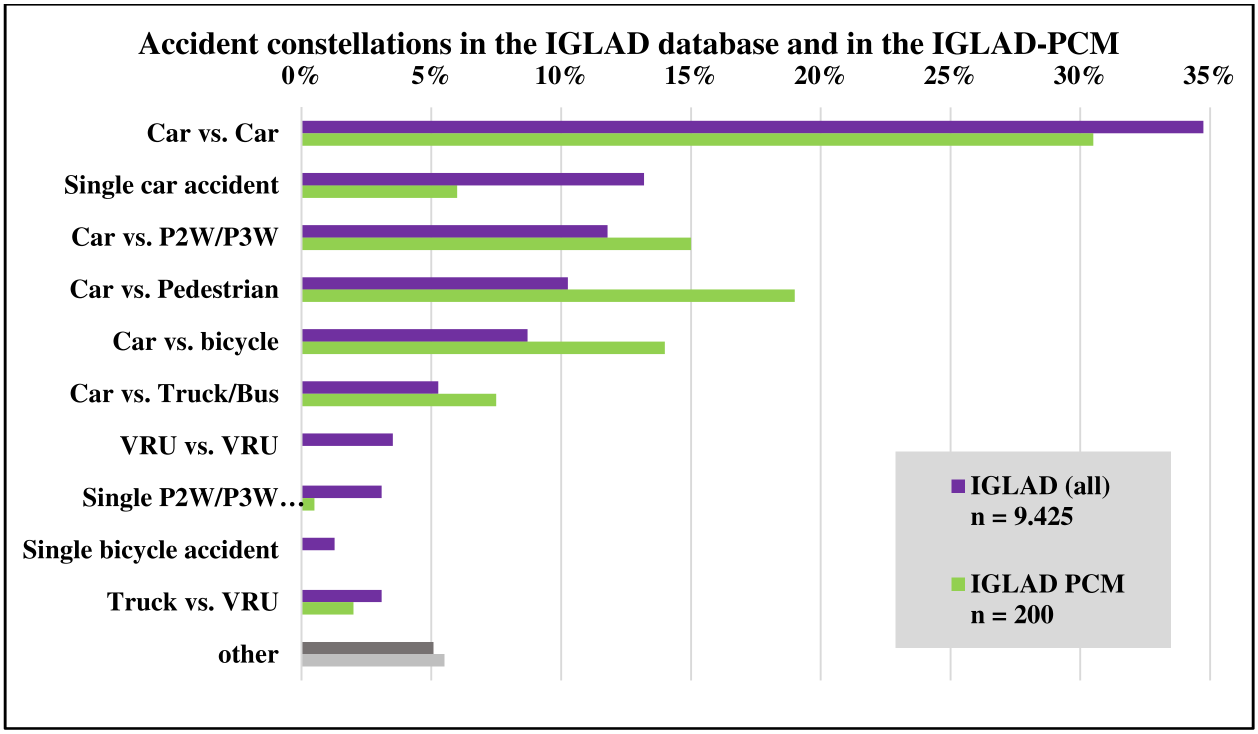


Figure 13. **Accident constellations in the IGLAD database and in the IGLAD-PCM**

Figure 14 shows the accident severity distribution of the 200 PCM cases. The distributions for most countries is similar to the one in Figure 1. For the Austrian, Chinese as well as the Italian data the accidents in the IGLAD PCM is even more representative for the related country's national statistics. This is due to the fact that the vast majority of PCM cases were selected from the latest IGLAD release (Phase IV / 2021; containing accidents of the years 2019 and 2020). Representativeness aspects are becoming more and more important within IGLAD since years and thus, most countries provide IGLAD data with a good level of representativeness.

