Verband der Automobilindustrie

Forschungsvereinigung Automobiltechnik

FAT-Schriftenreihe 368

Objective assessment of database quality for use in the automotive research and development process – Part 2

Objective assessment of database quality for use in the automotive research and development process – Part 2

Forschungsstelle

Verkehrsunfallforschung an der TU Dresden GmbH

Autoren

J. Ziegler, P. Miklis, D. Grosche, T. Unger

Das Forschungsprojekt wurde mit Mitteln der Forschungsvereinigung Automobiltechnik e.V. (FAT) gefördert.

Acknowledgments

The Research Association for Automotive Technology (FAT) has commissioned the Traffic Accident Research Institute at TU Dresden GmbH (VUFO) to work on the mentioned research project. Vehicle manufacturers and numerous suppliers have joined in the FAT to conduct pre-competitive and joint research under the framework of the German Association of the Automotive Industry (VDA).

The initiation of this research project was originated from the FAT AK3 working group by Dr.-Ing. Michael Düring (Volkswagen AG) and Michael Wagner (Continental AG). The two FAT project leaders were always valuable idea generators. Their constructive criticism and approaches contributed significantly to the success of this research project. Our first gratitude belongs to them.

Our second gratitude belongs to all members of FAT AK3 working group, who contributed to the development of the Global Safety Database with the research on international road accident databases and statistics and many approaches during the interim meetings.

I. Table of contents

II. Abstract

Road traffic accidents remain to be a leading cause of death worldwide with roughly 1.3 million fatalities annually [1] . To develop new safety approaches according to real-world challenges, accurate information is needed. Therefore, road safety experts are constantly looking for real-world data to answer the open challenges and ultimately reach “Vision Zero”.

The Global Safety Database (GSD) offers access to an one of its kind up-to-date repository of road traffic accident statistics and databases on a meta-data level for road safety analyses.

One main objective is the compilation of international data sources, for which a data management system has been developed. In addition to the inventory of road accident data sources, a questionnaire created by road safety experts is used to check the applicability of data sources for specific questions. Therefore, an automated and dynamic matching process enables comparing variables representing the questions with the existing data source content in the GSD. The results are stored in a result matrix which indicates the proportion of variables that correspond to the variables necessary to answer the research question for each data source investigated.

In order to identify similarities and differences in road safety within the countries, a clustering methodology is developed to point out the possibilities and limitations of projecting information from the initial countries to other areas. The assessment of the representativeness of the individual data sources is the basis for the clustering.

From a general perspective, the GSD is an essential tool pushing forward the worldwide harmonisation of traffic accident statistics and databases. Knowledge about the real-world accident scenery by bringing important databases together empowers the data-driven development which is eventually a key bringing us one step closer to a road system without casualties, the achievement of the Vision Zero.

III. Task

The task of this research project is based on the results of the FAT publication 343 “Objective assessment of database quality for use in the automotive research and development process – Part 1” and is subdivided into several tasks, which are described more in detail in the following report:

- Identification of data sources in the context of vehicle safety and road traffic accidents
- Contact and consultation of data providers with a detailed interview on the investigation methods, data content and quality management of the data source
- Brief descriptions of the researched accident databases and statistics
- Analysis of individual data sources and countries for a comparison in terms of clustering possibilities of certain data sources on regions
- Implementation of a web-based dynamic database and development of proposals for authorisation and usage concepts (e.g., for FAT members and external users); implementation of individual access authorisations, possibility of updating and checking by different user roles

IV. Abbreviations and indices

Symbol	**Description**
ARDD	Australian Road Deaths Database
BRICS	Brazil, Russia, India, China and South Africa (Association of Emerging Economies)
CASR	Centre for Automotive Safety Research
CIDAS	China In-Depth Accident Study
DESTATIS	Federal Statistical Office
DPAC	Defective Product Administrative Center
ECIS	Enhanced Crash Investigation Study
FAT	Research Association for Automotive Technology (Germany)
FRSC	Federal Road Safety Corps
GIBDD	Gosavtoinspekcija (State Traffic Safety Inspectorate)
GIDAS	German In-Depth Accident Study
GSD	Global Safety Database
IAAT	Investigação Avançada de Acidentes de Transito (Advanced Traffic Accident Investigation)
ID	Identification
IRSMS	Integrated Road Safety Management System
MORTH	Ministry of Road Transport and Highways
NAIS	National Automobile Accident In-depth Investigation System
NPA	National Police Agency
PDF	Portable Document Format
PTW	Powered two-wheelers
RASSI	Road Accident Sampling System India
RENAEST	Registro Nacional de Acidente e Estatísticas de Trânsito (National Registry of Traffic Accidents and Statistics)
RTMC	Road Traffic Management Corporation
SC	Steering Committee
UAG	Ulykkesanalysegruppe (Accident Analysis Group)
USA	United States of America
VRU	Vulnerable road user
VUFO	Traffic Accident Research Institute at TU Dresden GmbH
WHO	World Health Organization

1. Introduction

In the research and development of the automotive industry, numerous interdisciplinary aspects are considered, particularly in the field of vehicle and traffic safety. Relevant issues of assisted, connected and automated driving and the further development of passive and integral safety systems require reliable data sources. The sometimes very heterogeneous traffic and accident situation in different countries and continents requires taking as many as possible data sources from several countries/regions into account as possible. However, it is not always obvious which data source is suitable for which kind of research question or development approach.

The aim of the presented work is to develop a unified meta database that contains all necessary information for the research and development departments on a meta based level (no raw data) for several countries. The database is designed on a dynamic platform that allows updates to the data sources to have an immediate effect on the query with the questionnaire.

One of the main objectives is the research on international data sources in the field of traffic and vehicle safety. This includes national road accident statistics based on police accident data as well as highly detailed investigations in smaller regions (so-called In-depth data sources).

In addition to the development of a meta database, a questionnaire is used to check the applicability of the developed meta database for specific questions from the German automotive industry. The objective assessment between the meta database and questionnaire requires to develop a matching process. The aim of the matching process is to indicate a percentage of coverage of necessary variables in the selected data sources for each question. The results of the matching process are collected in a result matrix.

Based on the developed meta database and the matching process, the result matrix offers possibilities for objective assessments of data sources and provides the opportunity for data providers to improve their data quantity and data quality. Furthermore, the meta database can act as a platform to bring several data providers from different countries together and to encourage the global harmonisation of traffic accident data sources. For this purpose, an usage and management system will be set up in order to improve and control data quality and the search for new data providers and road safety experts.

In addition to the development of the database structure and user administration, the representativeness of individual data sources is examined and compared with the country-specific accident figures and investigation methods. The developed clustering procedure supports the process of searching for representatives in regions or for individual countries. The data analysis of the meta database is used to highlight special features from data sources around the world.

2. **Research on data sources**

The research on data sources for road traffic accidents differs between national databases and in-depth databases. The national data sources consist of data collected by the police or similar national road traffic or road safety authorities. These data sources contain a huge number of accidents and provide a macroscopic view of the accident scenario. National data sources focus on collecting general accident-related data to monitor the overall accident situation and assess high level safety topics and support regulatory efforts.

In-depth data sources take a different approach, in which data providers primarily investigate how the accident may have occurred. However, their sole goal is to get as much information about the accident to answer the “why” and “how” without forcing or supporting a criminal prosecution. These data sources are mostly characterized by a smaller number of cases, but a much higher level of detail in the investigated road traffic accidents and thus provides a microscopic view of the accident scenario.

Depending on design criteria such as the sampling plan and the number of investigated cases, well designed in-depth data sources offer the opportunity to extrapolate their accident scenario to the national level in order to give representative statements.

Figure 2-1 shows an overview of the difference of national and in-depth data.

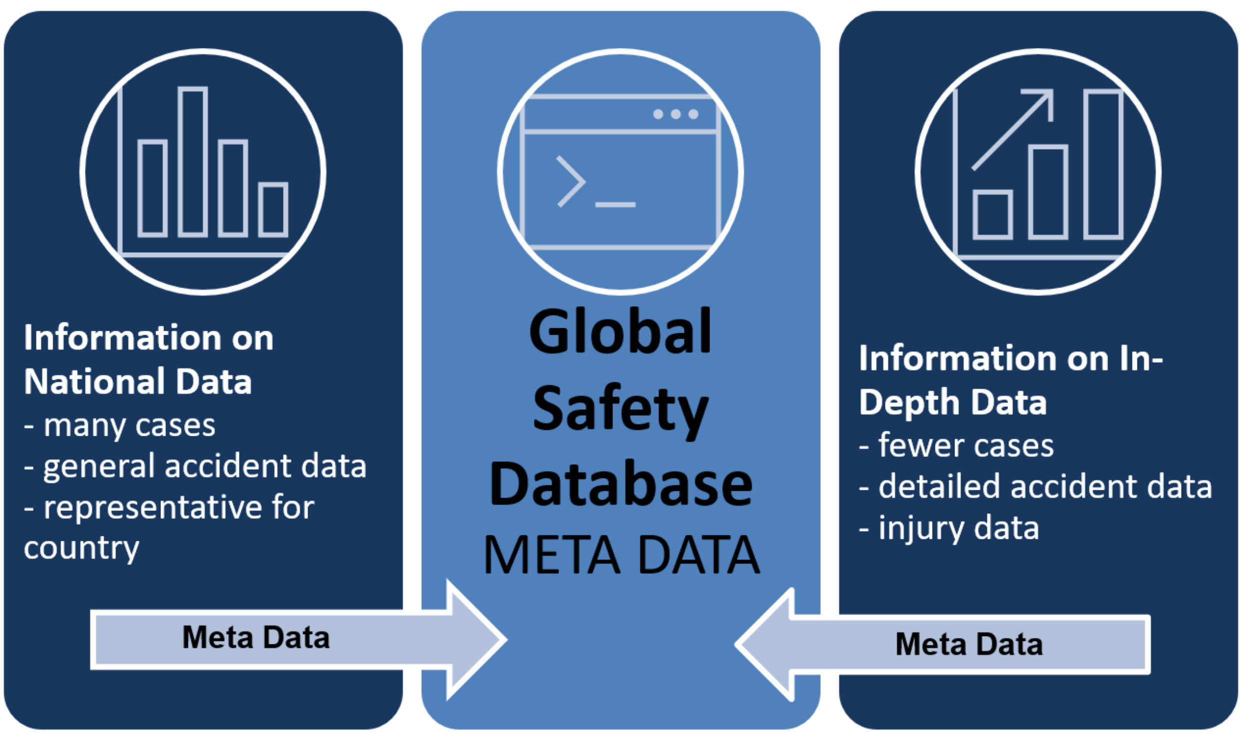


Figure 2-1: Differences in national and in-depth data sources

The Research Association for Automotive Technology (FAT) defined specific countries for a focused research on data sources. The first project part "Objective assessment of database quality for use in the automotive research and development process" [2] investigated Germany, France, Greece, Czech Republic, Sweden, Denmark and the United States of America. The second part focused on countries preferably outside of Europe. As the Association of Emerging Economies include interesting countries such as Brazil, Russia, India, China and South Africa (BRICS-Countries), they were chosen to be covered. This choice was extended by countries on all continents in order to cover wide areas of the planet. The continents Europe and America were deliberately given less attention, as these two continents were part of the first project.

As a result, the following countries were selected: Australia, Indonesia, Japan, Norway and Nigeria + BRICS countries.

2.1. National data sources

Based on the country selection, the following section investigates national data sources in particular. Basically, all data sources share a common parameter base, such as timeline, general accident properties, participant types and injury levels. However, differences can be found on closer inspection. Therefore, the codebook, the general conditions of investigation, the context and the definitions have been reviewed.

A first overview of the research of national data sources is given in Table 2-1.

Table 2-1: Investigated national accident data sources

Country	Geographical coverage	Accidents/year	Timeline	Surveys	Comment
Brazil (RENAEST)	Country	890,000	2021-today	Police report	Online statistics
Russia (GIBDD)	Country	130,000	2005-today	Police report	Online single case access
India (MORTH)	Country	450,000	2009-today	Police report	Annual reports
China (NAIS)	7 cities	1,200	2011-today		Single case investigation; in-depth
South Africa (RTMC)	Country	10,000	2005-today	Police report	Report mainly based on fatal crashes only
Australia (ARDD)	Country	1,500	1989-today	Police report	Single cases; only fatal accidents and fatalities
Indonesia (IRSMS)	Country		2015-today	Police report	Single cases
Japan (NPA)	Country		1948-today	Police report	Reports online; fatal and injury accidents

Country	Geographical coverage	Acci-dents/year	Timeline	Surveys	Comment
Nigeria (FRSC)	Country	9,000	1988-today	Police report	Annual reports
Norway (NORSTAT)	Country		1939-today	Police report	Fatal and injury accidents; no suicide cases and accidents due to ill health included

All examined data sources cover an entire country, either because they consist of police reported accidents or because the database is maintained by the authorities responsible for traffic accidents themselves. The main exception is the Chinese data source NAIS (National Automobile Accident In-Depth Investigation System). This source is managed by the Defective Product Administrative Center (DPAC) at the national level, but only examines traffic accidents in selected cities and thus does not cover the entirety of accidents in the country. Also, it does not only consider police reports, but also in-depth investigation for individual cases.

Most sources contain information regarding time, weather, lighting conditions, type of road and of the area. In addition, some sources contain information about the type of vehicle or the type of accident and sometimes even about the type of injury caused by the accident. Exceptions in Table 2-1 are the countries South Africa [3] and Nigeria [4] , as the reports focus on different aspects related to transport infrastructure and not only on accidents. Therefore, these reports only include the number of accidents that occurred during the year and the number of fatalities in these accidents.

Another aspect for national data sources is the ratio of accidents per year. Most sources have up to 10,000 cases per year, which allows them to achieve a high degree of representativity.

Many sources shown in Table 2-1 provide their results periodically via publicly available reports, which can be accessed in PDF data format. Some exceptions are RENAEST from Brazil [5] and GIBDD from Russia [6]. The Brazilian data source provides its results through a Ministry of Infrastructure website in the form of a dashboard. There, all Brazilian states can be viewed individually or as a whole and filtered according to various parameters. The Russian data source is also similarly available. However, after selecting the desired region, the desired information can be retrieved for each individual accident. The information consists of the accident location and time, involved vehicles and persons and the accident type with a scheme. Furthermore, the data on road and weather conditions and vehicle data can be retrieved. Finally, there is information about the participants themselves, where they were sitting and whether they were wearing seat belts, for example. Personal data, such as names and age or injury data, is not available.

2.2. In-depth data sources

In-depth accident data sources are mainly investigated by interdisciplinary research units. They often consist of a technical team supplemented by a medical and/or psychological unit. The former investigates the technical aspects of the accident (e.g., infrastructure, vehicle) and the medical part investigates the injuries suffered by the persons in the accident and the physical/psychological backgrounds as well as the mental consequences caused by the accident.

Most data providers of in-depth data sources reconstruct their accident scenario in order to know the most probable accident sequences. Due to the increased effort in the data recording and reconstruction, the number of accidents is smaller than in national data sources.

According to the defined country selection, the following in-depth data sources have been researched (Figure 2-2).

