ABSTRACT

An important method for the development of strategies and measures to prevent accidents and mitigate the injury severity is the analysis of accident databases. However, the variety of research questions requires different kind of information. To assess accident situations, examine trends, or similar analysis, databases at a base level such as national statistics are available for many countries. On the other hand, the identification of accident and injury causation and the evaluation of countermeasures require a higher level of detail. For this reason, several in-depth accident data collection projects emerged worldwide in recent years. Unfortunately, due to different standards for data collection and coding comparative analysis of in-depth data from different countries is difficult or even impossible. This paper describes the approach taken by the IGLAD (Initiative for the Global Harmonization of Accident Data) project to handle these shortcomings by using a common data scheme and investigates the opportunities and limitations for weighting and extrapolation to national statistics. The methods used to process and merge the different data samples are described and an overview of the current status in terms of case counts, marginal distributions and the participating countries from Europe, Asia, Australia and North and South America providing data for the project is given. As an application example, the IGLAD dataset with accidents from 2007 to 2015 was used to analyze the distributions of accident types, presence of safety systems, characteristics of injury severity for each country and provide country comparisons. Also capabilities for pre-crash analysis were assessed. As a result, exemplary statistical assessment of injury probability, descriptive statistics for comparison between different countries were given as a result of the analysis. A pilot study about a more detailed analysis of the pre-crash phase has already been conducted which would allow for analysis of the potential benefit of safety systems in different countries. The authors discuss limitations, special characteristics and bias of the data samples from the individual countries. An outlook is given on the future development of the project, now preparing its fourth data release and on further extension of the data. Summarizing, the paper gives an overview of IGLAD as a new field crash data set and shows its unique opportunities for road accident analysis in a global scope, which are not provided by any other accident data source.
HISTORY, STATUS AND OUTLOOK

Since its start in late 2011, the IGLAD project (initiative for the global harmonization of accident data) has come a long way. The goal of the project is to build up a database of so-called in-depth accident data on an international level. While most of the countries worldwide provide basic national statistics about the number of road fatalities or injured persons on a very high and aggregated level, in-depth data provides details about single cases, their environment, participants, collisions, injuries and safety systems. So far, no data that can be compared between different countries worldwide or even in the same data format has existed. The IGLAD project took this momentum and strives for a uniform and international in-depth accident database, which is built up from the bottom on the basis of already existing databases. This is accomplished by creating a well-defined and simple layer on top of all participating databases, which serves as a common denominator of them. A more detailed description of the technical aspects can be found in [1].

History
IGLAD was initiated by Daimler AG, ACEA and different research institutes and announced as a working group at the FIA Mobility Group in October 2010. Supported by FIA and ACEA, the goal of the group is to define a common standardized accident data set as an effective foundation for developing and measuring road safety policy endorsements and interventions. It shall also establish how this data set helps to achieve the goals of the “European Road Safety Action Programme” [2] and the „Decade of Action for Road Safety“ [3].

The first IGLAD working group meeting in March 2012 comprised a more detailed discussion on the common data scheme and steps necessary for a standardized data set. A common data scheme has been drafted and as a proof of concept, a pilot study has been conducted where each data supplier converted a small set of accidents into the current version of the common data scheme data. This should show the feasibility of the approach and give a small preview of the resulting data set that could be provided by the IGLAD project. The nine countries taking part in the pilot study were: USA, India, Germany, Sweden, France, Spain, Austria, Poland, and Italy.

By end of 2012, the basic project setup had been accomplished and first technical and organizational issues had been solved, so that the first project phase could be started. Target of phase 1 was to build an initial database with at least 100 cases per country. Phase 1 was funded by ACEA and finished in mid of 2014 resulting in a first dataset of 1550 cases from 10 different countries.

Phase 2 of the project started in 2014. From now on, the project was self-containing with an own project structure and funding model. A consortium agreement was set up that reflects the different roles of all involved parties. As there is no umbrella organization for this international project, an administrator was established who could care for the correct flow of data and financial resources (figure 1). A steering group is responsible for strategic decisions and a technical working group cares about the maintenance of the database, scheme, codebook and related questions.

![Figure 1. Organizational structure of the project for phase 2 (2014 - 2016).](image)

The interesting part of the organisational structure is on the member and data provider side. Members are parties that can buy data and data providers deliver data. Of course there are parties that are both at the same time, there are data providers that are owners of their data repository and there are data providers that act in the name of another consortium or even only recode other data. This leads to different constellations in terms of financial compensation. As IGLAD is non-profit and for research purpose, special attention has to be drawn on fair balance between the data providers and members. The corresponding funding model is shown in figure 2.
Figure 2. The funding model balances interests in the project for phase 2 (2014 - 2016).