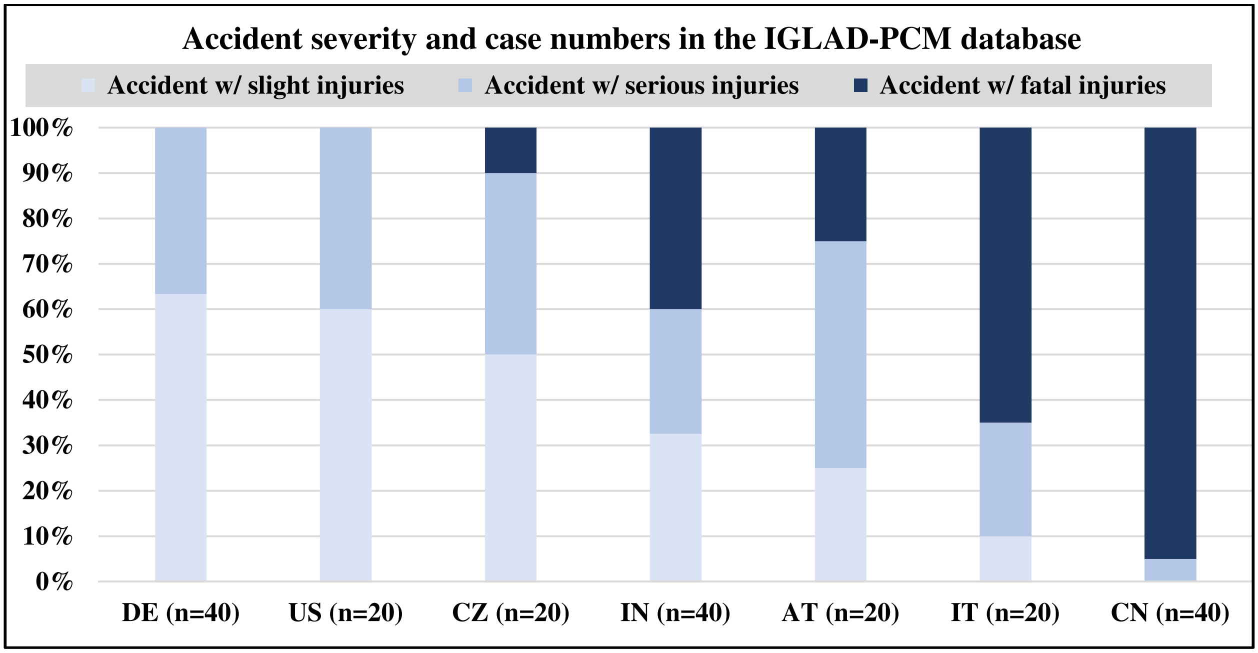


Figure 14. **Accident severity and case numbers in the IGLAD-PCM database**

Applications of the PCM format

Accident and/or traffic scenarios which are stored in the PCM format can be used in several ways. Most of them base on retrospective analysis to support the development of future vehicle safety systems. Figure 15 shows some examples, which are briefly explained below.

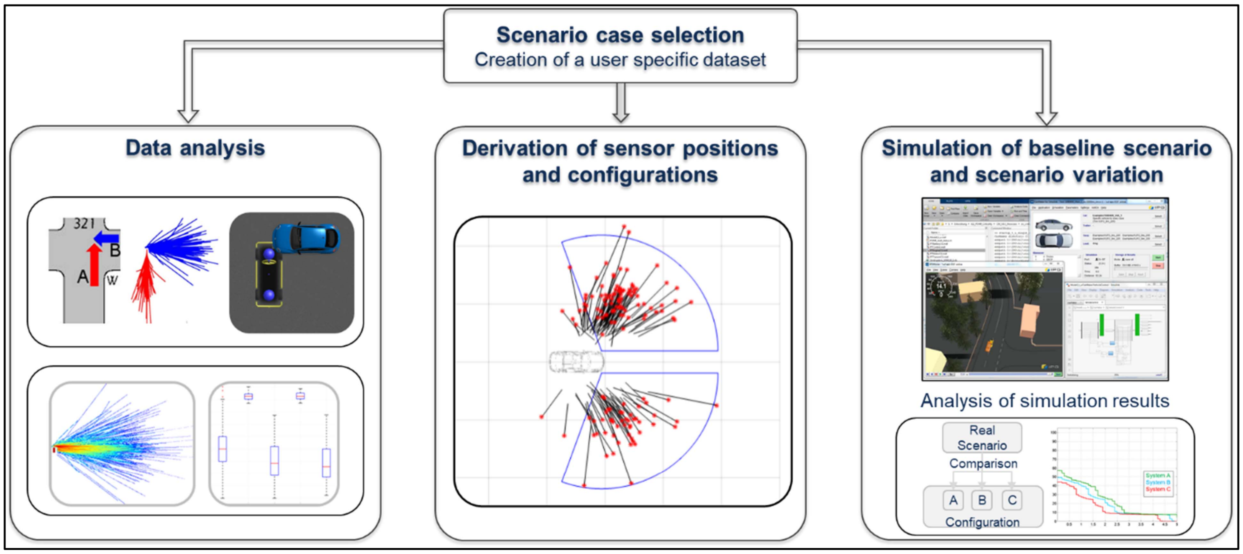


***Figure 15*. Selected application possibilities of PCM data**

Within the field of (retrospective) data analysis, sub-samples from the overall database are examined by means of descriptive analyses. These are, for example, individual or grouped accident types. From these, scenario groups can be formed and analyzed. For example, the left part of the figure above shows the center of gravity trajectories of the accident participants of accident type 321 in the upper section. This allows the approach direction of the accident participants to be analyzed. In addition, the corresponding speeds and accelerations of the participants can also be analyzed, as in the boxplot in the lower section. This means that the following questions can be answered: Which road user is where at which point of time and which maneuver is he carrying out?