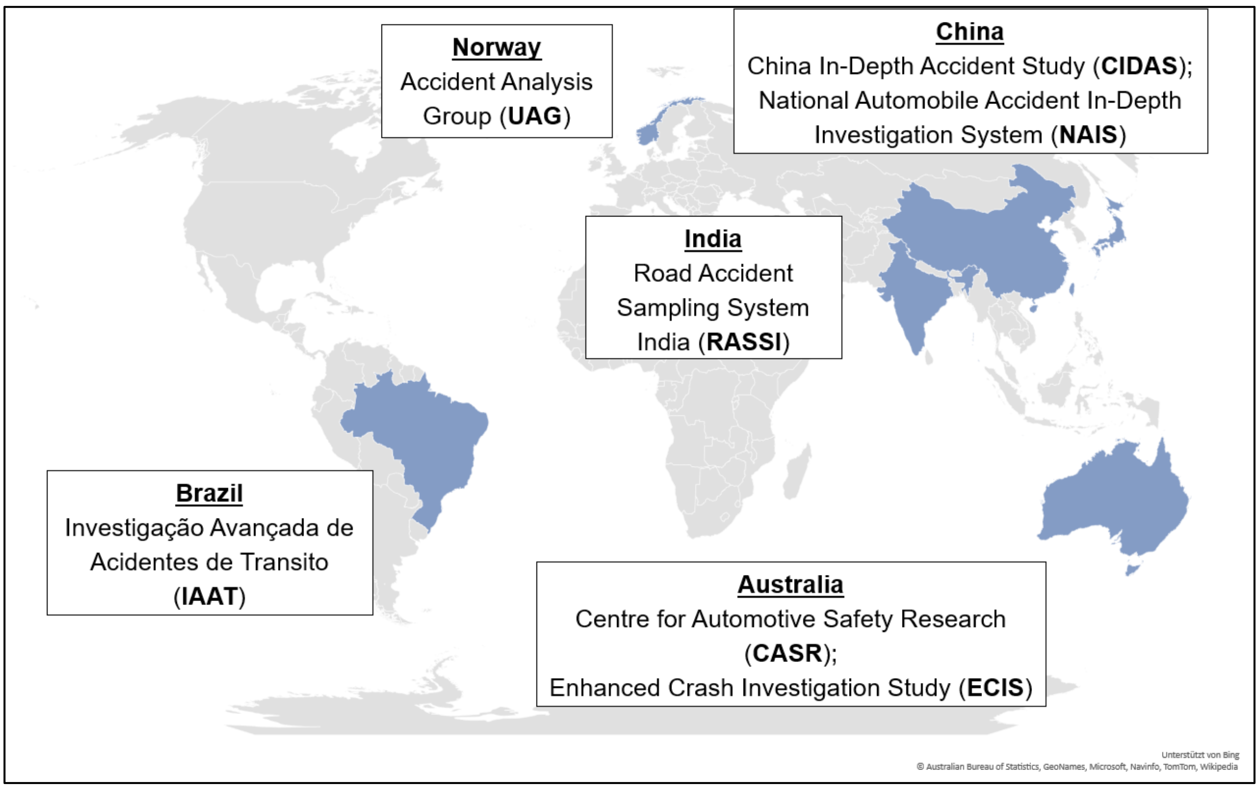


Figure 2-2: Researched in-depth data bases

A first overview of the research of in-depth data sources is given in Table 2-2.

Table 2-2: Investigated in-depth accident data sources

Country	Geographical coverage	Acci-dents/year	Timeline	Access	Comment
Brazil (IAAT)	Campinas, Sao Paulo	100	2016-today		
India (RASSI)	6 areas	500	2011-today	Members only	
China (CIDAS)		100	2011-today	Members only	
Australia (CASR)	Adelaide + 100 km	100	1976-today		
Australia (ECIS)	Victoria	200	2014-2016		Accident study; crashes in remote locations and at night are generally excluded
Norway (UAG)	Country	100	2005-today	public	Fatal accidents

The main difference between in-depth and national data sources is the geographical coverage. While in national data sources the entire country is considered, in-depth data sources usually only consider specific areas. As an example, the IAAT database from Brazil can be mentioned here. The survey locations of the accidents and data are limited in this data source to the cities of Campinas and Sao Paulo.

Many in-depth data sources, unlike national data sources, collect detailed data on injuries and medical treatment of participants in accidents. To reconstruct these accidents, a lot of information is needed. Therefore, the in-depth data sources examined show that far more variables are considered here. As an example, the national and in-depth data sources of Germany are shown in Figure 2-3 with the differences in the number of variables and accordingly the depth of the data.

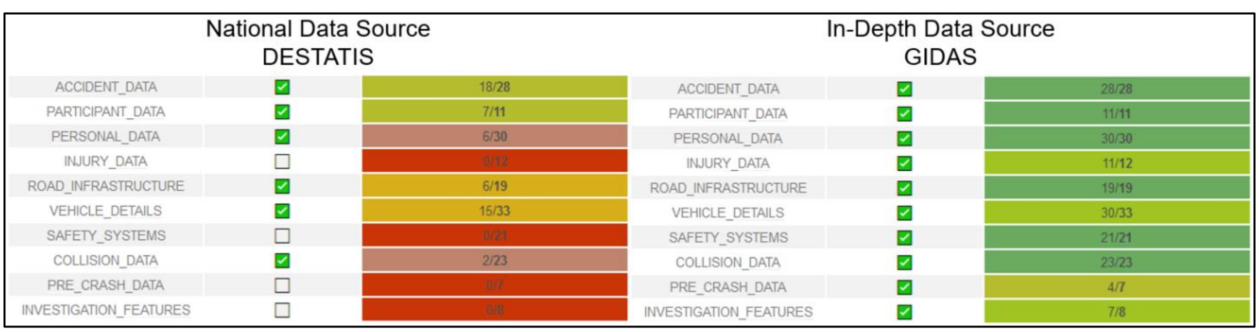


Figure 2-3 Differences in variables of national and in-depth data

In-depth data sources are often not publicly available. Exceptions in the Table 2-2 are the data sources UAG [7] from Norway and ECIS [8] from Australia. In the case of ECIS, the data source is a project with a duration of two years and a total of 200 accidents from the state of Victoria. The results are made publicly available in the form of a report.

The in-depth investigation of road traffic accidents in Brazil (IAAT) is operated by SAE Brazil in cooperation with Robert Bosch Latin America [9]. In India (RASSI) the data is maintained by JP Research India [10]. The responsible research unit for Chinese in-depth data (CIDAS) is the China Automotive Traffic and Research Center [11]. Two in-depth data sources were found in Australia. The Centre for Automotive Safety Research investigates accidents in the area around Adelaide (CASR) [12] and has its own database on accidents. In the state of Victoria, the in-depth accident investigation was conducted with a cooperation of the Transport Accident Commission and the Monash University Accident Research Centre (ECIS).

2.3. Research process

As a first step, considered countries were investigated regarding national accident data sources. Primarily official authorities for road safety or police authorities were contacted. In addition to searching for official authorities, studies or in-depth databases were searched and contacted. With the incoming answers, new contacts or data sources kept coming up. These new sources were also investigated via contacts or via online searches. In some cases, individual discussions with the responsible persons for a specific data source brought new insights whereas in other cases, contact was further developed via email and information was exchanged. Thereby, the response rate and available information was much better in some countries than in others.

A particular problem that can be mentioned here is mainly the response rate. It was often not possible to establish contact. Another problem was that after contacting the competent authority, the person responsible for the data source did not respond. In two cases, the database was being relaunched or redeveloped at the time being and the promised help was postponed until the time of completion.

On the other hand, exchanges with well-connected countries have also been successful. In addition to personal conversations, codebooks were sometimes provided. However, even checking codebooks had its complications. Language barriers were a problem, especially in countries with their own idiosyncrasies. But here, too, it was possible to turn to the responsible authorities and individuals of the respective country for help.

Figure 2-4 shows an overview of the research process.

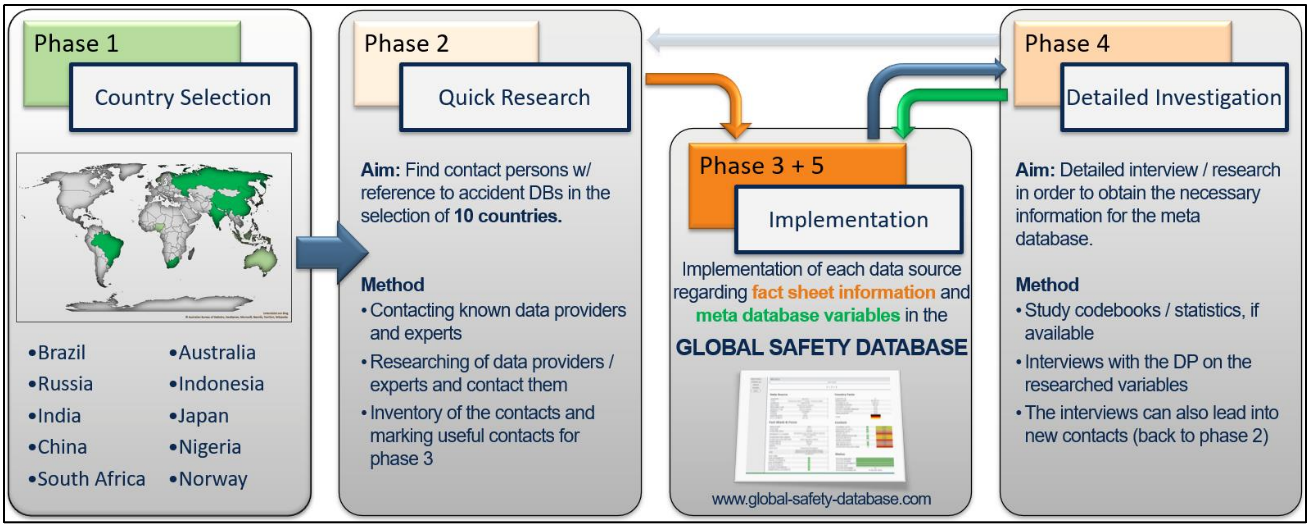


Figure 2-4: Overview research process

It was possible to contact all relevant data source providers in the selected countries, mainly via e-mail. Some of these contacts led to a personal meeting, in other cases new contacts were arranged. However, contact with most countries was rather difficult, as in some cases the contact could either not be established or the responsible department did not respond. In contrast, contact with the countries Australia, India and Japan was very good and helpful.

A general remark from almost all data source responsible was that they did not want to release their codebooks. However, they were very cooperative in personal contact and in asking specific questions about desired variables. Thus, we were able to solve issues understanding or getting access to the meta data for many data sources. In other cases, the metadata could also be derived from published reports from the respective data source.

3. **Representativeness and clustering**

One essential aspect of the investigated data sources is representativeness. This concludes the investigation method for road traffic accidents in certain countries or regions which is directly linked to the possibility of extrapolation the data to other regions or countries of the world.

For this purpose, the individual investigated data sources in the Global Safety Database (GSD) are analyzed and the data providers are interviewed about the own assessment of representativeness. Based on the findings, a quantitative and qualitative evaluation method is developed to provide information about the representativeness for each in-depth data source, when the national data source is known.

Based on the methodology for the assessment of representativeness, a cluster method is developed for the evaluation of the global road safety by using different safety factors. The result should provide a tendency which countries or regions can be projected to other countries or regions due comparable or similar safety standards. The goal of this approach is to point out the limits of comparability and to identify further steps to transfer the countries already researched to other countries or regions which are mainly unknown.

3.1. Representativeness

The representativeness is defined diversely, depending on the scientific mission and the goal to be aimed for. Within the GSD, the representativeness is defined by two key aspects:

- the existence of a “suitable” sample plan and
- the unbiased chance of any accident to be part of data base.

A data source is representative if it is possible to draw conclusions from a sample to the totality, i.e. when certain elements of the totality have the same chance of being part of the sample. Therefore, the official road traffic accident statistics or national data sources collected by the police form the totality and represent the road accident scenario for a country or region.

The sample of an in-depth data source corresponds to accident investigations in a certain region, which may also be investigated by the police, but are mostly carried out by in-depth collection units, which do not include all police-recorded accidents.

One example of a sampling in-depth data source is the GIDAS database [13]. GIDAS collects only accidents with personal injury according to a fixed sampling plan. However, GIDAS works with a time-based sampling, which leads to the fact that not all accidents meeting the criteria will be part of the data base. Due to that fixed sampling plan in the GIDAS database, every police-reported accident has the same chance to become an accident of the GIDAS database.

The criteria for the GIDAS sampling plan are the following:

- The accident must be an accident with personal injury.
- The accident must happen within the GIDAS investigation area.
- The accident must occur within the shift periods.
- Always, the last police-reported accident is investigated.

In addition to the data source to be assessed (sample), the assessment of representativeness also requires the appropriate totality, which contains similar or even the same criteria. Criteria are a set of variables that describe the accident scenario and should be present in the sample as well as in the totality.

Accident-describing criteria can be, for example:

- Accident severity (fatally, seriously, slightly injured)
- Accident location (urban, rural, motorway)
- Kind / Type of accident (descriptive criteria for the accident event)
- Number of participants (single accident, accident with at least 2 participants)
- Kind of road users (4-wheeled vehicle, PTW, bicycle, pedestrian, others)
- …

Depending on the country or region, the criteria described above can be defined differently or other criteria have to be used. The most important aspect is that the criteria must be identical between the two analysed data sources (total source and sample source).

For the quantitative assessment of representativeness, the number of accidents, participants or persons depending on the chosen criteria have to be known from the sample data sources as well as from the totality. In various in-depth accident databases, this information is usually not freely available. Most of the variables from in-depth data sources underly restrictions such as being a member or an official user. Within the framework of this project, the access to the following in-depth data sources according to our definition are available: GIDAS (Germany), CRRS (USA)

Due to lack of information on the number of accidents depending on the chosen criteria especially for the in-depth data sources, a qualitative assessment is more feasible. Nevertheless, a quantitative methodology has been developed to show a method to assess representativeness.

In the following example, the GIDAS database is evaluated quantitatively in relation to the official road accident statistics for Germany 2020 [14] by considering five criteria described above. Only accidents with personal injury are considered from the German statistic. For this evaluation, GIDAS data is used from January 1999 until June 2021, which includes all fully documented and reconstructed accidents with personal injury (Table 3-1).

Table 3-1: Distribution of accidents and participants in the official road accident statistics for Germany 2020 and GIDAS (un-/weighted)

	Official German accident statistics 2020		**GIDAS 07/21** (unweighted)		**GIDAS 07/21** (weighted)	
ACCIDENT LOCATION*						
Urban	183.412	69,3%	31.443	78,8%	183.412	69,3%
Rural	65.850	24,9%	6.250	15,7%	65.850	24,9%
Motorway	15.237	5,8%	2.222	5,5%	15.237	5,8%
	264.499		**39.915**		**264.499**	
ACCIDENT TYPE						
1xx	54.128	20,5%	7.246	18,2%	54.128	20,5%
2xx	35.730	13,5%	6.301	15,8%	35.730	13,5%
3xx	56.524	21,4%	10.302	25,8%	56.524	21,4%
4xx	10.665	4,0%	3.535	8,9%	10.665	4,0%
5xx	8.824	3,3%	1.714	4,3%	8.824	3,3%
6xx	60.810	23,0%	7.994	20,0%	60.810	23,0%
7xx	37.818	14,3%	2.823	7,0%	37.818	14,3%
	264.499		**39.915**		**264.499**	
ACCIDENT CATEGORY*						
Slightly injured	210.674	79,6%	27.681	69,3%	210.538	79,6%
Seriously injured	51.243	19,4%	11.493	28,8%	51.195	19,4%
Fatally injured	2.582	1,0%	741	1,9%	2.766	1,0%
	264.499		**39.915**		**264.499**	
NR. OF PARTICIPANTS						
Single accident	63.946	24,2%	7.410	18,6%	61.402	23,2%
accident w/ ≥2 participants	200.553	75,8%	32.505	81,4%	203.097	76,8%
	264.499		**39.915**		**264.499**	
KIND OF ROAD USERS*						
4-wheeled vehicle	321.597	65,4%	50.742	65,9%	347.120	69,3%
PTW	38.319	7,8%	5.821	7,5%	38.789	7,7%
Bicycle	100.159	20,4%	13.973	18,1%	82.544	16,5%
Pedestrian	26.130	5,3%	5.433	7,0%	25.329	5,1%
Other / Unknown	5.278	1,1%	1.170	1,5%	7.214	1,4%
	491.483		**77.139**		**491.483**	

** Criteria with special penetration on the accident scenario*

In addition to the information on the number of accidents, the number of injured persons and the number of participants, further boundary conditions are defined for the quantitative assessment of representativeness. The official road accident databases and the chosen criteria with parameters from these statistics are considered as the totality (=1). The assessment is determined by a score calculated from the ratio of the sample database to the totality depending on the criteria and parameters within the criteria. The result is shown in a radar chart with the maximum and minimum deviation.

