

AUTHOR



Dipl.-Ing. Johann Ziegler is Research Associate of the Traffic Accident Research at TU Dresden GmbH (Vufo) (Germany).



Dipl.-Ing. Henrik Liers is Managing Director of the Traffic Accident Research at TU Dresden GmbH (Vufo) (Germany).



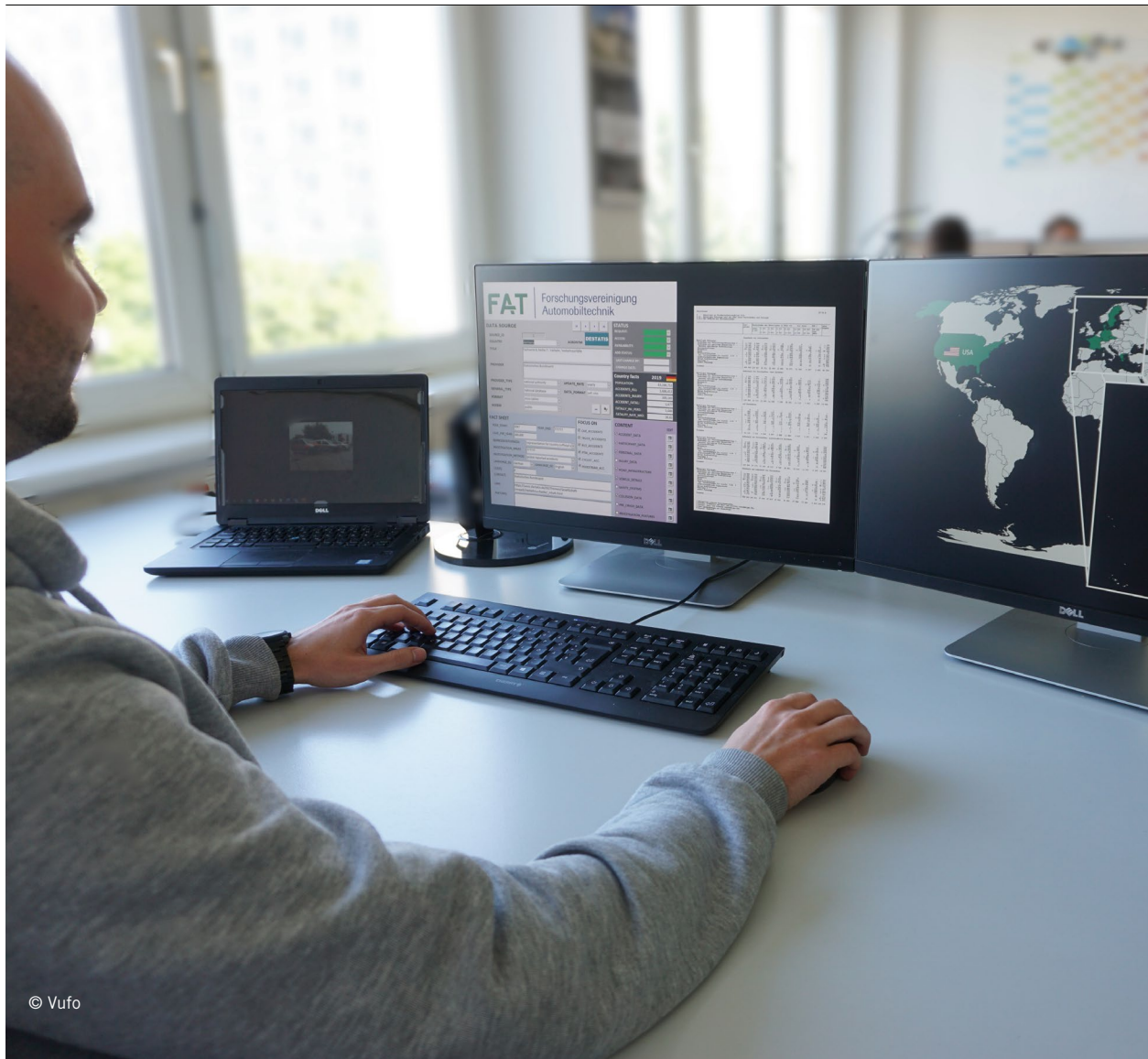
Dipl.-Ing. Albine Chanove is Research Associate at the Fraunhofer Institute for Transportation and Infrastructure Systems (IVI) in Dresden (Germany).