Another possibility is the derivation of sensor positions and specifications. This means that an ego vehicle is defined and the position of several participants from different scenarios is examined. Thus, for example, the position or also the specification (opening angle, range, etc.) of a sensor can be examined at a defined time or in a time range. Especially in connection with criticality measures like TTC (Time to Collision) this can be an important tool. Important research and development questions can be answered, as for example: Where should a sensor be positioned on the ego vehicle and which specification should it have in order to provide a high benefit?

In addition, it is possible to transfer scenarios into vehicle dynamics solvers. This offers the possibility to implement a virtual system in the ego vehicle and then to analyze the influenced scenario in relation to the baseline scenario in order to evaluate the effectiveness of the system. This can be used to answer the question if a system has an impact on the accident scenario, and if so, what type and magnitude of impact?

Field-of-View and Speed Analysis

PCM pre-crash data can be used in a variety of ways in the context of automotive engineering. One possible way is the application for deriving system requirements for ADAS. The data provides key performance indicators for sensors, algorithms, and actuation controls. The goal is to make system requirements correct, complete and relevant, and thus support the development of safe and robust products.

Since the PCM only provides a description of the pre-crash trajectories but does not contain metadata on the traffic accidents, the PCM needs to be considered as an add-on to the IGLAD accident data. A two-step approach to analyzing IGLAD accident data and PCM pre-crash data is proposed as shown in Figure 16.

1. Deriving of relevant accident scenarios in IGLAD, such as car-vs-car or car-vs-VRU. Each accident is considered from the viewpoints of the participants in the conflict situation, to generate a set of accident scenarios. For the corresponding method and the used scenario catalog, refer to (20). The car-vs-car accidents can be targeted from the causer or from the non-causer perspectives, therefore the number of accident scenarios that are addressed by ADAS is twice the number of accidents. Car-vs-VRU accidents are only analyzed from the car perspective, as cars are considered responsible for VRU protection. The resulting complete set of IGLAD accident scenarios is the basis for a car safety system development.

2. Analysis of pre-crash characteristics, such as object positions, orientations, and speeds. Subsequently the pre-crash scenarios are retrieved from the PCM and analyzed from the perspective of the ego car. Therefore, the positions, orientation and speeds are transformed into the ego space. This allows for an analysis of sensor fields-of-view in different accident scenarios such as run-up, crossing or turning. The speed information is used for designing real-world test cases. Thus, the validation of safety systems can be supported by virtual simulation of PCM test data.

The current PCM data set, in data year 2021, contains 84 car-vs-car pre-crash scenarios. Furthermore, there are 11 car-vs-motorcycle, 14 car-vs-bicycle, and 21 car-vs-pedestrian pre-crash scenarios.