Depending on the application and user group, factors / criteria with special penetration on the accident scenario may be determined. Within the framework of the project, these criteria are marked with this symbol (*).

The comparison of the relative shares within the categories results in the ratio as well as the maximum and minimum value for this category (Table 3-2).

Table 3-2: Ratio between official road accident statistics for Germany 2022 and GIDAS (un-/weighted) by category with maximal and minimal value

	RATIO Official German accident statistics 2020 vs. GIDAS (unweighted)	MAX	MIN	**RATIO** Official German accident statistics 2020 vs. GIDAS (weighted)	MAX	MIN
ACCIDENT LOCATION*						
Urban	1,1	1,1	0,6	1,0	1,0	1,0
Rural	0,6			1,0		
Motorway	1,0			1,0		
ACCIDENT TYPE						
1xx	0,9	2,2	0,5	1,0	1,0	1,0
2xx	1,2			1,0		
3xx	1,2			1,0		
4xx	2,2			1,0		
5xx	1,3			1,0		
6xx	0,9			1,0		
7xx	0,5			1,0		
ACCIDENT CATEGORY*						
Slightly injured	0,9	1,9	0,9	1,0	1,0	1,0
Seriously injured	1,5			1,0		
Fatally injured	1,9			1,1		
NR. OF PARTICIPANTS						
Single accident	0,8	1,1	0,8	1,0	1,0	1,0
accident w/ ≥2 participants	1,1			1,0		
KIND OF ROAD USERS*						
4-wheeled vehicle	1,0	1,4	0,9	1,1	1,3	0,8
PTW	1,0			1,0		
Bicycle	0,9			0,8		
Pedestrian	1,3			1,0		
Other / Unknown	1,4			1,3		

Based on calculation for the ratio per parameter as well as on the determination of the minimum and maximum value, a radar graph is established for better visualization.

The Figure 3-1 compares the unweighted GIDAS data set with official road accident statistics for Germany from 2020. The radar graph shows that each category has variance. The largest range between maximum and minimum values is found for accident type (Δ=1,7) and accident category (Δ=1,0).

The weighting of the GIDAS data set is necessary to achieve an approximation to the official German road accident scenario. Depending on the accident reporting by the police and accident selection, there may be an overweight and/or underweight of accident scenarios. Based on a weighting process, representative statements can be made with the GIDAS data set.

The Figure 3 2 illustrates that for categories accident location, accident type and accident category the differences between the official road accident statistic for Germany 2022 and GIDAS are minimal. The basis is that the GIDAS data is weighted on these three mentioned categories. This type of weighting has also a positive influence on the range between minimum and maximum value for the number of participants.

The range for the type of road users remains the same in this weighting process. It is conceivable to extend this weighting process to include the mentioned category in order to become more representative to the German accident scenario with the weighted GIDAS data set.

The challenges in the quantitative assessment of representativeness are that each national statistic and in-depth data source must be considered separately. Depending on the country-specific variables in the accident investigation, different schemes are necessary or country-specific properties must be brought into an existing scheme.

Obtaining the data basis for the national statistics as well as for in-depth data sources can also cause challenges. Almost every in-depth data source does not provide freely accessible information on the number of investigated accidents according to specific categories. In-depth data sources are often not freely accessible and are reserved for the members only. Consequently, a data lake is caused by the non-freely accessible data sources. Data and access lacks also exist on the national level.

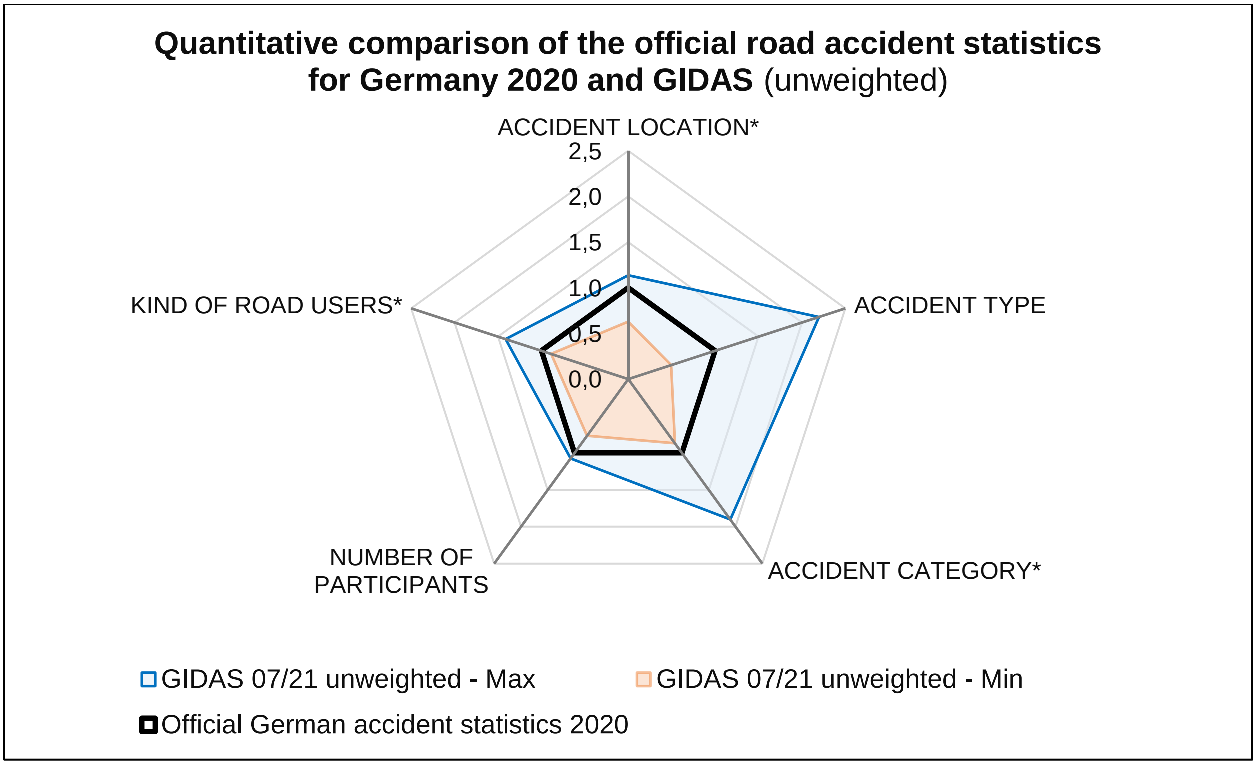


Figure 3-1: Quantitative comparison of the official road accident statistics for Germany 2020 and GIDAS (unweighted)

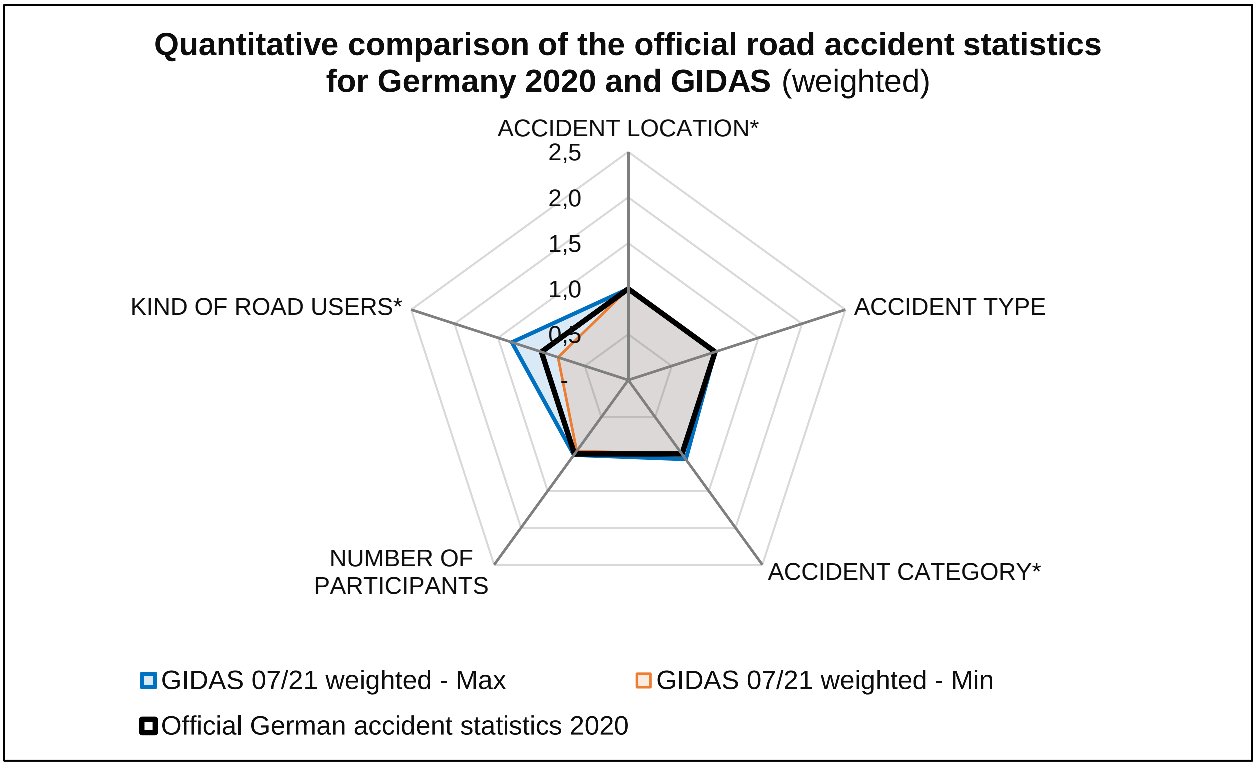


Figure 3-2: Quantitative comparison of the official road accident statistics for Germany 2020 and GIDAS (weighted)

Due to lack of information/data on the number of accidents depending on the chosen categories and comparative databases, a qualitative assessment is more feasible, especially if the GSD is used for the consideration. This method is not based on the number of accidents, or the number of persons involved in the accident, but rather the availability of information / variables is compared.

For this purpose, a collection of variables is extracted from the GSD and compared for the national as well as for the in-depth data sources of each country. The variable set can be chosen according to requirements of the analyses. Within the framework of this project, the following variables from the GSD were used for the qualitative assessment (Table 3-3).

Table 3-3: Variables set from the GSD for the qualitative assessment of representativeness

Variable description	Variable GSD	Classification
Accident month	ACCMONTH	Accident data
Weekday	WDAY	
Daytime	DAYTIME	
Accident location	LOCATION	
Maximum accident severity	ACCSEVERITY	
Accident type	ACCTYPE	
Number of participants	NR_PARTICIPANTS	
Weather conditions	CLOUD_FOG_WIND	
Participant type	PARTTYPE	Participant data
Age of the person	AGE	Personal data
Gender of the person	GENDER	
Injury severity per person	IN_SVERITY	

As a result of the research on data sources (Research on data sources), it was found that not every country has national and in-depth data sources. Consequently, the qualitative assessment is only carried out for the following countries (Table 3-4).

Table 3-4: Countries from the GSD research for the qualitative assessment

Countries	Nr. of national data sources	Nr. of in-depth data sources	Qualitative Assessment
Germany	3	9	possible
Sweden	2	2	possible, but few information missing
Denmark	2	0	not possible
Greece	1	1	possible
France	1	2	possible
Czech Republic	1	1	possible
USA	4	5	possible
Brazil	1	1	possible
Russia	1	0	not possible
India	1	1	possible
China	0	2	not possible
South Africa	1	0	not possible
Australia	3	3	possible, but few information missing
Indonesia	2	0	not possible
Japan	1	1	possible
Norway	2	1	possible
Nigeria	1	0	not possible

The methodology for the qualitative assessment is based on comparing the predefined variables (Table 3-3) between the national data source and the in-depth data source. In this process, it is only queried whether a variable is present in both data sources. If a variable is present in both the national and in-depth data source, it may be possible to use this variable for a representativeness assessment of the in-depth data source to the national statistics. The following fictional example illustrates the methodology (Table 3-5).

Depending on the variables on which a representativeness assessment is carried out, some countries (e.g., Germany, Greece, France, Czech Republic, USA, Brazil, Australia, Japan) show better preconditions than other countries (e.g., Sweden, India, Norway). If the representativeness assessment is limited to the available variables and defined, then statements for representativeness can also be possible with these data sources, which have currently fewer matches to the official data sources or statistics.

The described qualitative assessment method is based only on the availability of variables. The parameters within the variables are not considered in this qualitative method. But in a quantitative assessment of representativeness, the parameters within a variable have to be taken into account.

Table 3-5: Qualitative assessment of national data sources and In-Dept data source on a schematic basis

	ACCMONTH	WDAY	DAYTIME	LOCATION	ACCSEVERITY	ACCTYPE	NR_PARTICIPANTS	Weather	PARTTYPE	AGE	GENDER	INJ_SEVERITY	Total	**SHARE OF MATCHES**
GERMANY														
National data source	1	1	1	1	1	1	1	1	1	1	1	1	12	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	12/12 = **100%**
SWEDEN														
National data source	1	1	1	1	1	1	0	0	1	0	0	0	7	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	7/12 = **58%**
GREECE														
National data source	1	1	1	1	1	1	1	1	1	1	1	1	12	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	12/12 = **100%**
FRANCE														
National data source	1	1	1	1	1	1	1	1	1	1	1	1	12	
In-depth data source	0	0	0	1	1	1	1	1	1	1	1	1	9	9/12 = **75%**
CZECH REPUBLIC														
National data source	1	1	1	1	1	1	0	1	1	1	1	1	11	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	11/12 = **92%**
USA														
National data source	1	1	1	1	1	1	1	1	1	1	1	1	12	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	12/12 = **100%**
BRAZIL														
National data source	1	1	1	1	1	1	0	0	1	1	1	1	10	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	0	11	9/12 = **75%**
INDIA														
National data source	1	0	0	1	0	1	0	1	1	1	1	1	8	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	0	11	7/12 = **58%**
AUSTRALIA														
National data source	1	1	1	1	1	1	0	0	1	1	1	0	9	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	9/12 = **75%**
JAPAN														
National data source	1	1	1	1	1	1	1	1	1	1	1	1	12	
In-depth data source	1	1	1	1	1	1	1	1	1	1	1	1	12	12/12 = **100%**
NORWAY														
National data source	1	1	0	0	0	0	1	0	0	1	1	1	6	
In-depth data source	1	1	1	0	1	0	1	0	1	1	1	1	9	6/12 = **50%**

3.2. Clustering

The assessment methodology for the representativeness is used to compare national and in-depth data sources within a country. We investigate the feasibility of clustering by comparing and analyzing similarities and differences in road safety between two or more countries.

The main objective of the cluster analysis is to find out which countries have the closest similarity to a country selected for the analyses (target country) based on road safety standards. But it is not intended to carry out a full clustering calculation. The basis of the clustering method within the GSD is a simple comparison with rank assignments.

The clustering is carried out according to the following safety standards:

- Population density *in inhabitants / km²*
- Gross Nation Income *in US dollars per capita*
- Fatality rate I *in deaths per 1 Mio. inhabitants*
- Fatalities by road user *in percentage*
- Vehicle rate *in vehicles per 1 Mio. inhabitants*
- Distribution of reg. vehicles *in percentage*
- Fatality rate II *in deaths per 1 Mio. vehicles*
- Safety standards in traffic *speed limitation, belt usage, helmet law*

The basis of the road safety standard is the latest WHO global status report on road safety for 2018 [1] . In total, 175 countries are mentioned in der WHO report, which are considered in the clustering process within this project. Not every country contains full information to all road safety standards. The boundary conditions for data collection by the WHO are mentioned in the report and can be found there.

In order to identify the countries with the closest similarity to regions or other countries, a ranking system is developed. A methodology checks within the safety standards which country is closest to the target region or country. The relative ratio between challenging country and target country/region identifies the deviations. The countries with the closest distance to the target earn the highest scores. The score depends on the number of countries to be compared. The average score of all safety standards results in the final rank for each challenged country. The highest average score gets the ranking position 1 and is closest to the target country.