Dr.-Ing. Ludwig Drees is Specialist for Integral Safety at BMW AG in Munich and Chair of the FAT working group 3 “Biomechanics and Accident Research” (Germany).

Systematic Approach to the Objective Evaluation of Traffic Accident Databases

Assisted, automated and connected driving requires data sources in the development as well as safeguarding process. In the FAT project “Objective assessment of the quality of databases for the use of these in the research and vehicle development process” (FAT publication series No. 343), the Traffic Accident Research at TU Dresden GmbH (Vufo) together with the Fraunhofer Institute for Transportation and Infrastructure Systems (IVI) is developing a meta database that can pull information from current and future databases as well as from databases adjusted to research questions.



© Vufo

1	MOTIVATION
2	CONCEPTION OF THE META DATABASE
3	SEARCH FOR DATA SOURCES
4	METHODOLOGY OF THE OBJECTIVE EVALUATION
5	EXAMPLE APPLICATIONS
6	SUMMARY AND OUTLOOK

1 MOTIVATION

Extensive road traffic data sources are essential to consider safety aspects in vehicle development. Numerous data sources exist worldwide around road traffic accidents that are used for the development, validation, and field observation of new safety systems. 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 objective inventory and evaluation of these data collections. Therefore, in addition to a content inventory, the usability for research and development activities, the level of detail contained, and methodological and organizational characteristics are also of interest. The comparison of the available content with a specific questionnaire enables an objective evaluation of data sources. A meta database that can be expanded and adapted for future content makes it easier for researchers to identify suitable databases for future research questions as well.

2 CONCEPTION OF THE META DATABASE

A project was initiated by Working Group 3 of the German Research Association for Automotive Technology (FAT) to create a meta database [1] based on the database structure of the interdisciplinary German In-Depth Accident Study (Gidas) [2], one of the world's leading accident research projects. The meta database contains, on the one hand, high-level information about the researched data sources and on the other hand a content-related part in which the available information is inventoried. Each part comprises several tables that are linked by primary keys. The meta database not only contains information on the parameters and contents available in a data source, but also meta data that directly determine the suitability of the data source for certain research questions or the reliability of statements derived from it. For the usability and evaluability of data sources, their metadata are elementary for determining representativeness.

Further parameters are dedicated to the methods of data collection and the possibilities of accessing them. For the derivation of exposure variables or basic figures, country-specific key figures are stored, for example, the number of traffic accidents per year as well as information on demographics, vehicle fleet, and infrastructure. The storage of content-related aspects takes place in thematically sorted tables at accident, participant, person, and injury level. The developed meta database in the current version

contains 15 tables with 237 variables. All entities of the database are described in a codebook written in English.

3 SEARCH FOR DATA SOURCES

Both national sources and in-depth accident data sources are researched. The former are based on data from police accident investigations, they usually contain very high accident figures and offer users the possibility of a macroscopic view on the country's accident situation. Except for the dark field of unreported accidents, they represent a complete survey of accidents involving personal injury and are considered to be representative of the respective country. The main focus of police recording is the collection of accident-related data for the purpose of preserving evidence and observing the general accident situation. In-depth data collections, on the other hand, pursue a different, mostly interdisciplinary and scientifically motivated approach. Compared to national data sources, they are characterized by a significantly lower number of cases but a substantially higher level of detail. This enables microscopic analyses of the accident situation. In the current project, in addition to the draft and implementation of the meta database, the project also conducts detailed research for the countries listed in **TABLE 1**. A distinction is made between fully researched data sources and those whose existence is known but for which no further information has yet been extracted.