Other improvements compared to phase 1 was a simplified and unified data processing using the software Unidato as a common data acquisition tool. This allowed for extended automatic quality control using an extensive list of plausibility checks and streamlined the process of merging the data. There were also improvements in the codebook, the quality of sketches and some variables were added. The first data of phase 2 was released in 2015 containing 800 cases from 9 countries. The second dataset of phase 2 was released in 2016 with 850 cases from 9 countries. The third and last dataset of phase 2 is currently being prepared and about to be released shortly. It will contain 1,000 cases from 10 countries. This marks the end of the second phase, which was finished by the end of the year 2016 covered by the first consortium agreement. A new consortium agreement has been drafted with minor changes in the funding model and other parts and it is about to be signed by all parties of the consortium, ensuring the continuation of the project for another three years until 2019.

Status
The total database as of end of phase 2 includes 3,100 cases from 11 different countries. The 12 data providers that delivered data for it are: VUFO GmbH and BASi (Germany), Applus IDIADA Group (Spain and Czech Republic), Uni Firenze (Italy), Uni Adelaide (Australia), JP Research (India), NHTSA (USA), LAB (France), SAFER (Sweden), VSI at Graz University of Technology (Austria) and SHUFO (China), see also figure 3. CATARC as an additional data provider from China has joined the project for the third dataset of phase 2. There are also two promising data providers planned for the third phase of the project. An in-depth data project has recently started in Brazil and has already been accepted as a new data provider for the first data set of phase 3. This opens up access to accident data in South America as a new continent for the project. Just like SHUFO in China, it is a data spot that natively collects IGLAD data right from the beginning. Also, Korea as an important country in the Asian region is about to join with data from KATRI, the Korea Apparel Testing and Research Institute, which will also be collected according to the IGLAD standard right from the beginning.

Figure 3. Data providers and case counts at end of phase 2 with a total of 3100 cases.

Also the number of members is steadily increasing and currently accounts for 20 by the end of phase 2. For current information and contact details see the project’s webpage [4].

Outlook
A pilot project has been finished in 2015 that delivered more detailed information about the pre-crash phase in IGLAD [5]. The resulting pre-crash matrix contains a trajectory and various vehicle dynamics values of each participant up to five seconds before the first impact. To obtain this, the type of information needed for conducting a simulation of the accident has been investigated, which in turn generated the pre-crash data. In a next step, the subset of the IGLAD data was determined, which was ready for a pre-crash simulation. The pilot
study concluded with a simulation of these accidents and the generation of the pre-crash matrix. This data enables a benefit assessment of the performance of safety systems, especially assistance systems, in real accidents in all countries that are part of the IGLAD database. Figure 4 shows the scaled sketch of an accident in the IGLAD data. The sketch is one of the major sources of information for building the pre-crash-matrix. Below, a simulation of this accident is shown, differing from the original accident in that the car is equipped with a virtual assistance system for cross traffic situations. An additional benefit of this work is the improvement of the reconstruction data and the sketches in the IGLAD database. However, the integration of the pre-crash phase into IGLAD is still work in progress.

Generally, as an improvement for upcoming releases of the database it is planned to increase the number of countries of the participating data providers and the volume of the data, for example adding photographic documentation in addition to the sketches of each accident.

**REPRESENTATIVENESS**

Several types of analysis have been conducted with the data so far. One important focus is to check the representativeness of the data by comparing it to national data of the IGLAD data provider countries. Some results are presented here using the IRTAD [6] data for generating marginal distributions. IRTAD summarizes numbers from national statistics in selected countries world-wide and presents them in a common format while also trying to compensate for national differences, e.g. in the definition of fatalities. Fatalities are also the focus of IRTAD, which makes the comparison with IGLAD a bit difficult, as only few fatalities are contained there, yet. However a first analysis showed some interesting results.

Figures 5 and 6 show a country comparison of some grouped variables of the IGLAD data. The blue line denotes the percentage of the group within the IGLAD data and the red line shows the national statistics. If both lines overlap then the IGLAD data is likely to be representative with respect to the national statistics related to this variable.

Generally, there is a good correlation between IGLAD and the national statistics. Larger deviance is observed with passenger cars. However, this is not unexpected, as some of the country samples of IGLAD are known to be somehow biased. The reason is a different selection criteria of some of the original data. However, if representativeness is needed, then there is always the possibility to weight the results by extrapolating variables in IGLAD that are also present in the national statistics. The general long-term target is to improve the IGLAD samples in each country with respect to representativeness, quality and quantity.