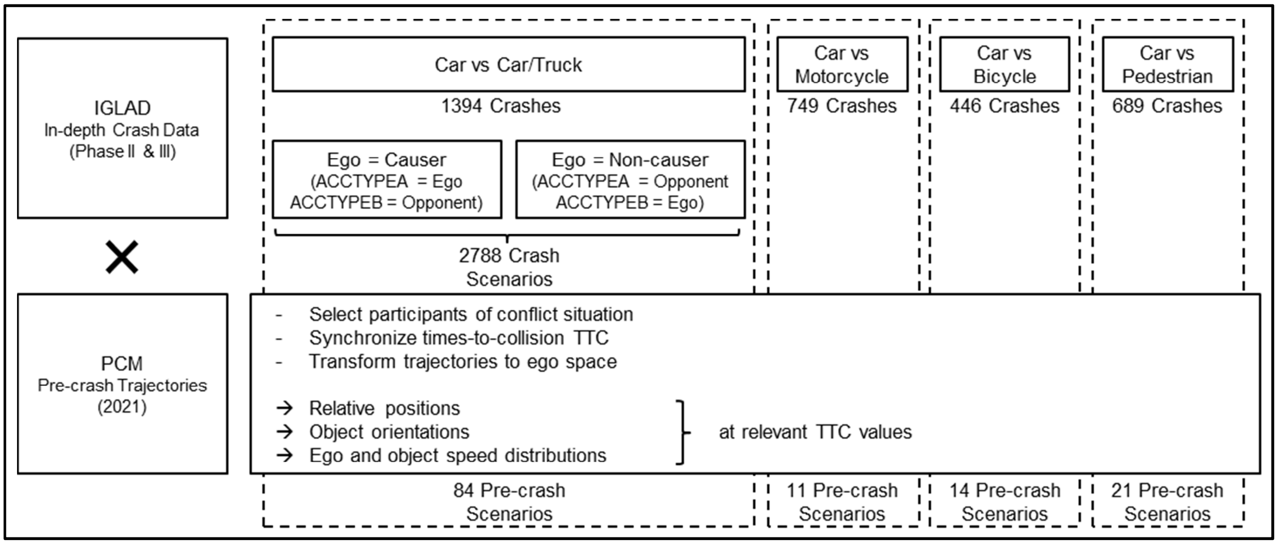


Figure 16. **Method for deriving pre-crash scenarios from IGLAD and PCM.**

PCM data analysis results of car-vs-car accidents are shown in Figure 17. The left-hand side picture depicts the relative positions and orientations of other cars at time-to-collision TTC = 2s. To put the object positions into perspective, generic fields-of-views for typical sensors are drawn: LRR (long range radar), CAM (camera), SRR (short range radar). The count for each scenario type and for each country is given. The most frequent scenarios are “run-up” L1 (13 cases), “rear-end” L4 (13 cases), “crossing from right” C1 (13 cases), and “oncoming same lane” On1 (12 cases). The right-hand side picture shows the speed histograms of ego and object car. Ideally, once the IGLAD-PCM data sample grows, the field-of-view visualization and speed analysis should be performed on a scenario-basis, to get requirements for the different safety functions.

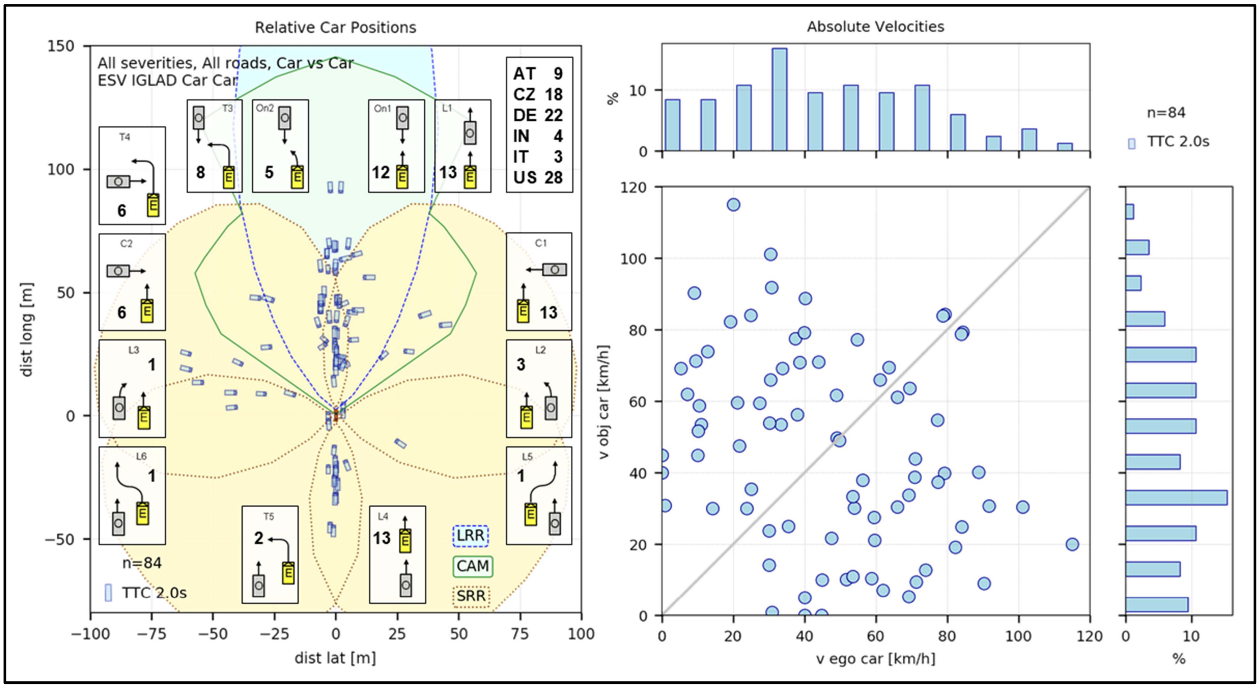


Figure 17. **Field-of-view and speed analysis of car-vs-car pre-crash scenarios**

The car-vs-VRU accidents are also analyzed and visualized in Figure 18 (car-vs-motorcycle), Figure 19 (car-vs-bicycle), and Figure 20 (car-vs-pedestrian). Again, each figure shows the field-of-view and speed analysis. Also, the counts per scenario type and country within the IGLAD-PCM data sample are given.