The example of the Scandinavian countries shows which country is the closest to Norway by using four of eight safety standards (Table 3-6). Depending on the chosen safety standards, Sweden has the closest approximation to Norway. By using all safety standards, Sweden is also the closest country to Norway.

In addition to the selection of road safety standards, a weighting procedure has also been implemented to give special significance to some safety factors. However, for general consideration, the weighting has been deactivated, so that each road safety standard has the same significance.

Table 3-6: Calculation of the ranking for the cluster methodology using the example of Norway on four of eight safety standards

	Population density		Score	Fatality Rate		Score	Vehicle Rate		Score	...
Norway	**14.1**			**26.31**			**730,307**			
Sweden	23.3	+65%	3	26.54	+1%	4	582,606	-20%	4	
Denmark	137.0	+871%	1	38.58	+47%	3	532,275	-27%	3	
Finland	16.4	+16%	4	46.85	+78%	2	940,142	+29%	2	
Iceland	0.4	-97%	2	583.12	+2116%	1	7,673,373	+951%	1	
Fatalities by road user	Drivers/Passengers of 4-wheeled vehicles			Drivers / Passengers of 2- or 3- wheelers			Cyclists			...
Norway	**50%**			**17%**			**9%**			
Sweden	54%	+4%	3	16%	-0,7%	4	8%	-0,8%	3	
Denmark	48%	-2%	4	16%	-0,9%	3	15%	+6%	2	
Finland	65%	+15%	2	9%	-8%	1	10%	+0,7%	4	
Iceland	72%	+22%	1	11%	-6%	2	0%	-9%	1	

	Ø SCORE	TOTAL SCORE	RANK
Norway			
Sweden	**3,5**	**4**	**1**
Denmark	2,7	3	2
Finland	2,5	2	3
Iceland	1,3	1	4

According to the described clustering method, the Table 3-7 shows which countries are most similar to the countries from the project country selection by bordering countries, continents and worldwide.

Table 3-7: Countries with the closest similarity according to the chosen road safety standards by bordering countries, continents and worldwide

	Bordering countries	Continent	Worldwide
Brazil	Suriname	Suriname	Iran
Russia	Kazakhstan	Iran	Iran
India	Pakistan	Myanmar	Myanmar
China	Kazakhstan	Iraq	Iraq
South Africa	Botswana	Tunisia	Tunisia
Australia	New Zealand	New Zealand	Canada
Indonesia	Timor-Leste	Timor-Leste	Timor-Leste
Japan		Qatar	Netherlands
Nigeria	Benim	Ghana	Tajikistan
Norway	Sweden	Sweden	Sweden

The clustering process not only map which country is most similar to the target country. It also offers the possibility to find out which country within a region most closely represents the region according to the average values of the safety standards. For this purpose, the countries of a region under consideration are combined and the mean value of the respective safety standards is calculated based on the country-specific individual values.

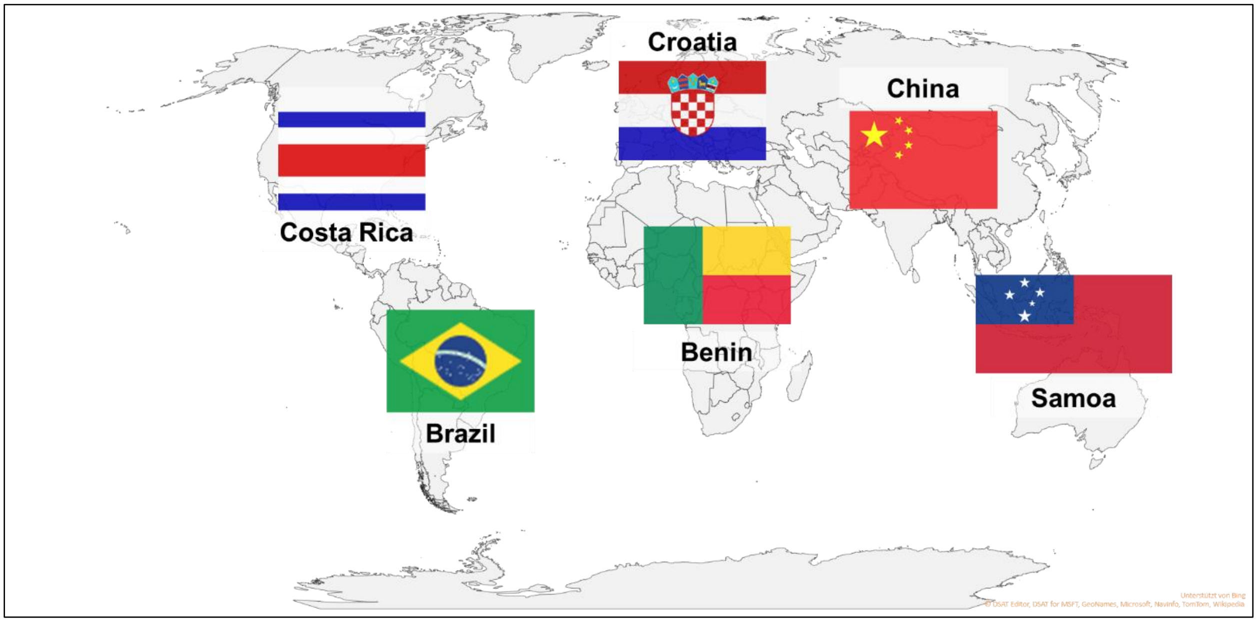
In the example, all countries of Scandinavian are combined to form Northern Europe. The global average value for this region is given from the individual values in the Table 3-8. The individual scores are then compared with the average and the scoring points are calculated. According to the country selection, Sweden is the most suitable representative for North Europe.

Table 3-8: Procedure for clustering within a region using the example of Northern Europe

North Europe	Population Density		Score	Fatality Rate		Score	Vehicle Rate		Score	...	ØSCORE*	RANG
	Ø 38.2			**Ø 33,96**			**Ø 683.350**			...		
Norway	14.1	-63%	3	26.31	-23%	3	730.307	+7%	5	...	3.0	2
Sweden	23.3	-39%	5	26.54	-22%	4	582.606	-15%	4	...	3.6	1
Denmark	137.0	+258%	1	28.58	-14%	5	532.275	-22%	3	...	3.0	2
Finland	16.4	-57%	4	46.85	+38%	2	940.142	+38%	2	...	3.0	2
Iceland	0.4	-99%	2	583.12	+1617%	1	7.673.373	+1023%	1	...	2.3	5

*further safety standards are taken into account in the calculation

This variant of the cluster method for determining representatives for selected regions is carried out for each continent. The result is shown in Figure 3-3.



Area	*Representative*	*Area*	*Representative*
Europe	Croatia	**World**	
North	Sweden	*Asia*	China
East	Latvia	*Africa*	Benin
South	Croatia	*North America*	Costa Rica
West	Netherlands	*South America*	Brazil
Central	Czech Republic	*Oceania*	Samoa

Figure 3-3: Representative by region according to the clustering method

Based on the WHO global status report on road safety for 2018 [1] , it is possible to compare countries and/or regions with each other. However, the described clustering method is also linked to boundary conditions that influence the result.

The final score as well as the average values for the region are based on individual values, whereby outliers, such as Iceland, influence the methodology. With a relatively large area and few inhabitants, Iceland has many values that would have to be defined as outliers. In a further development, it should be considered how such outliers should be considered in the clustering in the future. Furthermore, the average seems to dominate in the current clustering scheme. The weighting of the individual safety standards / data is essential.

For a solid argumentation for future developments in road safety, up-to-date data is mandatory. The current clustering calculation is based on data from 2018, of which the accident data is from 2016. That means current influences (e.g., covid pandemic, climate change, energy transition) can rarely be considered. Furthermore, some of the countries listed in the WHO global status report on road safety do not provide safety standards and/or accident data. Consequently, some data are assumed, calculated and extrapolated by WHO. [1]

3.3. Scientific analysis

The current GSD meta database contains 103 data sources from 40 countries. Two thirds of the inventoried data sources are national databases or statistics. The remaining data source are in-depth databases. Depending on the countries commissioned as well as the success in researching data sources, the content of each data source differs. In the following countries, the research has collected a considerable quantity of information: Australia, Brazil, China, Czech Republic, France, Germany, Greece, Japan, Netherlands, USA

By filling the GSD with meta information from the individual data sources, some variables are coded more or less frequently. Table 3-9 shows which variables are triggered more or less frequently in the GSD.

Table 3-9: Variables from the GSD with the most or less frequently coding

Most coded variables in the GSD		**Most non-coded variables in the GSD**	
GSD variable	***Description***	***GSD variable***	***Description***
ACCYEAR	Accident year	PARKSYS	Parking/maneuvering systems
PARTTYPE	Participant type	Vehicle2X	Vehicle to X communication
AGE	Age of the persons	VIDEO	Accident video
ROADTYPE	Type of road	LASERSCAN	Laser scan of the accident scene
ACCMONTH	Month of the accident	TRAFFICDENS	Traffic density
GENDER	Gender of the persons	AIS1990	Abbreviated Injury Scale from 1990
ACCTYPE	Type of the accident	INJIMAGE	Existence of pictures from medical imaging techniques
ACCTIME	Time of the accident	MUEMAX	Maximum possible deceleration
WDAY	Weekday of the accident	REGISTDATA	Data of registration paper
DAYTIME	Daytime of the accident	AUTOMTION_LEVEL	Automation level

Depending on the efforts of the data providers, it can be shown which variables are the "standards" and which variables bring the differences between the data sources. These variables are partly special in the investigation of accident data (e.g., investigation of road traffic accident by video sequences of roadside cameras), which makes them and their data source unique.

The GSD has been developed as a tool for researching suitable data sources on new topics in road safety. In the past as well as in the present, every research and development department used its own researched sources of accident data worldwide. The GSD aims to centralise this effort. In addition, the GSD also provides the research and development departments information on countries that may not be known yet.

The latest publication of the accident scenario in the European Union (EU) from 2022 reports that the share of fatally injured cyclists, motorcyclists and pedestrians increase. [15] From 2010 to 2020, the proportion of pedestrian fatalities increased by 0.6% and cyclist fatalities increased by 3.4% compared to the total accident scenario. The shares of fatalities of car drivers and others indicate a decreasing trend. The proportion of motorcyclist (PTW) fatalities increases by 0.9%.

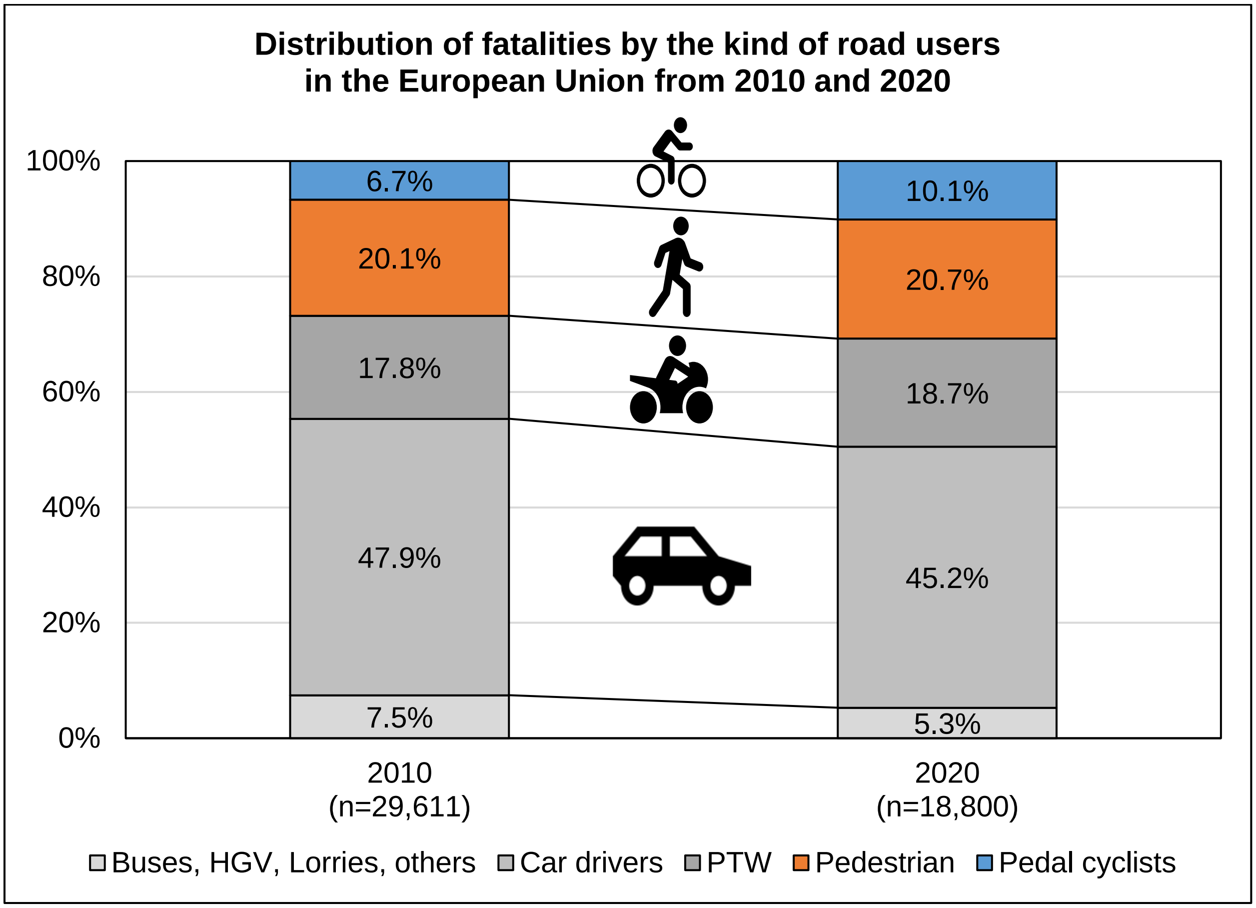


Figure 3-4: Distribution of fatalities in the European Union from 2010 and 2020 [15]

For the development of new safety systems and safety functions, the accident scenario with the vulnerable road users (VRU) will become more important in the future. VRUs includes cyclists, motorcyclists and pedestrians. For this purpose, the meta database and the extension of the GSD with individual questions could serve as basis on the existence of pedestrian, cyclist and PTW data in the EU and worldwide. Cross-comparisons between different countries or regions in the EU or worldwide can help to identify new safety aspects.

One of these fundamental aspects for the safety of cyclist and motorised two-wheelers is the helmet usage and the injury severity in accidents with and without helmet. The example demonstrates how search criteria (variables) can be used for the GSD to find the suitable data sources for the VRU challenges. In addition, the results are objectively evaluated by the matching process.