4 METHODOLOGY OF THE OBJECTIVE EVALUATION

A central concern of the project is the objective assessment of the quality of the numerous existing data sources. To evaluate the metrics, the suitability of the data sources for answering current and future research questions is examined – also with a view to a practical application of the meta database. For its compilation, member companies of the FAT were asked to collect questions relevant to them from the field of design, development, and evaluation of road safety measures and to make them available to the research participants. In this way, an interdisciplinary catalogue of questions was created with more than 190 questions, of which 120 are content-based research questions and the rest are questions about the characteristics and metadata of the researched data sources.

Country	Number of data sources		Number of fully researched data sources	
	National	In-depth	National	In-depth
Denmark	2	-	2	-
Germany	3	9	2	1
France	1	1	1	-
Greece	1	1	1	1
Sweden	1	2	-	1
Czechia	1	1	1	1
USA	4	3	4	2
Other	17	15	-	-
Total	30	32	11	6

TABLE 1 Number and type of data sources researched by country (© Vufo)

The research questions are analyzed semantically to create a mapping of the content associated with the question to the parameters of the meta database. The requirements for the data source contents are thus defined for each question. For simplified handling, these requirements for the existence of corresponding content are stored in binary codes at the parameter level. Subsequently, an automated matching process compares the requirements for the questions with the existence of data source content and stores the result in a result matrix. Finally, the content of this matrix indicates for each data source the proportion of variables that correspond to the variables necessary to answer the research question. **FIGURE 1** shows in a cross comparison of five selected data sources how many of the 120 research questions can be answered completely or to what percentage. With the data source from “EU Country I”, for example, 115 of the questions included in the catalogue can be answered completely.

It should also be noted that no assessment is made of the completeness and plausibility of the actual (accident) data. This quality check as well as the decision whether a data source is used for an analysis even with a lower percentage of answered questions and thus missing information is up to the user.

5 EXAMPLE APPLICATIONS

In the development phase of new safety systems, consumer protection protocols and legislative regulations, retrospective and increasingly prospective analyses of road accident data sources are standard tools. This often involves cross-comparisons between the measures and situations of different countries and regions, which provide a global view of the issue in question and, where appropriate, also help to identify measures that have already been tried and tested. In addition, most companies in the automotive

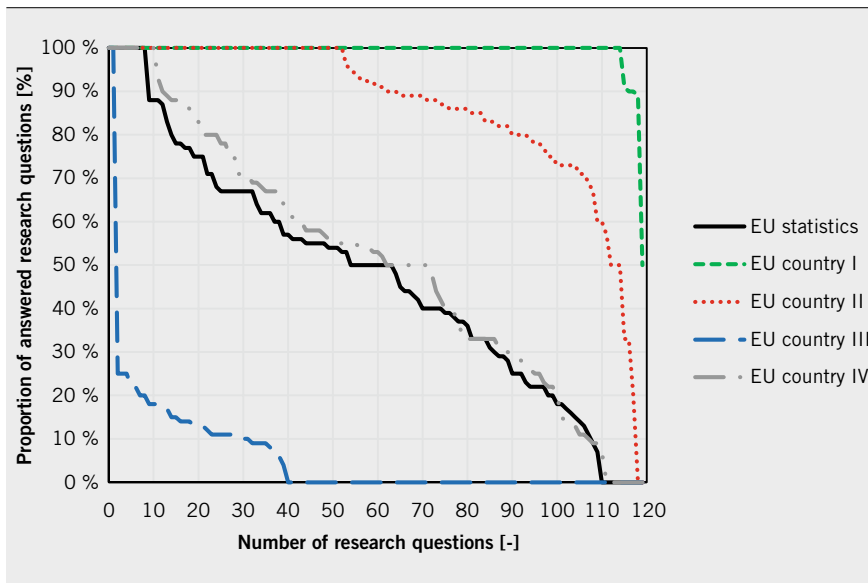


FIGURE 1 Visualization of the usability of five selected data sources from one EU country at a time for answering the research questions stored in the meta database (© Vufo)