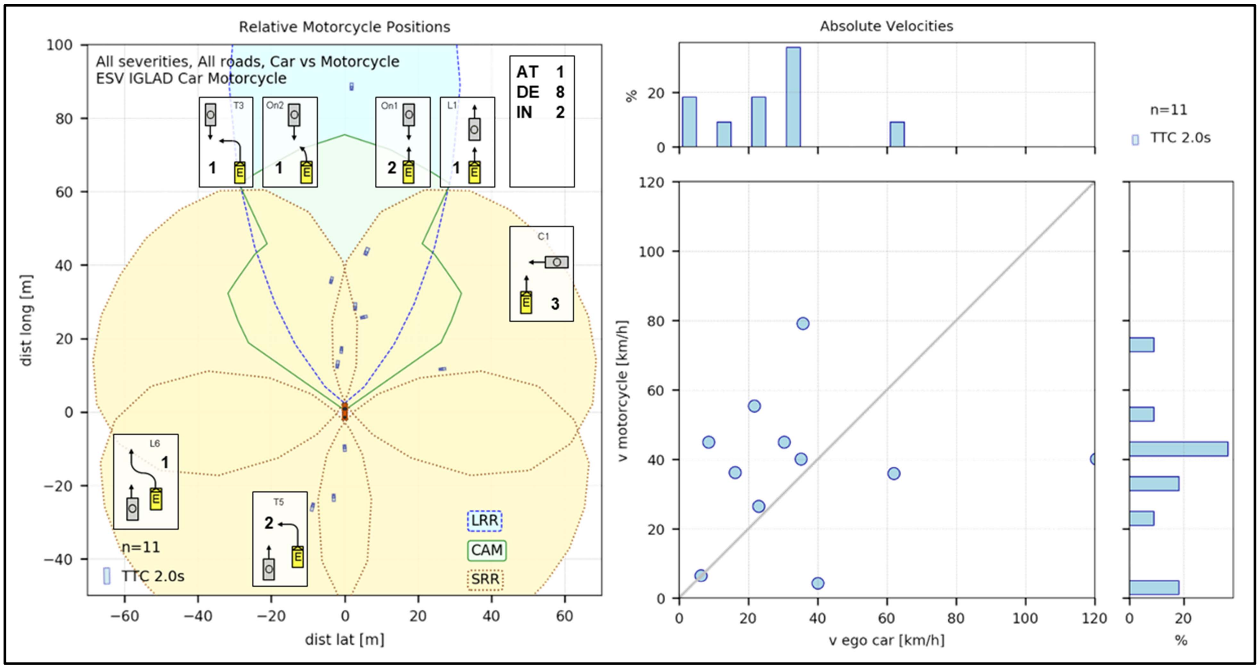


Figure 18. **Field-of-view and speed analysis of car-vs-motorcycle pre-crash scenarios**