In order to map the trend of the VRU accident scenario in different countries and data sources, the GSD (Figure 3-5) is filtered by data sources that contain the following information:

	GSD variable
• Accident year (Derivation of a trend)	[ACCYEAR]
• Accident scenario = Accident type	[ACCTYPE]
• Information on the accident scenario of VRUs included	
○ Cyclists	[CYCLIST_ACCIDENTS]
○ Motorcyclists	[PTW_ACCIDENTS]
○ Pedestrians	[PEDESTRIAN_ACCIDENTS]
• Information on the accident opponent	[PARTTYPE]
• Injury severity	[INJ_SEVERITY]
• Helmet usage	[HELMETUSE]

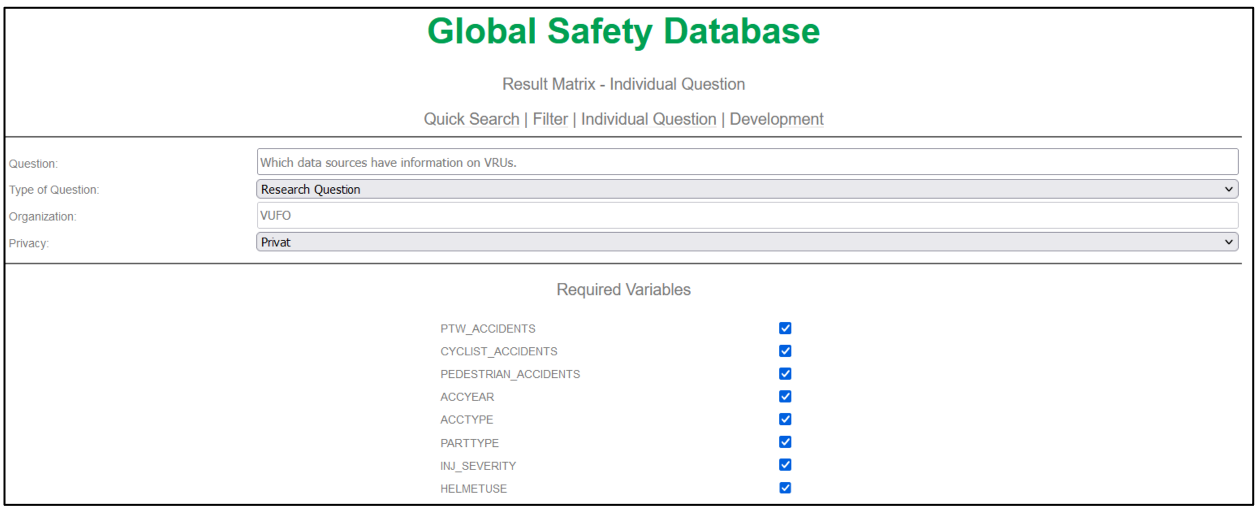


Figure 3-5: Extract from the GSD on individual questions on VRU accidents

In the area of the result matrix on the GSD webpage, an own research question can be generated under the tab "Individual questions". Therefore, the required variables must be selected (Figure 3-5). A detailed instructions are described in the appendix “Adding Individual Question”.

In order to preserve the competitive equilibrium between the data providers, Table 3-10 shows which countries are best suited to answer the individual question for VRU accident scenario by 100% coverage. Normally, the comparison is done between the data sources. Consequently, the number of countries in Table 3-10 may be smaller than the number of data sources.

Table 3-10: Countries with 100% coverage for the query on the VRU accident scenario

Continent	***Country / Countries***
Africa	South Africa
America	USA
Asia	India, Japan
Europe	Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Spain

The Figure 3-6 shows which other countries are included in the result matrix for VRU accident scenario compared to Table 3-10. According to the matching process, the other countries have a coverage of less than 100%. The countries coloured in grey in Figure 3-6 have a coverage of 0% or they are not inventoried in the GSD yet.

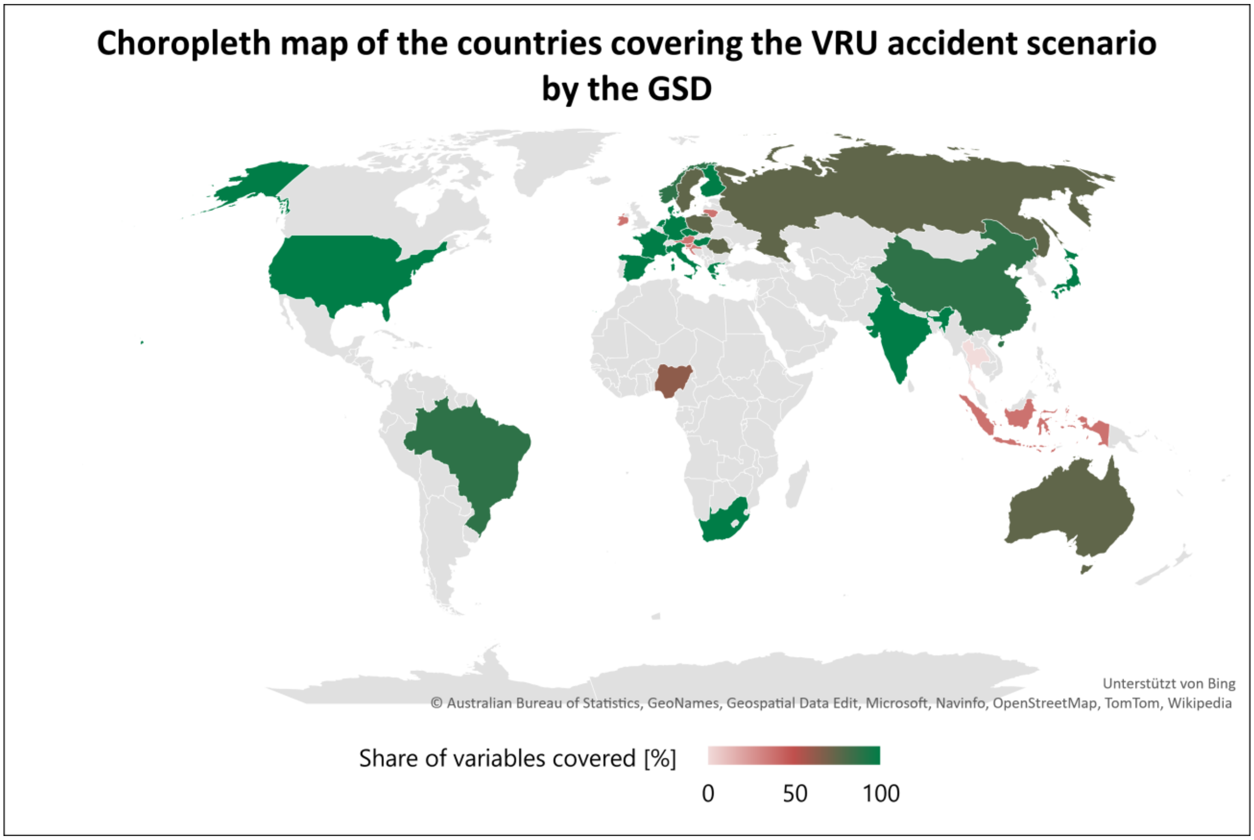


Figure 3-6: Choropleth map of the countries covering the VRU accident scenario by the GSD

The next step could be to check if the access to the selected countries and data sources is already available or have the access to be provided by contacting the data provider. For this purpose, the meta database contains the information on the contact details (e.g., link to the data source, contact person). Further processing and acquisition of the data is the responsibility of the analyser.

By using the cluster method, the countries of the database can be checked to ensure that the results for the VRU accident scenario could also stand for other regions or even for an entire continent. Consequently, a continental comparison would also be conceivable as long as the raw data are available and the representativeness is known.

4. **Dynamic and web-based database**

The GSD is a one of its kind platforms to gain up-to-date insights into existing real-world traffic accident statistics and databases worldwide for research and development of future vehicle and road traffic safety systems.

Therefore, vehicle and road traffic safety experts are constantly looking for real-world data sources to tackle/answer the open challenges. However, it is not always obvious which data source is suitable for which type of research question or development approach, as there is currently no harmonized overview of real-world traffic accident data sources nor an objective evaluation available. To our knowledge, the GSD is the first application to enable its users to efficiently identify suitable databases for specific research questions worldwide.

For this purpose, it is necessary to offer users a web-based database with real-time analysis of data sources and compare them dynamically with a questionnaire. The world of road safety experts and other interested persons will be provided a web-hosted website, where authorized persons can make changes to the database or the questionnaire. If one of the following changes are made…

- Changes in the data sources or questionnaire, or
- Add new data sources or questions to the GSD,

… than the adjustment is addressed dynamically without the administration having to intervene. For this purpose, user groups must be defined that are allowed to read and/or change certain content depending on their user rank. Therefore, the GSD administration needs to ensure the data quality and establish a control process to prevent uncontrolled growth with incorrect or inaccurate information.

The development and establishment of the GSD has faced various challenges, which have been considered and addressed by the project partners in consultation with the project owner. The main requirements of the GSD are to generate a high coverage of data sources, statistics and information of road safety worldwide and to provide a high quality of data at the meta level.

Every potential user should receive the same access rights (read/write) regardless of the organisation, while complying with the contractual boundary conditions. Advancement opportunities in the user rank can be requested by anyone.

Another constraint is the design of the website, which should be user-friendly on the one hand and self-explanatory on the other. Furthermore, every change in the GSD will be tracked and checked to ensure data quality. Therefore, a Steering Committee (SC) of GSD experts is established and a control procedure is developed. Every user is allowed to make suggestions for improvement, which will be checked by the SC.

All boundary conditions must be observed in compliance with data protection law, corporate identity (CI) and compliance guidelines. A manual for working with the GSD is provided in the appendix (Manual "Global Safety Database").

4.1. Utilisation and authorisation concept

The GSD is meant to be used by everyone, who is interested in accident data and statistics regardless of the organisation. The general scheme of the GSD's usage and authorisation management system is shown in Figure 4-1.

The access request is done via online form, where interested users deposit personal data and create an account with their own password. Thereby it is completely uninteresting to which user group the new member belongs. Some examples for the user group are given in Figure 4-1, although there may be more user groups (e.g., private persons, legislation).

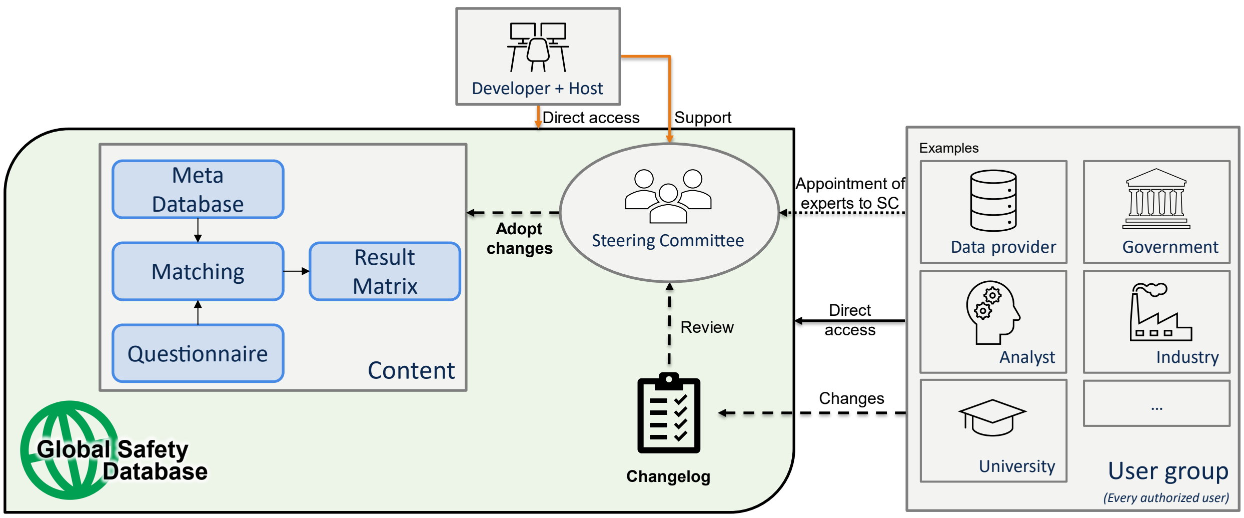


Figure 4-1: Utilisation and authorisation concept of the Global Safety Database

All new members of the GSD are first given the user access of a guest. Persons with these statuses have the possibility to see all public accessible data sources (e.g., national statistics), but not to change. The non-public accessible data sources (e.g., mainly in-depth data sources) as well as non-public questions in the questionnaire are not visible for guests.

The guest status is intended to give new members of the GSD a first impression. Members of the GSD who are interested in obtaining a higher rank must apply to the SC. The following ranks are distinguished in Table 4-1.

The next level after the guest access is the authorised access, which is granted after application to SC and is activated either by the head of SC or by the developer on assignment of the SC. This user group can read and write in the GSD as well as use the question catalogue to the full extent. Only the administrative objects are not visible.

Every change that is made by authorised users and higher ranks in the meta database as well as in the questionnaire is recorded in the change log and will lead in a review. Steering committee members have access to the changelog and are responsible for reviewing the changes. Therefore, a distinction is made between small (e.g., two-man rule) and larger (clarification within the SC) reviews.

Table 4-1: User ranks and permissions in the GSD

User classification	Only read	Read & Write	Full access of meta database	Add new data sources	Full access to questionnaire	Add new ques-tions	Access to the changelog	Access to the user management	Access to review panel
Developer	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Head of Steering Committee	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Member of Steering Committee	-	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Authorised GSD member	-	Yes	Yes	Yes	Yes	Yes	No	No	No
Guest	Yes	No	No	No	No	Yes	No	No	No

Each change in the GSD meta database as well as in the questionnaire is automatically a small review and must be checked by two SC members. If both SC members agree, the small review is either accepted or discarded. Therefore, a review panel has been developed. If the two SC members do not agree, the small review becomes a larger review and must be reviewed by the entire SC. In large review the head of SC have the final power of decision.

Requests to change of variable or parameter definitions or requests for new variables are submitted by the GSD users, recorded in the changelog and reviewed by the SC.

The SC is a group of appointed experts for road traffic accident databases and statistics. Further experts can be appointed by the SC based on experience with the GSD by checking the members activities via the changelog. The head of the SC is determined by the SC members. Main tasks of the SC members are

- Review of changes in the meta database in regular intervals
- Exchange on larger reviews in a group of international road safety experts
- Checking requested access authorisations
- Exchange with the developer of the GSD
- Representative representation of the GSD in other committees/working groups
- Responsible for the content on the GSD front page

The developer of the GSD can be also the host of the website, or it works separately. The main task of the developer is to implement reported bugs, variable or parameter changes and user management from the SC into the GSD. At the same time, the developer has to give feedback to the SC about possible irregularities and problems with the GSD.

4.2. Structure of the GSD website

The structure of the GSD website is subdivided into six main sections (Figure 4-2), with different access rights depending on the user rank (see 4.1- Utilisation and authorisation concept). The front page of the GSD can be reached by https://www.global-safety-database.com/ and serves as a basis of information for interested persons. The latest news and publications of the GSD are published at the front page and all persons are informed about the benefits of the GSD. Registration and later the login to the internal page can be accessed via the front page.

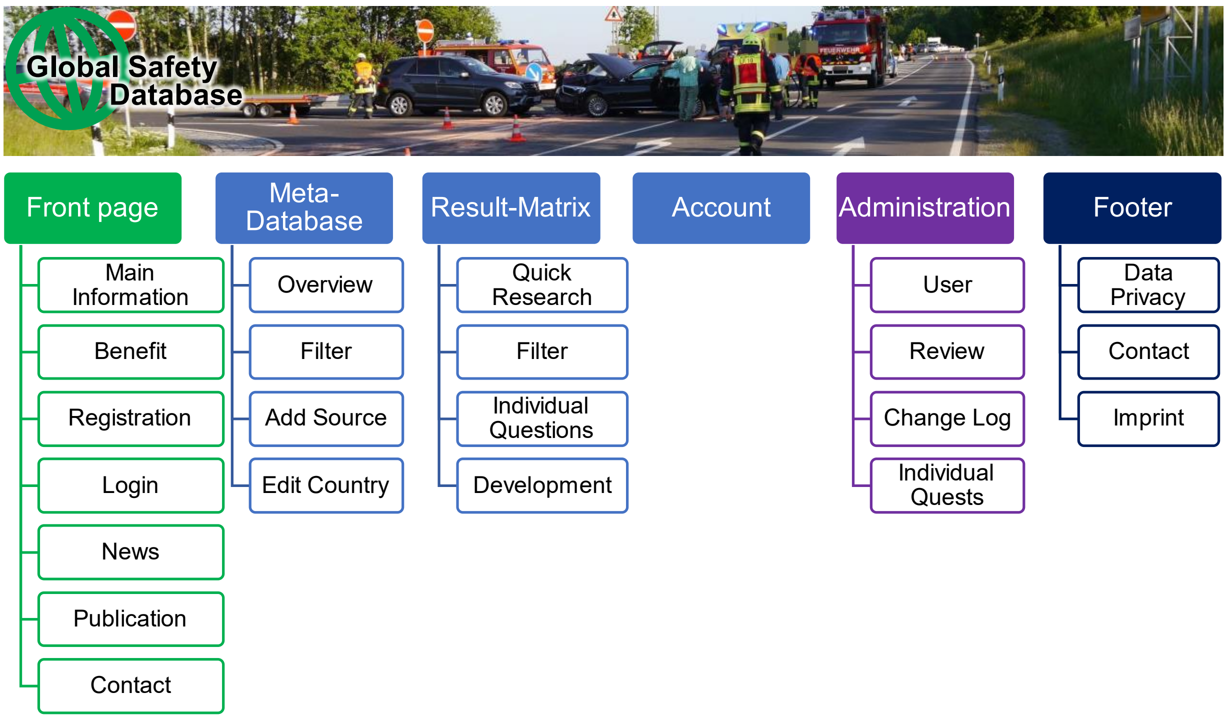


Figure 4-2: Structure of the Global Safety Database

Every registered user has access to the meta database of the GSD. Depending on the rank of the user, visibility is restricted distinguishing between public and non-public data sources. The definition therefore can be found in 4.1- Utilisation and authorisation concept. Regardless of user rank, the structure in meta database is visible, but partly restricted for guests.