**DATA ANALYSIS**

As an example of usage of the data, various analysis was conducted with the current IGLAD dataset with accidents from 2007 to 2015. Especially country comparisons are well suited for IGLAD data analysis. In the following charts, the country names are abbreviated with the two letter ISO 3166-1 codes for country names [7]. The analysis is based on 2,900 of the 3,100 accidents of IGLAD, the remaining cases were still subject to quality checks.

**Accident types and safety systems**

As a first example, accident types and safety systems were analyzed. The accident type is a very common variable also in other in-depth databases.
The IGLAD accident type is based on the GDV categories [8]. For this analysis, the ego vehicle (participant A) of each accident was selected. For simplification, it was also grouped into 30 groups, the groups are developed by Autoliv based on Najm et al [9]. For simplification these groups were then grouped into seven more general groups.

It can be observed that some countries have typically frequent accident types, like “Lange Change” in Spain and India covering half of the number of accidents. Currently, there are hardly any other sources of data that are able to provide this information, especially not in a uniform and harmonized manner.

For a closer analysis of passive safety systems, only the driver of the vehicle was selected.

Looking at the distribution of accident types with passenger cars and SUVs only, the top three countries look quite similar compared to the overall accident type distribution:

- Loss of control: AT (46%), AU (39%), SE (34%)
- Straight: CN (35%), DE (35%), SE (25%)
- Turning: US (20%), IT (18%), CN (17%)
- Crossing: US (29%), AU (28%), DE (20%)
- Lane Change: SP (61%), IN (58%), IT (31%)

In most countries belt usage is above 80%, indicating good protection of occupants and making increased implementation of seat belt reminders not an obvious priority. The low belt usage for CN and IN for example reflects the values in the WHO report [10] with 37% and 26% respectively.
In all countries, except IN, IT, SE and US at least 80% of passenger cars or SUVs are equipped with front driver airbags (figure 11). However, in the US 95% of passenger cars and 91% of light trucks on the road in 2012 were reported to have a frontal driver airbag [11] and therefore the US number reported in IGLAD seems low.

The occurrence of curtain airbags is not as common as frontal driver airbags. This is not surprising as this type of airbag was introduced to the market later, but shows potential to increase protection level. Only in AT has above 50% equipment rate in passenger cars or SUVs (figure 12).

The occurrence of curtain airbags is not as common as frontal driver airbags. This is not surprising as this type of airbag was introduced to the market later, but shows potential to increase protection level. Only in AT has above 50% equipment rate in passenger cars or SUVs (figure 12).

The existence of ESC for passenger cars and SUV’s seems somewhat low (figure 14), but almost not present at all in loss of control accidents (figure 15), which indicates that the system is effective. In EU and US it is mandatory for all new cars and LTVs since a few years back to be equipped with ESC [11], therefore it seems a bit strange that IT and US does not have any passenger cars and SUV’s with ESC, unless the accidents in IT and US in the IGLAD database involved much older passenger cars and SUV’s. Passenger cars on the road in the US in 2012 were reported to be equipped at 20% with ESC while light trucks were equipped with 22%.
Pedestrians

Out of the data set of 2,900 accident cases 375 (12.9%) cases with impacts against a pedestrian were identified. In total 390 pedestrians were recorded in the dataset (figure 16). 129 fatal and 102 severe injured pedestrian accidents are in the dataset. In 141 cases only slight injuries were recorded. In three cases no injuries to the participants are present.

312 (83.2%) accidents were recorded at urban and 63 (16.8%) at rural sites. 233 (62.1%) out of 375 pedestrian accidents happened during the day. 53 (14.1%) accidents in the data set happened at darkness lightning conditions and 20 (5.3%) at dawn/twilight. At electric light conditions 67 (17.9%) accidents are present in the data set. Mainly accidents take place at dry road conditions (288 out of 375, 76.8%). However, 81 (21.6%) of pedestrian accidents were recorded at wet road conditions. In 47 accidents out of 81 (58.0%) it was still raining.

Pedestrians coming from left are most frequent in fatal accidents (38%) having a total share of 31.9%. Pedestrians coming from right are most frequent in general (36.8%). Interestingly there are less pedestrian fatally injured (32.6%). Further accidents can be identified in situations in which the pedestrian is walking longitudinal (8.3%). Fatally injured in this situation counts for 12.4% of the pedestrians.
Looking to the body regions itself the head has the highest share of AIS 3+ injuries (23.1%) compared to the other body regions (figure 20).