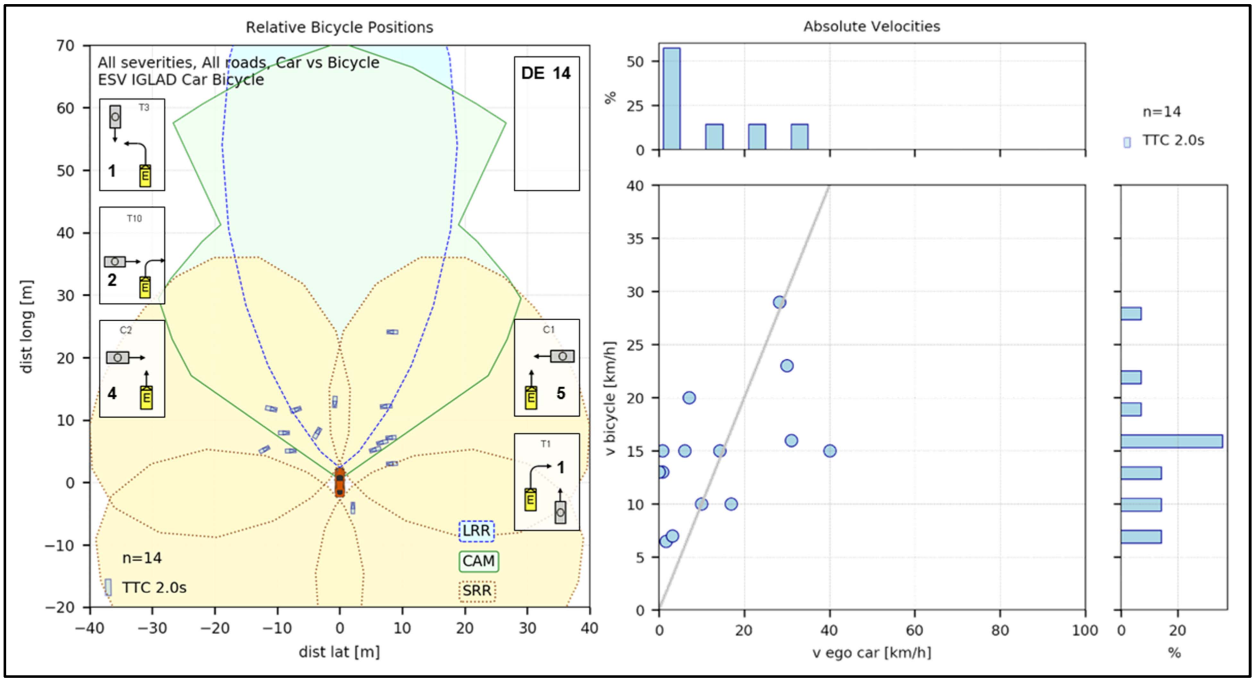


Figure 19. **Field-of-view and speed analysis of car-vs-bicycle pre-crash scenarios**

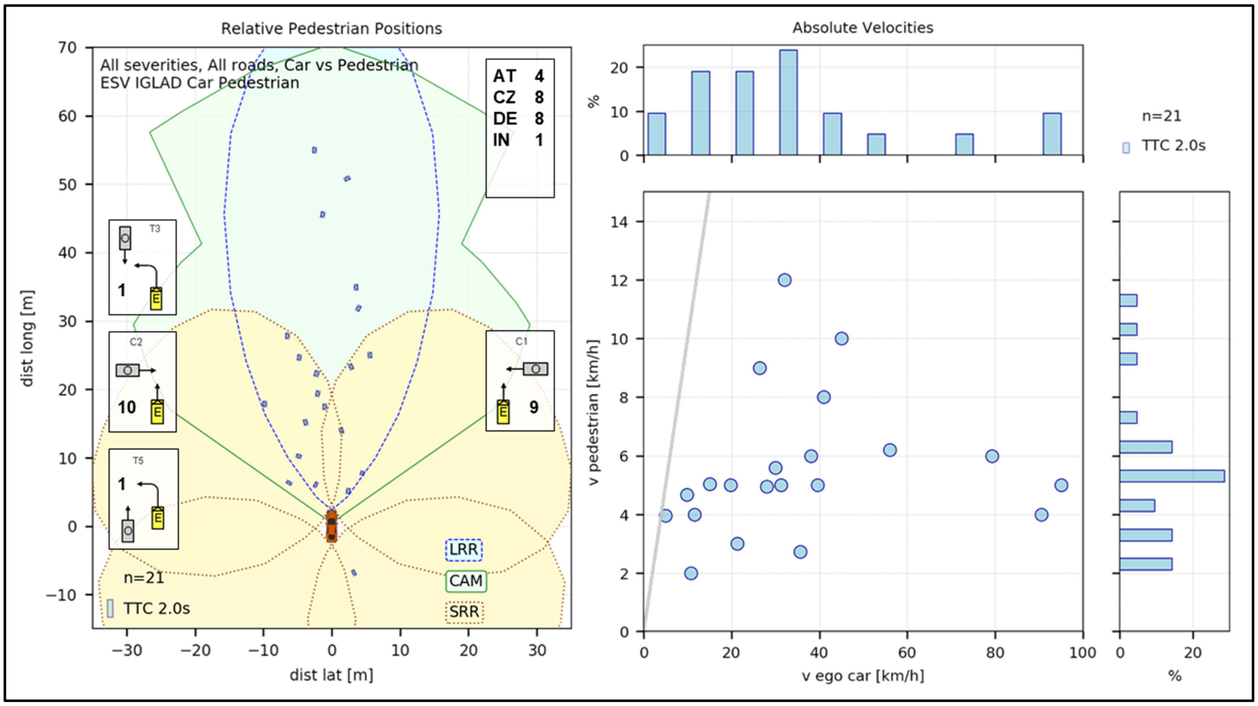


Figure 20. **Field-of-view and speed analysis of car-vs-pedestrian pre-crash scenarios**