Authorized user will be able to change data or add new data sources, which will be recorded in the changelog. In addition to the GUI for changing and adding data, the user can also use the GUI for filtering data sources. The export of data can be offered for single data sources.

The questionnaire is visible and usable for everyone. In contrast to the guest account, every authorized user has access to the whole questionnaire with private and public questions. Guests only see public questions from the questionnaire.

Public questions are a combination of individual private questions with the same scientific background. Private questions are too specific to share with the guests. Each user or organization is obliged to provide “Public Research Questions”, when using the matching algorithm for their own private research questions. Each coding of public research question is subject to review in the SC.

The matching process, which compares the questionnaire with the content of the meta database, is a function in the background of the website. By implementing a matching algorithm, the dynamic comparison is feasible. Whenever a change in the meta database or in the questionnaire is made, the dynamic algorithm automatically takes all changes into account and creates a new result matrix when the file is queried. The result matrix indicates the percentage of the respective data source that covers the variables needed to answer a certain question.

The scientific method for the matching process and the result matrix is described in detail in the project phase I [2] . The major difference between project phases I and II is the dynamic algorithm, which is integrated into a web-based platform. The project phase I is based on a static environment with a correspondingly static result matrix. After each change in the metadata database or in the questionnaire in the project phase I, the matching process must be carried out manually.

In the header “Account”, the users find their access data and can change their password. In addition, all individually questions that have been created with this account are collected there. In the footer of the website, the usual legal requirements are linked and explained.

The main difference between the normal GSD users and the SC members is the GUI for the administration. This interface provides the SC members a tool in which user administration can be carried out and records from the changelog can be reviewed. In addition, new questions can be viewed and checked.

By using the contact form, GSD users can report problems or irregularities. Furthermore, the contact form should also be used to apply for an authorized access to GSD or for a SC membership.

5. **Conclusion and Outlook**

The GSD aims to investigate, inventory and objectively evaluate numerous accident data sources from different countries all over the world. Thus, it provides data for top relevant questions of todays and tomorrows road safety. For this purpose, a meta database is designed in project phase I [2] and filled with data sources of different kinds and origins. Within the project phase II the following countries are researched: Brazil, Russia, India, China, South Africa, Australia, Indonesia, Japan, Norway and Nigeria

The subsequent objective evaluation is based on a matching process, which matches the content of current research questions of a questionnaire with the availability of variables in each data source. The result is stored in a so-called result matrix that shows the availability of necessary variables per data source for each and every research question.

An innovative methodology for clustering is used in order to identify countries or regions which data may be used to be transferred to other countries with similarities evaluating the road safety factors. Therefore, the representativeness of data sources especially of in-depth data sources is an essential aspect as knowledge of one country or region may accelerate the road safety improvement in other regions of the world as well.

As the GSD comprises only meta data, representative statements may also only be formulated on a qualitative level. The GSD is designed to only store meta data of data sources. This meta data is not sufficient for a qualitative assessment which requires raw data. Within the GSD, a qualitative assessment has been co-developed to show the possibilities how representativeness could be assessed with raw data if it is accessible.

In order to support the development process, the GSD is meant to be used by everyone, regardless of the position or organisation. The GSD is freely accessible and the data quality as well as user management is ensured by the voluntarily voted Steering Committee.

From a global perspective, the GSD is one essential tool to push forth the worldwide harmonisation of traffic accident statistics and databases. The GSD is designed to be an ongoing meta database as well as questionnaire that engages its users by sharing information on global data sources for road traffic accidents and upcoming research questions for road safety. Thus, a community of experts may be formed and grow steadily. Furthermore, the GSD aims to support the increasingly data-driven vehicle development process by greatly reducing the effort finding required and suitable data sources to answer the top relevant research and development questions. This novel community is a key puzzle piece for enabling right developments and decision for future generations of safety measures, regulations, and systems. Thus, a Vision Zero with no road traffic casualties is getting closer and closer.

Appendix

List of literature

[1] World Health Organization. “Global status report on road safety 2018” Geneva: World Health Organization; 2018; Licence: CC BY-NC-SA 3.0 IGO; ISBN 978-92-4-156568-4

[2] Ziegler, J.; Liers, H.; Chanove, A.; Pohle, M. “Objective assessment of database quality for use in the automotive research and development process”. FAT publication series 343, Berlin, 2021

[3] Road Traffic Management Corporation, “State of Road Safety Report 2020-2021” Republic of South Africa: Department: Transport, 2021

[4] Boboye O., “FRSC Statistical Digest”, Federal Road Safety Corps National Headquarters, Abuja, 2021

[5] Online Dashboard, “Registro Nacional de Acidentes e Estatisticas de Trânsito”, [Online] [Cited: December 20, 2022] https://www.gov.br/infraestrutura/pt-br/assuntos/transito/arquivos-senatran/docs/renaest

[6] “State Inspectorate for Road Traffic Safety”, [Online] [Cited: December 18, 2021] http://www.gibdd.ru/

[7] Statens Vegvesen, [Online] [Cited: December 20, 2022] https://www.vegvesen.no/fag/fokusomrader/trafikksikkerhet/ulykkesdata/analyse-av-dod-sulykker-uag/, Norway

[8] Accident Research Centre, “Enhanced Crash Investigation Study – Report 1”, Monash University, 2020

[9] Schulze, O., “Investigação Avançada de Acidentes do Trânsito”, SAE Brazil, 2016

[10] Painter, M., “An introduction to Road Accident Sampling System – India”, JP Research, 2018

[11] China Automotive Technology and Research Center, “CATARC In-depth of Automotive Industries Development in China”, 2019

[12] Baldock, M.; Woolley, J.; Ponte, G.; Wundersitz, L.; Lindsay, V., „In-Depth Crash Investigation at the Centre for Automotive Safety Research”, University of Adelaide, South Australia, 2009

[13] The Research Association of Automotive Technology (FAT), Federal Highway Institute of Germany (BASt). “German In-Depth Accident Study (GIDAS)“. [Online] [Cited: December 08, 2022.] www.gidas.org.

[14] Statistisches Bundesamt (Destatis), „Fachserie 8 Reihe 7, Verkehr, Verkehrsunfälle, 2020,“ Statistisches Bundesamt (Destatis), 2021.

[15] European Commission (2022) Annual statistical report on road safety in the EU, 2021. European Road Safety Observatory. Brussels, European Commission, Directorate General for Transport

List of figures

List of tables

Manual "Global Safety Database"

The GSD is subdivided into main sections, shown in Figure 4-2 and in the table below.

HOME	• Main information of the GSD • Current number of data sources and countries • Current news and publications • Contact information • Partners involved in the GSD

META-DATABASE	The tab META-DATABASE is divided into four different subchapters:
	• Overview
	  ○ Overview of each data source
	  ○ Select bar for each data source
	• Filter
	  ○ Filter by Continent/Country
	  ○ Filter by Country Facts (Population, Accidents (all), Accidents (injury), Accidents (fatal)
	  ○ Filter by accidents by vehicle type
	• Export
	  ○ Export of data sources and associated records
	• Add Source (no GUEST rights)
	  ○ Input mask for new data source
	• Edit Country (no GUEST rights)
	  ○ Change mask for existing country

RESULT-MATRIX	The tab RESULT-MATRIX is divided into four subchapters:
	• Quick Search
	  ○ Overview of each question
	  ○ Select bar for each question
	• Filter
	  ○ Filter by Question
	  ○ Filter by Data Sources
	• Individual Question
	  ○ Input mask for new question with all required variables selectable
	• Development (just user level “Developer”)
	  ○ Detailed matrix information of each question

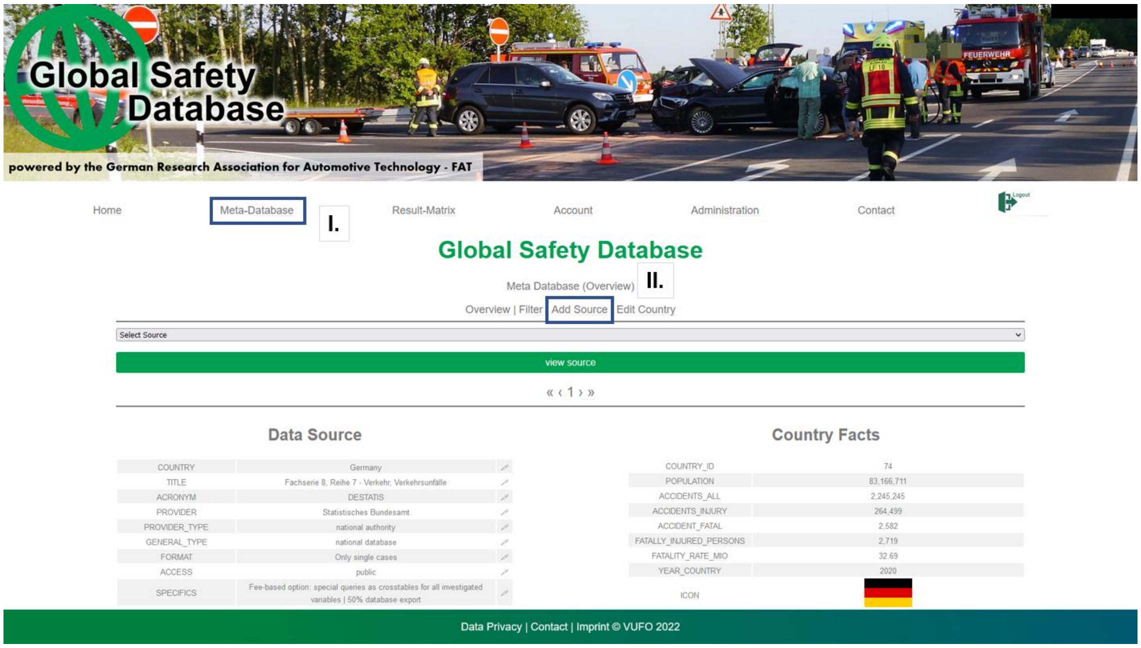
ACCOUNT	• Information about the user (name, e-mail address, questions entered into GSD) • Possibility to change password

ADMINISTRATION (only SC)	The tab ADMINISTRATION is divided into four subchapters:
	• User
	  ○ Overview of all users registered
	• Review
	  ○ Recommended changes to the variables
	  ○ Confirmation or Rejection of these changes
	• Change Log
	  ○ Overview of the changes in variables of each source
	• Individual Quests
	  ○ Overview of individual questions by user ID

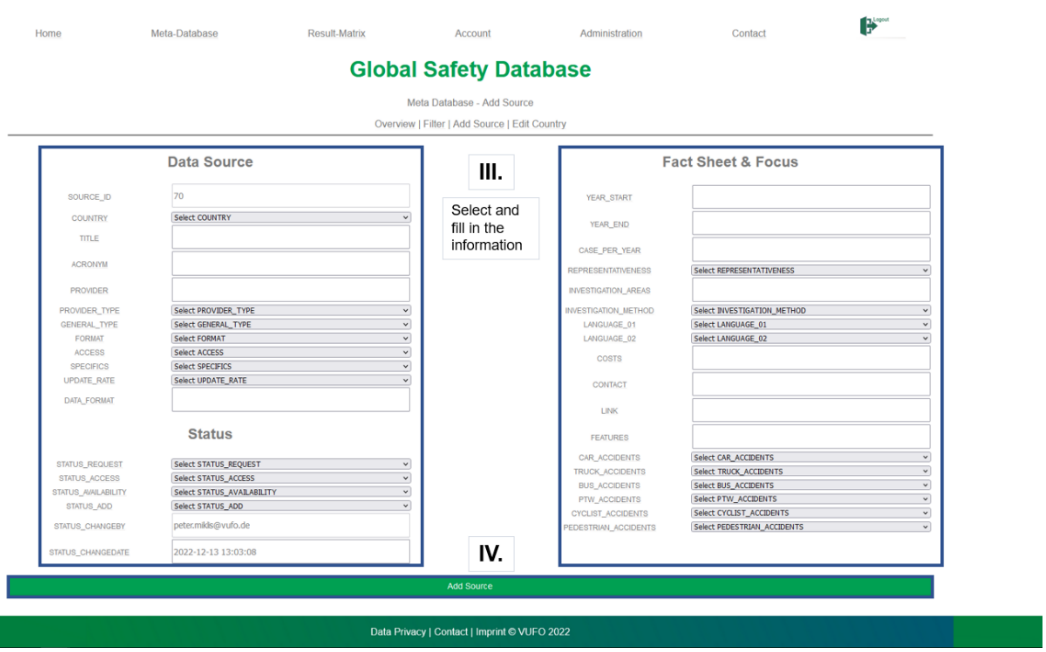
CONTACT	• Contact information

Creating a new data source

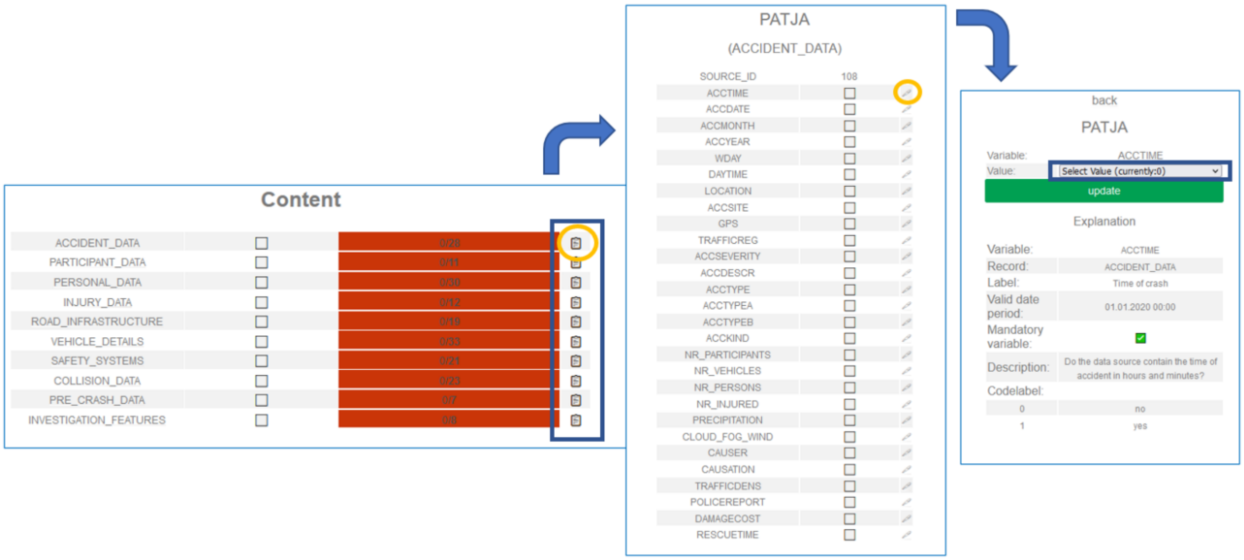
The first step in creating a new data source is to select the “Meta-Database” tab [I], followed by selecting the subchapter “Add Source” [II].



Then all necessary information must be entered, either by selecting it or manually writing it in [III]. The last step is clicking the button “Add Source” to submit the changes to the Steering Committee [IV]. All changes must be approved by the Steering Committee or the Head of Steering Committee.