The average collision speeds in pedestrian accidents were found to be at 34.7 km/h (SD=22.4) and the median is at 32.5 km/h. Quite similar to the total average collision speed are passenger cars (34.7 km/h, SD=21.5) and light trucks & vans (34.5 km/h, SD= 20.8). The average collision speed of motorized two wheelers is above the total average speed and amounts to 52.6 km/h (SD=30.4). However, collisions with motorcycles are not so frequent in the data set. The collision speed for trucks and busses is below the total average speed and amounts to 25.2 km/h (SD=29.2).

Cyclists
Out of the data set of 2,900 accident cases 234 (8.1%) cases with impacts against cyclists were identified. Further 32 accidents with electric bicycle or tricycle are present in the dataset. In total 241 cyclists and 32 electric bicycle or tricycle are in the dataset (figure 23). The electric bicycle are almost associated to the CN data provider. One accident was associated to AT. Subsequently accidents with electric bicycle are not considered in more detail.
27 out of the 234 accidents with cyclists ended up in fatally injuries. 60 cases are reported with severe and 146 cases with slight injuries. One accident present was reported without injured participants. 206 (88.0%) accidents were recorded at urban and 28 (12.0%) at rural sites. 190 (81.2%) out of 234 cyclist accidents happened during the day. 27 (11.5%) accidents in the data set happened at electric lightning conditions and 14 (6.0%) at dawn/twilight. Only two accidents were recorded at darkness. Mainly accidents take place at dry road conditions (206 out of 234, 86.8%). However, 28 (12.0%) of cyclist accidents were recorded at wet road conditions and three at glare ice/glazed frost.

Collisions with vehicles which turn into or cross a road are most frequent, independent of the injury severity (figure 24).

Collisions with vehicles which turn into or cross a road are most frequent, independent of the injury severity (figure 24).

Most frequent involved participant type in bicycle accidents in the dataset was found to be a passenger car (78.6%) (figure 25). Further relevant participants are vans (5.6%), trucks (incl. bus 4.6%) and motorized two wheelers (4.2%).

The most pronounced injured body region in the dataset was identified to be the lower (30.0%) and upper extremities (26.1%) and the head (13.5%) (figure 26).

Looking to the body regions itself the head has the highest share of AIS 3+ injuries (11.6%) compared to the other body regions (fig. 27).
The average collision speed in accidents with cyclists were found to be at 23.3 km/h (SD=21.0) and the median is at 18.0 km/h. The average collision speed of passenger cars is at 23.3 km/h (SD=21.0) and light trucks & vans is at 16.0 km/h (SD=14.0). The average collision speed of motorized two wheelers is above the total average speed and amounts to 41.2 km/h (SD=22.2). However, collisions with motorcycles are not so frequent in the data set. The collision speed for trucks and busses is at 23.0 km/h (SD=25.0).

Figure 28. Collision speed of motorized vehicles.

Without distinction of the cyclist age and the motion relative to the vehicle the driving speed is given in Fig. 29.

Collision type
The collision type of an accident describes the moving direction of the involved vehicles at the point of the first collision on the roadway. It is distinguished between the following collision types within the harmonized IGLAD data scheme:

1 - collision with another vehicle which starts, stops or is stationary
2 - collision with another vehicle moving ahead or waiting
3 - collision with another vehicle moving laterally in the same direction
4 - collision with another oncoming vehicle
5 - collision with another vehicle which turns into or crosses a road
6 - collision between vehicle and pedestrian
7 - collision with an obstacle in the carriageway
8 - leaving the carriageway to the right
9 - leaving the carriageway to the left
88888 - collision of another type

An analysis of the collision types (type 8 and 9 were treated in common) of the different countries for 2,895 of 2,900 accidents with specified collision type is shown in figure 30. The whole IGLAD database shows particularly high percentages for the type “turn into/crossing”. More than 30% of all IGLAD accidents happen due to a first collision on a junction. Where Australia shows 32%, China 50% Germany 39%, India 27%, Italy 30% and the United States 48%, this spots a very frequent scenario for the majority of the included countries.