DISCUSSION AND LIMITATION

The challenging task of storing data from reconstructed accident scenarios from different data providers in a uniform simulation database was successfully completed. Within one year, eight data providers from seven countries on three continents have transferred a partial data set of their IGLAD accidents into the PCM format V5.

The main challenges were, among others, the transfer of the dynamics data (trajectories, velocities, accelerations), the correct transfer of sketch objects and the first-time application of the PCM format specifications. Nevertheless, it was possible to generate consistent PCM files from the various basic data.

The establishment of an additional quality assurance process proved to be useful, in which the data provider together with VUFO was able to improve the quality of the delivered PCM data in an iterative way. Thus, it can be ensured that the existing 200 PCM cases are consistent, and the contents meet a certain quality standard.

Another benefit of the PCM creation is the fact that the original data and accident sketches are re-examined during the creation and quality check process, which may reveal further errors. However, as the quality checks are done after the PCM has been completely created by the data providers, the effort for correcting errors is rather high (especially for sketches) and could be mitigated with a continuous quality control during the generation of the data. An appropriate visualization tool or quality control of the processed data immediately at the end of the data collection is highly recommended to evaluate the plausibility and quality of the own data.

A limitation, besides the small number of cases so far, is the lack of representativeness of the IGLAD-PCM. As shown in Figure 13, so far mainly accidents involving passenger cars have been transferred to the PCM. These are clearly overrepresented, while accidents between VRUs or truck and bus accidents are underrepresented.

In addition, there are still some challenges. One is the automatic transfer of driving accidents. This main accident type is also underrepresented in the PCM. Another challenge results from the heterogeneous types of road markings and traffic signs around the world. Here, the IGLAD community is working together to create a common catalog of standard objects, at least for the most important and frequently used road markings and traffic signs.

CONCLUSION AND OUTLOOK

We have shown that the PCM specification makes it possible to harmonize pre-crash data from different countries and still take regional specifics into account. Since the PCM specification is an open data format, different scenario descriptions can be easily created and existing development tool chains can be supported. The initial release of the database includes only 200 accidents, but through annual updates the number of scenarios will constantly increase. The example has shown which applications of the IGLAD-PCM are possible.

This first release of the IGLAD-PCM was an important step forward for the IGLAD project. Thanks to the internal start-up funding from the IGLAD consortium, six new data providers have been enabled to convert their reconstructed accident data to PCM format, in addition to VUFO, which has already been delivering PCM data (within the GIDAS project) since 2011.

We have shown how the PCM data can be used to support data-driven system engineering of ADAS and AD systems. Pre-crash data are relevant for deriving key-performance indicators for sensors, controls, and actuation to develop safe and robust systems. Additionally, test scenarios are key for real-world validation by virtual simulation and to show how systems tackle critical crash cases.

The valuable experiences from this project encouraged us to follow-up with additional activities. There are several actions planned or have been already started, e.g.

- the creation of a catalogue of standard objects (mainly road markings) from different countries
- creating appropriate tools for data providers for the efficient and correct creation of PCM files
- enabling more/new data providers to create IGLAD-PCM
- transfer of PCM files into OpenDRIVE and OpenSCENARIO

In addition to the integration of further data sources or data providers, the last point mentioned is particularly important. In the future, the OpenX formats will be used as standard formats for the development, testing, validation, and Periodical Technical Inspection (PTI) of vehicles and functions. Corresponding real-world scenarios should therefore also be made available in these formats in the near future. For this purpose, VUFO already has approaches for conversion, and several hundred (GIDAS, not IGLAD) cases have already been successfully converted into the OpenX formats.

REFERENCES

Literaturverzeichnis

1. **Bakker, J., Jeppson, H., Hannawald, L., Spitzhüttl, F. et al.** *IGLAD - International Harmonized In-Depth Accident Data.* in: NHTSA (ed.), The 25th ESV Conference Proceedings, ESV Conference Proceedings, International Technical Conference on the Enhanced Safety of Vehicles, Michigan, USA, 5.-8.6.2017, NHTSA, 2017.