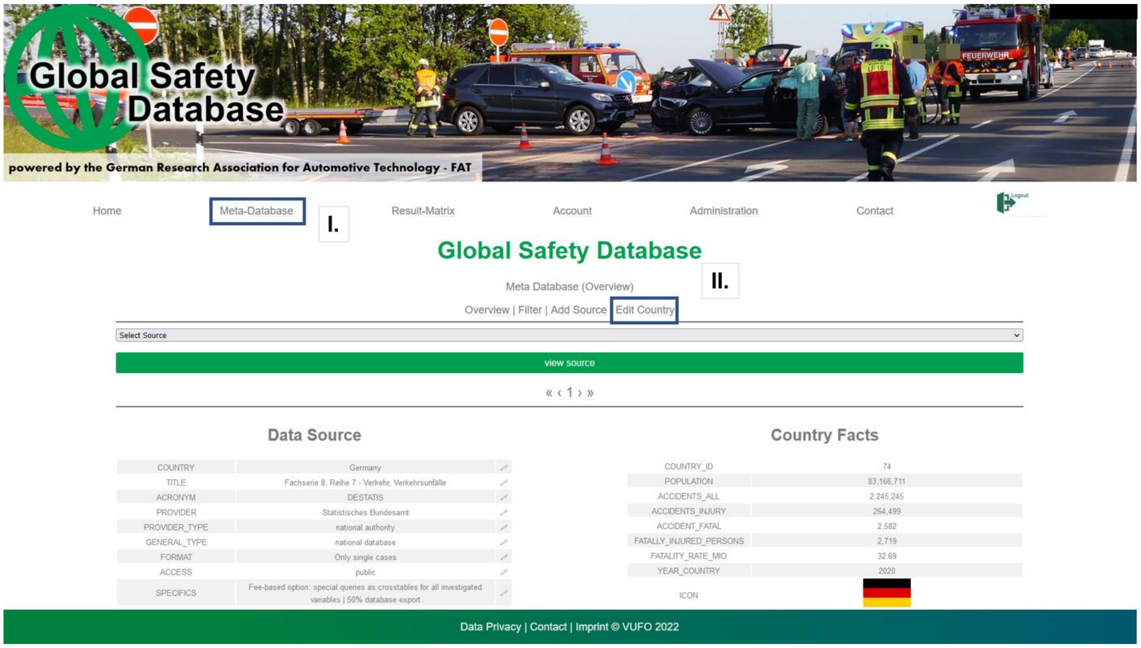


Once a new data source has been created, it can be found in the list of available data sources, after acceptance by the SC. In order for this to be evaluated subsequently, all available variables must then be entered individually in the “Content” and also confirmed by the SC.

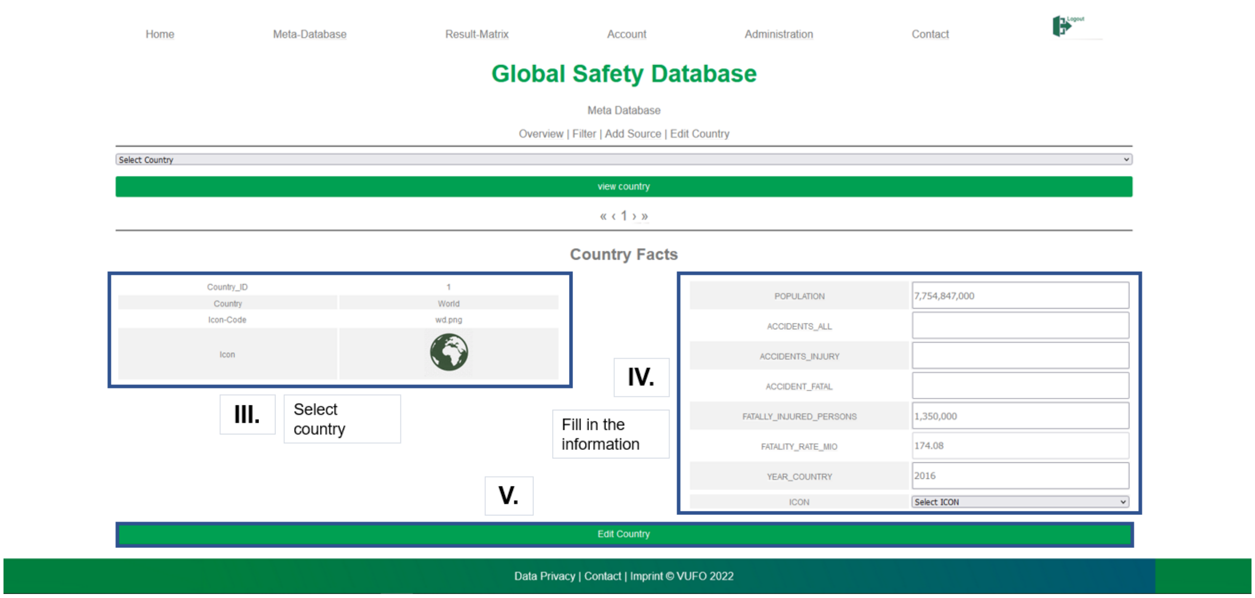


Editing country information

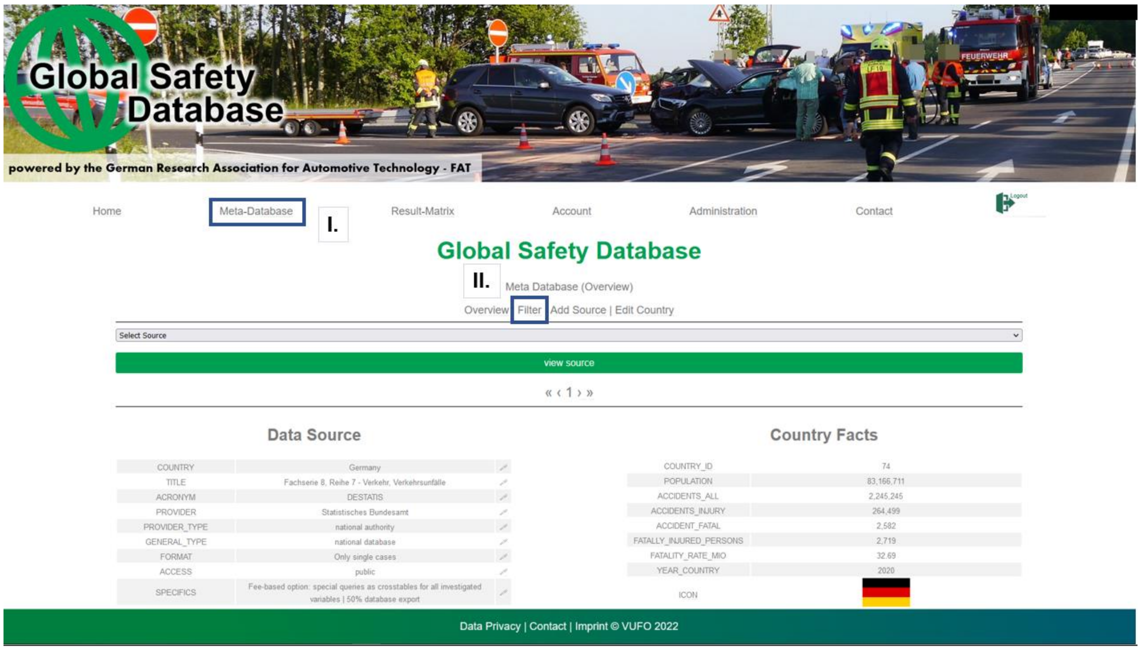
The first step in editing country information is to select the “Meta-Database” tab [I], followed by selecting the subchapter “Edit Country” [II].



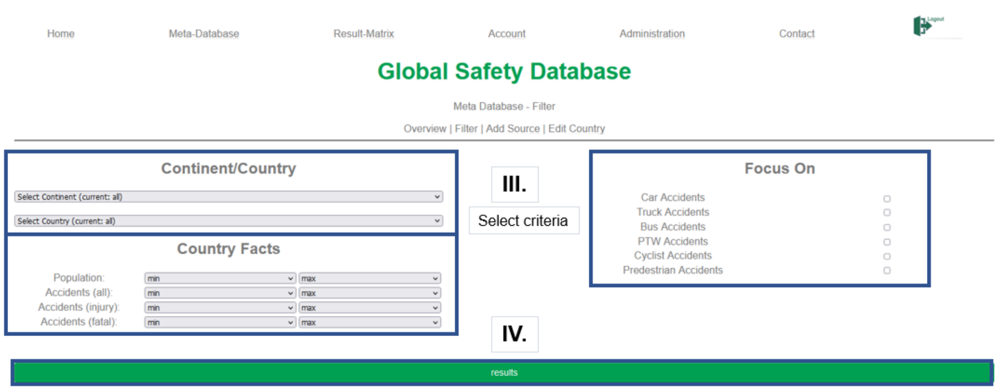
Then all necessary information must be entered [III]+[IV]. The last step is clicking the button “Edit Country” to submit the changes to the Steering Committee [V]. All changes must be approved by the Steering Committee or the Head of Steering Committee.



Searching for data source via filter option

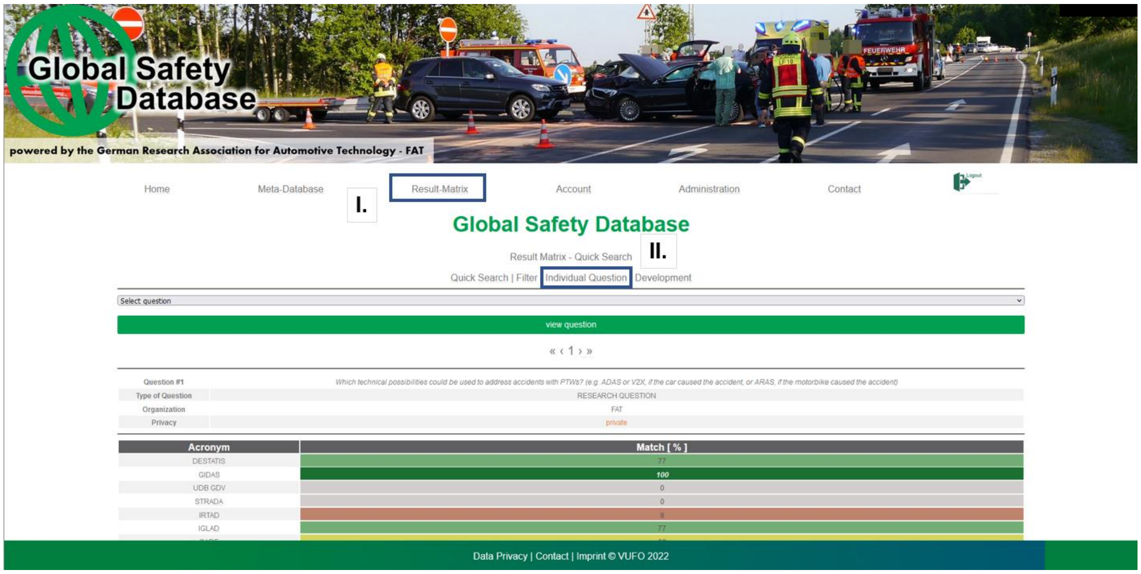


In the filter option, there are three main criteria according to which it is possible to filter (filter by Continent/Country, filter by Country Facts and filter by accidents by vehicle type). It is possible to filter by one criterion or by a combination of several criteria [III]. After setting the desired filters, the "results" button can be used to display the results [IV].

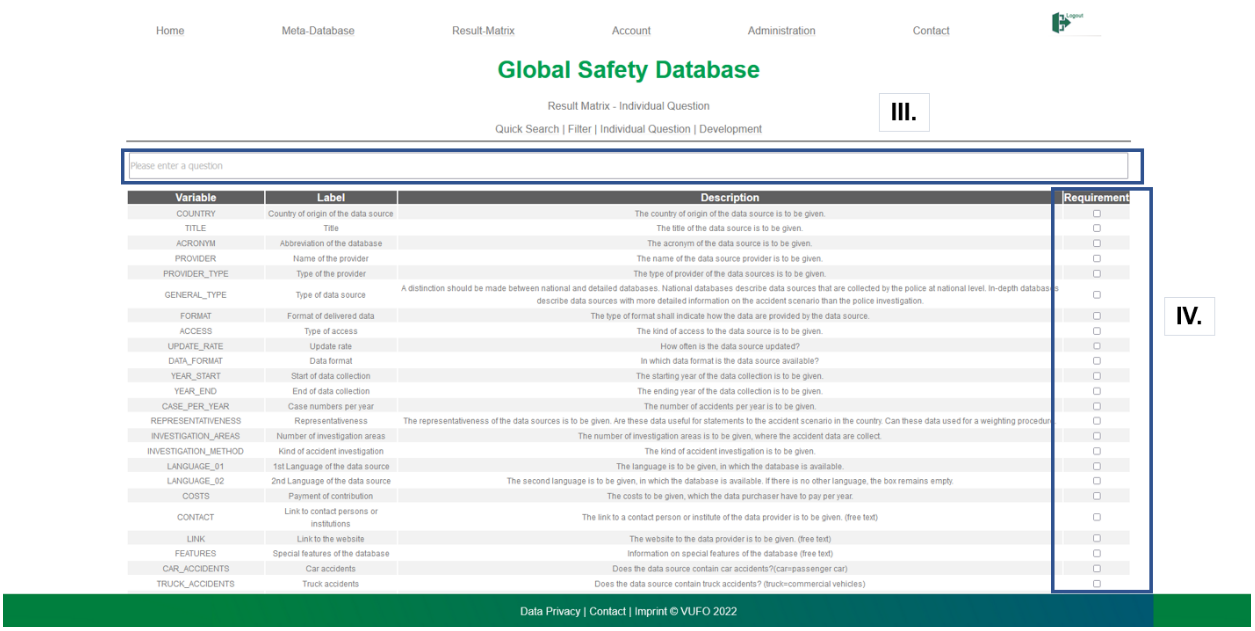


Adding Individual Question

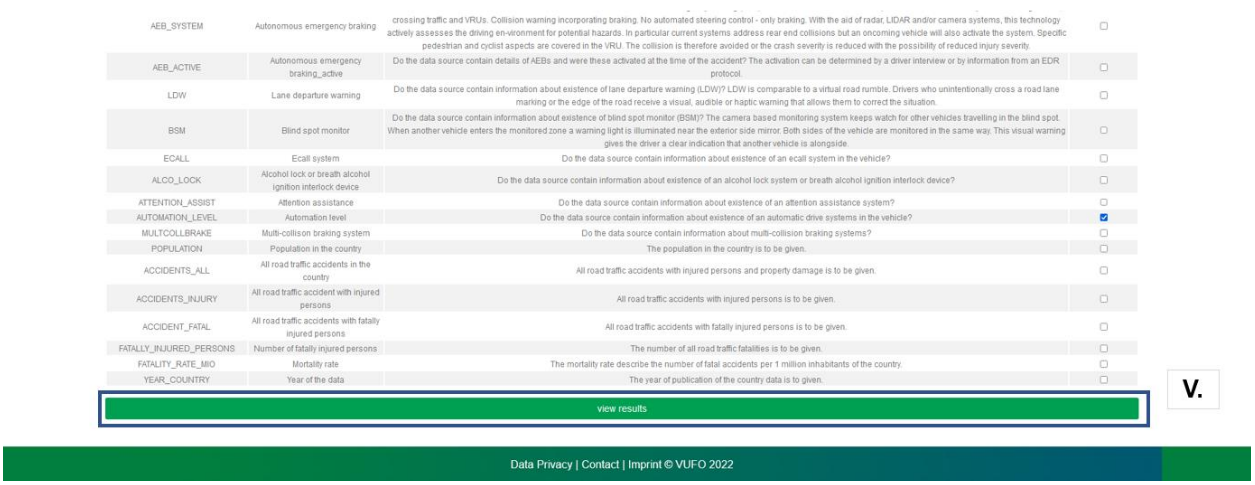
The creation of an Individual Question is done via the tab “Result-Matrix” [I] and the subchapter “Individual Question” [II].