Figure 30. Collision type per country, all vehicles.
The second largest group with 17% of all IGLAD accidents happen due to “leaving the carriageway”. Except in China, Italy and India, it is the 1st or 2nd most frequent collision type in all other countries. Although being half the proportion of “turn into/crossing” accidents, this group is very important to address regarding vehicle safety. Due to leaving the carriageway, they mostly come along with a collision against an obstacle next to the road (or a rollover) resulting in serious injuries. Safety systems like e.g. the Electronic Stability Control (ESC) are able to address some of these situations. ESC has a noticeable penetration of the market in general by now (see Figure 14), but anyhow countries with high penetration (Austria, China, Germany, France, Sweden, or Spain) do not all show small proportions of collision type “leaving the carriageway”. This could be addressed and might show some benefit for newer systems. Examples are “Lane Departure Warning System” or “Lane Keep Assist System”. With 16% of all IGLAD accidents the third most frequent group is a “collision with another oncoming vehicle”. Since overtaking is one of most frequent reasons for such collision type, newer advanced driver assistance systems and Car2Car communication systems can potentially address these situations. Especially India and Spain show noticeably high proportions. However, one has to keep in mind the small sample by now (n = 100 for both countries). Nevertheless, this collision type often results in very serious injuries due to the high relative speed between the two vehicles and the corresponding high Delta-v of each participant.

**Speed difference (Delta-v) distribution**

The Delta-v (Δv) is the vector difference between immediate post-crash speed and pre-crash speed of a participant. As this parameter correlates well with the injury severity, figure 31 shows the Delta-v proportions per MAIS class (Delta-v was grouped to classes of 10 km/h). The MAIS gives the maximum of all single values of the abbreviated injury scale (AIS) for a body region (AIS90 update AIS98) per occupant. The analysis is considering 4,313 occupants in cars (passenger car, SUV, light truck, van) with known Delta-v and known MAIS. The gray bars show the proportion of all occupants, yellow of all occupants with MAIS = 1, orange MAIS = 2, and dark orange all occupants with MAIS greater or equal to 3 (also known as MAIS3+).

![Figure 31. Delta-v (grouped) per MAIS, cars only.](image)

The figure shows that the majority of all car accidents happen in the area of low Delta-v. Also higher proportions of accidents with more severe injuries (especially MAIS3+) are located at the higher Delta-v area. That was to be expected as the Delta-v parameter is known to correlate well with the injury severity. Nevertheless, the most severe injuries are in the area of high Delta-v and thereby there is a big potential for saving lives and reducing number and severity of injuries in global traffic accidents. This information is hardly available for several countries and areas at once and in a uniform and harmonized manner outside the IGLAD database.

![Figure 32. Boxplot of Delta-v per MAIS group, cars only.](image)

Moreover, the IGLAD database enables the analysis of Delta-v distribution per country. Figure 32 shows a
boxplot (also box-and-whisker plot), including median, first and third quartiles and the whisker per country. Values are displayed in light orange for all MAIS = 1 and MAIS = 2 (MAIS 1-2) car occupants and in dark orange for all MAIS3+ car occupants. Case numbers for all groups are given on the top. Countries are sorted ascending by median value for MAIS 1-2 occupants. Values for the whole IGLAD database are between purple vertical lines. As the IGLAD database is existent since mid of 2014 its eligible case number is not sufficiently high for all countries. Especially all MAIS3+ values must be considered carefully. Countries and areas with a less modern vehicle fleet that do not include all current passive safety systems (see section accident types and safety systems) seem to have a lower Delta-v median for MAIS3+ injured occupants. This means, occupants injuries of MAIS3+ occur at a lower Delta-v. Corresponding to this, countries with modern vehicles like Austria, Germany, Spain or Australia, show very high Delta-v median values (e.g. around 60 km/h for MAIS3+ injured occupants), suggesting an effect of the introduction of passive safety systems. Nevertheless, it is important to assess all statements critically and carefully as most data shows large quantiles and whisker due to partially small available case number.

CONCLUSIONS

Global road safety becomes more important with the continuously growing vehicle fleet in the emerging markets. While the number of fatalities in countries with advanced technology and infrastructure is decreasing an opposite trend in emerging markets can be observed. The IGLAD project addresses this problematic from the perspective of the accident researcher where the global road accident data needs to be enlarged and harmonized. This should increase the capabilities to more quickly and efficiently identify measures for improvement of global road safety. The project has been started successfully, the first two phases have been already passed and a database of 3100 cases from 11 different countries has been built. A sustainable organization and funding model has been established. The next steps are further improvement of volume and quality and increased usage of the data. Some data analysis on typical topics for in-depth databases has been conducted and IGLAD has proven to be well suited for these analysis targets, especially when it comes to country comparisons. Finally, IGLAD continues to be a source of in-depth road accident data for different groups of researchers that strive for improving global road safety.

REFERENCES