2. **Bakker, J., Ockel, D. und Schöneburg, R.** *Multinational in-depth accident data: From concept to reality.* In: ESAR (Hg.): 6th International Conference on ESAR "Expert Symposium on Accident Research", June 20-21, Hannover, 2014.

3. **Ockel, D., Bakker, J. und Schöneburg, R.** *Internationale Harmonisierung von Unfalldaten.* Fortschrittsbericht des FIA / ACEA Projekts iGLAD (Initiative for the Global Harmonization of Accident Data). In: VDI-Berichte 2144: VDI Verlag (VDI-Berichte), S. 301–309., 2011.

4. **Ockel, D., Bakker, J., and Schöneburg, R.** *An initiative towards a simplified international in-depth accident database.* in: ESAR (ed.), 5th International Conference on ESAR "Expert Symposium on Accident Research", Hannover, 2012.

5. **Heinig, I., Bakker, J.** IGLAD Homepage. [Online] [Zitat vom: 08. December 2022.] http://www.iglad.net.

6. **Chalmers, SAFER Vehicle and Traffic Safety Center at.** https://www.saferresearch.com/. [Online] [Zitat vom: 24. 11 2022.]

7. **WHO.** *Global status report on road safety 2018.* World Health Organization, Geneva, Switzerland, ISBN 978-92-4-156568-4, 2018.

8. **European Commission.** *Annual statistical report on road safety in the EU, 2021.* 2022.

9. **OECD.** IRTAD road safety database. [Online] [Zitat vom: 20. November 2022.] https://www.itf-oecd.org/irtad-road-safety-database.

10. **Schubert, A., Liers, H., Petzold, M.** *The GIDAS Pre-Crash-Matrix 2016 - Innovations for standardized pre-crash-scenarios on the basis of the VUFO simulation model VAST. 7th* International Conference on ESAR "Expert Symposium on Accident Research", Hannover, June 09/10, 2016.

11. **Verkehrsunfallforschung an der TU Dresden GmbH.** https://www.vufo.de/pcm/. *PCM Specifications.* [Online] [Zitat vom: 18. 11 2022.]

12. **Petzold, M., Schubert, A., et. al.** *Pre-Crash-Matrix (PCM) Format Specification v5.0.* [PDF] VUFO GmbH, 2021.

13. **Tomasch, E., Steffan, H., and Darok, M.** *Retrospective accident investigation using information from court.* in: TRA (ed.), Transport Research Arena Europe 2008 (TRA), Ljubljana, Slovenia, April 21-24, ISBN 978-92-79-10039-0, 2008.

14. **Gruber, M., Kolk, H., Klug, C., Tomasch, E. et al.** *The effect of P-AEB system parameters on the effective-ness for real world pedestrian accidents.* NHTSA (ed.), The 26th ESV Conference Proceedings, International Technical Conference on the Enhanced Safety of Vehicles, Eindhoven, Netherlands, 10-13 June, NHTSA, 2019.

15. **Kolk, H., Tomasch, E., Haberl, M., Fellendorf, M. et al.** *Active safety effectiveness assessment by combination of traffic flow simulation and crash-simulation.* ESAR (ed.), 8th International Conference on ESAR "Expert Symposium on Accident Research", Hannover, 19.-20.04, 2018.

16. **Kolk, H., Sinz, W., Tomasch, E., Bakker, J. et al.** *Evaluation of a momentum based impact model and application in an effectivity study considering junction accidents.* ESAR (ed.), 7th International Conference on ESAR "Expert Symposium on Accident Research", 7th International Conference ESAR, Hannover, 09.-10.06.2016, 2016.

17. **Wimmer, P., Düring, M., Chajmowicz, H., Granum, F. et al.** *Toward harmonizing prospective effectiveness assessment for road safety: Comparing tools in standard test case simulations.* Traffic Injury Prevention 20 (sup1):S139-S145, 2019, doi:10.1080/15389588.2019.1616086.

18. **NHTSA.** https://www.nhtsa.gov/crash-data-systems/crash-investigation-sampling-system. [Online] [Zitat vom: 25. 11 2022.]

19. **Spitzbart, A.** *A Generalization of Hermite's Interpolation Formula.* American Mathematical Monthly, 67 (1): 42–46, doi:10.2307/2308924, JSTOR 2308924.

20. **Feifel, H., Wagner, M.** *Harmonized Scenarios for the Evaluation of Active Safety Systems based on In-Depth-Accident Data.* 8th International Expert Symposium on Accident Research (ESAR), Hanover. 2018.