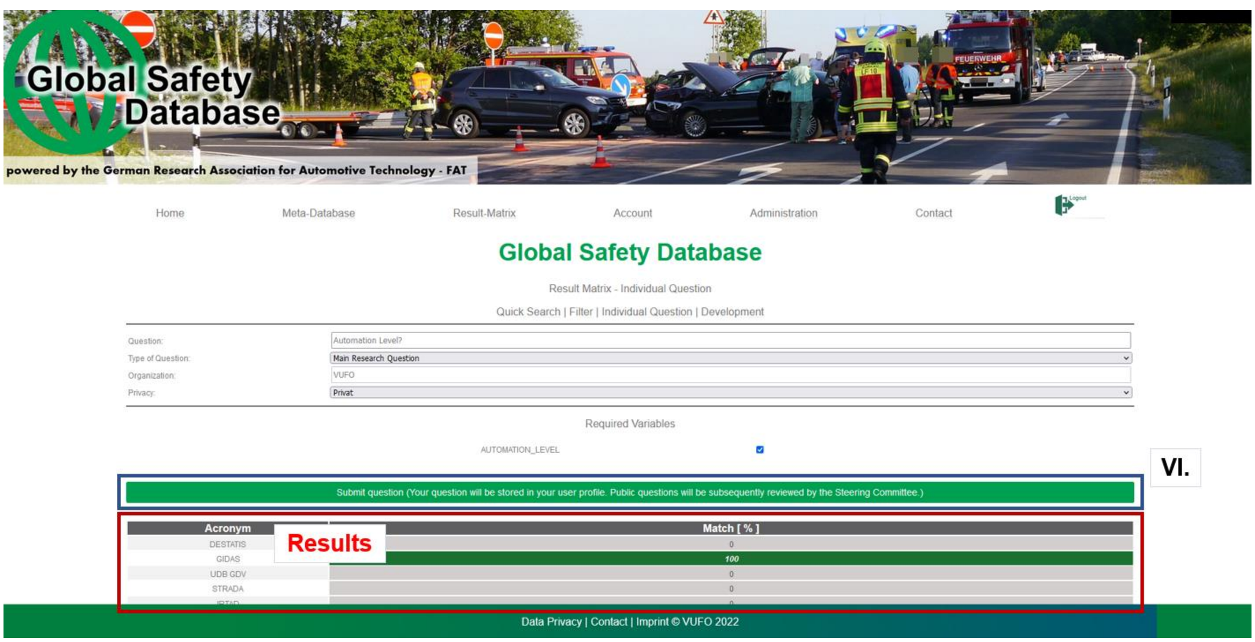
Then enter the question [III] and select all the variables required for the desired answer [IV].



To get the results of which database can answer the question, click the “view results” button [V].



All data sources and the corresponding hit rate are then displayed. Finally, there is still the possibility to submit this individual question to the SC [VI]. For this purpose, the type of question and the privacy settings must be adjusted as desired. After successful verification by the SC, this question is added to the existing question catalogue.

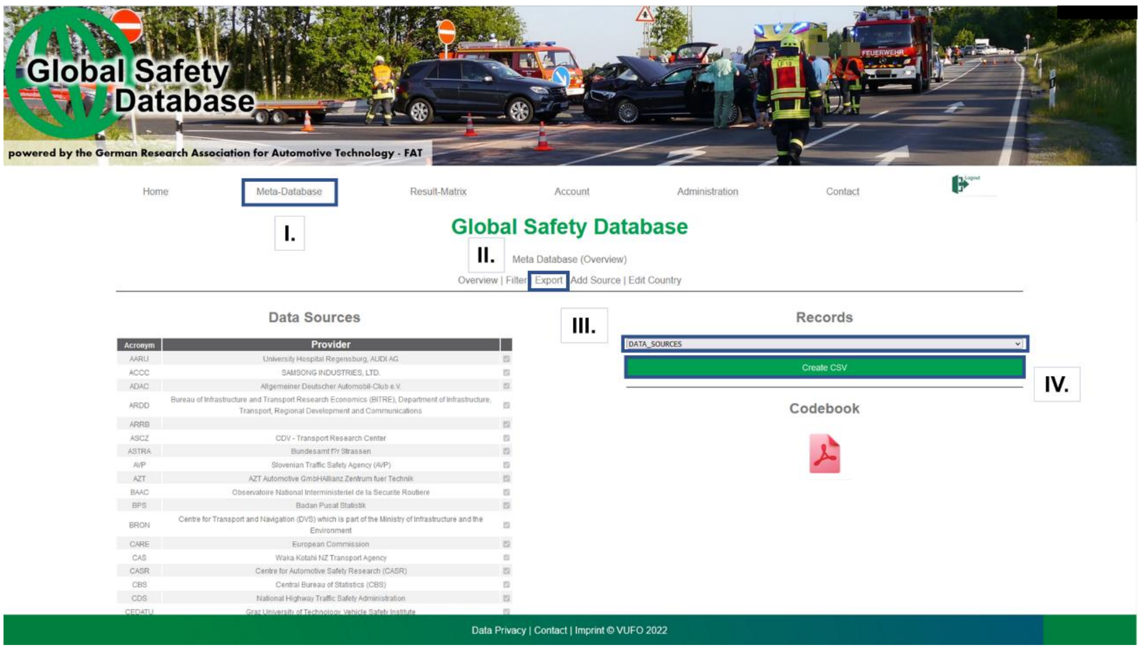


Searching for Research Questions

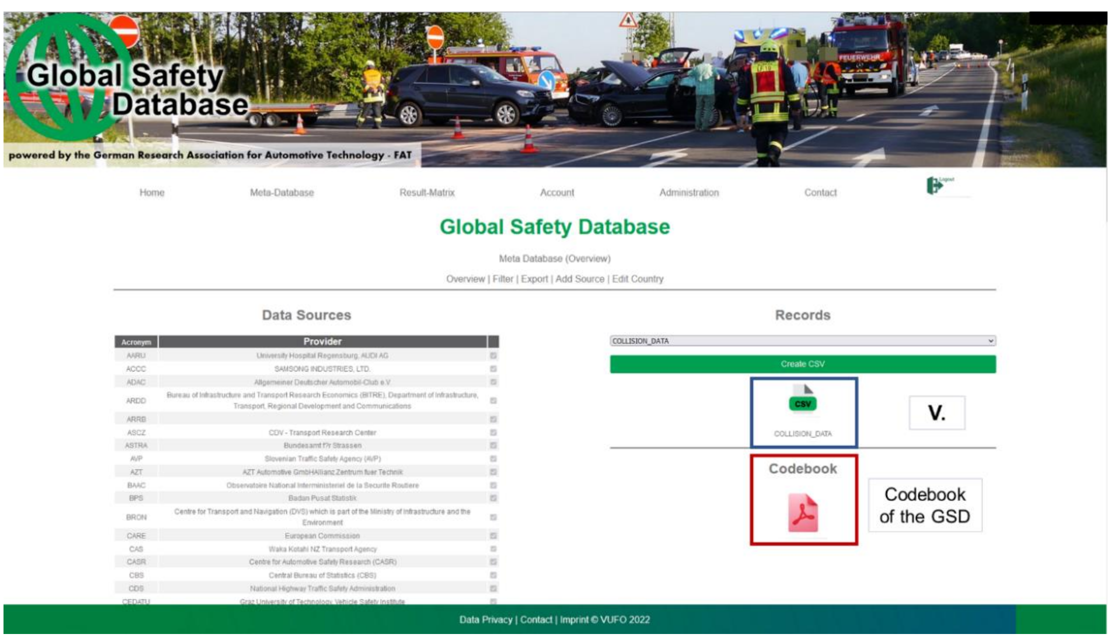
To search for questions already implemented in the GSD using various criteria, select "Result-Matrix" [I] and the "Filter" subchapter [II]. Then select the desired criteria [III] and click the "results" button to display the results [IV].

Export

To export the desired data sources click on the “Meta-Database” tab [I] and then on the “Export” subchapter [II]. If there are no further restrictions, all available data sources are selected first. If only certain data sources are to be considered, the data sources can be selected or deselected individually. When all the desired data sources have been selected, it is still necessary to select the records to be included [III]. To prepare the export file, click on the "Create CSV" button [IV].



After clicking the “Create CSV” button the file is prepared and can be downloaded [V]. Since the export file outputs the associated codes for each record, the GSD codebook can additionally be downloaded for assistance.



Bisher in der FAT-Schriftenreihe erschienen (ab 2019)

Nr.	Titel
316	MULTIC-Tooling, 2019
317	EPHoS: Evaluation of Programming - Models for Heterogeneous Systems, 2019
318	Air Quality Modelling on the Contribution of Brake Wear Emissions to Particulate Matter Concentrations Using a High-Resolution Brake Use Inventory, 2019
319	Dehnratenabhängiges Verformungs- und Versagensverhalten von dünnen Blechen unter Scherbelastung, 2019
320	Bionischer LAM-Stahlleichtbau für den Automobilbau – BioLAS, 2019
321	Wirkung von Systemen der aktiven, passiven und integralen Sicherheit bei Straßenverkehrsunfällen mit schweren Güterkraftfahrzeugen, 2019
322	Unfallvermeidung durch Reibwertprognosen - Umsetzung und Anwendung, 2019
323	Transitionen bei Level-3-Automation: Einfluss der Verkehrsumgebung auf die Bewältigungsleistung des Fahrers während Realfahrten, 2019
324	Methodische Aspekte und aktuelle inhaltliche Schwerpunkte bei der Konzeption experimenteller Studien zum hochautomatisierten Fahren, 2020
325	Der Einfluss von Wärmeverlusten auf den Rollwiderstand von Reifen, 2020
326	Lebensdauerberechnung hybrider Verbindungen, 2020
327	Entwicklung der Verletzungsschwere bei Verkehrsunfällen in Deutschland im Kontext verschiedener AIS-Revisionen, 2020
328	Entwicklung einer Methodik zur Korrektur von EES-Werten, 2020
329	Untersuchung zu den Einsatzmöglichkeiten der Graphen- und Heuristikbasierten Topologieoptimierung zur Entwicklung von 3D-Rahmenstrukturen in Crashlastfällen, 2020
330	Analyse der Einflussfaktoren auf die Abweichung zwischen CFD und Fahrversuch bei der Bestimmung des Luftwiderstands von Nutzfahrzeugen, 2020
331	Effiziente Charakterisierung und Modellierung des anisotropen Versagensverhaltens von LFT für Crashsimulation, 2020
332	Charakterisierung und Modellierung des Versagensverhaltens von Komponenten aus duktilem Gusseisen für die Crashsimulation, 2020
333	Charakterisierung und Meta-Modellierung von ungleichartigen Punktschweißverbindungen für die Crashsimulation, 2020
334	Simulationsgestützte Analyse und Bewertung der Fehlertoleranz von Kfz-Bordnetzen, 2020
335	Absicherung des autonomen Fahrens gegen EMV-bedingte Fehlfunktion, 2020
336	Auswirkung von instationären Anströmeffekten auf die Fahrzeugaerodynamik, 2020
337	Analyse von neuen Zell-Technologien und deren Auswirkungen auf das Gesamtsystem Batteriepack, 2020
338	Modellierung der Einflüsse von Mikrodefekten auf das Versagensverhalten von Al-Druckguss-komponenten mit stochastischem Aspekt für die Crashsimulation, 2020

339 Stochastisches Bruchverhalten von Glas, 2020

340 Schnelle, breitbandige Datenübertragung zwischen Truck und Trailer als Voraussetzung für das hochautomatisierte Fahren von Lastzügen, 2021

341 Wasserstoffkompatibilität von Aluminium-Legierungen für Brennstoffzellenfahrzeuge, 2021

342 Anforderungen an eine elektrische Lade- und Wasserstoffinfrastruktur für gewerbliche Nutzfahrzeuge mit dem Zeithorizont 2030, 2021

343 Objective assessment of database quality for use in the automotive research and development process, 2021

344 Review of non-exhaust particle emissions from road vehicles, 2021

345 Ganzheitliche Betrachtung von Rollwiderstandsverlusten an einem schweren Sattelzug unter realen Umgebungsbedingungen, 2021

346 Studie zur Abschätzung der Anwendungspotentiale, Risiken und notwendigen Forschungsbedarfe bei der Verwendung von Glashohlkugeln in Kombination mit thermoplastischem Schaumspritzguss, 2021

347 Typgenehmigungsanforderungen an Level-3-Autobahnsysteme - Hintergrundbetrachtungen zu technischen Anforderungen für eine automatisierte Fahrfunktion, 2021

348 Einfluss der Kantenbearbeitung von Aluminiumblechen auf das Restumformvermögen sowie die Festigkeitseigenschaften unter quasistatischer und schwingender Beanspruchung, 2021

349 Verstärkung dünner formgehärteter Bauteile mittels FVK-Verrippungen, 2021

350 HMI Anforderungen für den automatisierten Individualverkehr unter Berücksichtigung von Leistungsmöglichkeiten und -grenzen älterer Nutzer, 2021

351 Compatibility of polymers for fuel cell automobiles, 2021

352 Entwicklung einer gewichtsoptimierten Batteriegehäusestruktur für Volumenfahrzeuge, 2021

353 Charakterisierung und Modellierung des Deformations- und Versagensverhaltens von nicht-faserverstärkten Thermoplasten unter mehrachsiger Crashbelastung, 2021

354 Untersuchung zum thermischen Komfort im Pkw für den Grenzbereich des Luftzugempfindens, 2021

355 Anforderungen an die Güte, Verfügbarkeit und Vorausschau einer Reibwertschätzung aus Funktionssicht, 2021

356 Entwicklung einer standardisierten Prüfanordnung zur Bewertung der Übernahmeleistung beim automatisierten Fahren, 2022

357 Vorstudie zu Verkehrsemissionen - Räumlich und zeitlich aufgelöste Daten durch Schwarmmessungen, 2022

358 Produktivitätssteigerung und Kostensenkung der laser-additiven Fertigung für den Automobilbau, 2022

359 Analyse der Einflussfaktoren auf die Abweichung zwischen CFD und Fahrversuch bei der Bestimmung des Luftwiderstands von Nutzfahrzeugen mit Fokus auf den Ventilationswiderstand von Nfz-Rädern, 2022

360 Werkstoffmodelle und Kennwertermittlung für die industrielle Anwendung der Umform- und Crash-Simulation unter Berücksichtigung der thermischen Behandlungen beim Lackieren im Prozess bei hochfesten Werkstoffen, 2022

361 Compatibility of polymers for fuel cell automobiles, 2022

362 Ermüdung kurzfaserverstärkter thermoplastischer Polymerkwerkstoffe, 2022

363 Market research and definition of procedure to comparison of comfort measuring systems for a vehicle cabin, 2022

364 Methodische Ansätze zur Auswahl von Bordnetzstrukturen mit erhöhten Zuverlässigkeitsanforderungen, 2022

365 Fahrwiderstand von Lenk- und Liftachsen in Kurven und auf gerader Strecke unter realen Umgebungsbedingungen, 2022

366 Klimadaten und Nutzungsverhalten zu Auslegung, Versuch und Simulation an Kraftfahrzeug-Kälte-/Heizanlagen, 2022

367 Experimentelle und numerische Untersuchung des selbsttätigen Losdrehens von Schraubenverbindungen mit konstanten und variablen Amplituden und Entwicklung einer Bewertungsmethode, 2022

368 Objective assessment of database quality for use in the automotive research and development process – Part 2, 2023

Impressum

Herausgeber	FAT Forschungsvereinigung Automobiltechnik e.V. Behrenstraße 35 10117 Berlin Telefon +49 30 897842-0 Fax +49 30 897842-600 www.vda-fat.de
ISSN	2192-7863

Verband der Automobilindustrie e.V. (VDA)
Behrenstraße 35, 10117 Berlin
www.vda.de
Twitter @VDA_online

Forschungsvereinigung Automobiltechnik e.V. (FAT)
Behrenstraße 35, 10117 Berlin
www.vda.de/fat

Verband der
Automobilindustrie

Forschungsvereinigung
Automobiltechnik