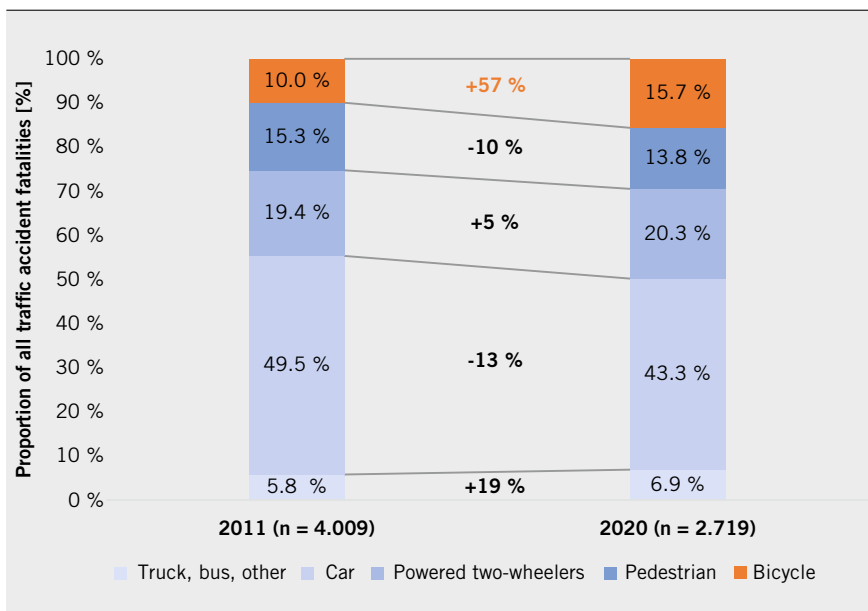


FIGURE 2 Distribution of the types of road users in all road accident fatalities within Germany (2011 and 2020) (© Vufo)

Country/region	Number of data sources	Type of data source	Accidents with cyclists	Helmet use	Impact direction cyclist/vehicle	Single injuries
Worldwide	1	In-depth	Yes	Yes	Yes	No
EU statistics	1	National	Yes	Yes	Yes	No
EU country I	1	In-depth	Yes	Yes	Yes	Yes
EU country II	1	In-depth	Yes	Yes	Yes	Yes
EU country III	1	In-depth	Yes	Yes	No	No
EU country IV	1	National	Yes	Yes	Yes	No
EU country V	1	National	Yes	Yes	No	No
North America	4	National and in-depth	Yes	Yes	3 Yes 1 No	1 Yes 3 No

TABLE 2 Results for selected parameters on cyclist safety in different data sources (© Vufo)

industry operate in many countries and on several continents, which requires an analysis of suitable data sources from different countries or economic regions.

A specific example will be used to briefly outline a selected application possibility. With a view to the German accident situation and the accelerated increase in cycling due to the Covid-19 pandemic, the focus will be on this type of road users. **FIGURE 2** illustrates the increasing proportion of cyclists killed in the last decade, based on figures from the German Federal Statistical Office (Destatis) [3]. Interesting aspects for analyses on cyclist safety are, in particular, the impact constellation with regard to injury causation and the use of bicycle helmets with regard to potential protective measures. With the help of simple queries, all data sources with the relevant information can be efficiently listed in the meta database. In the example in **TABLE 2**, it can be seen that eleven of the data sources included provide information on accidents involving cyclists and the aspect of helmet use. Five data sources remain for further analyses of the impact constellation, while more in-depth investigations based on individual injuries are only possible with three data sources.

Since access to individual or raw data sets is often limited and data use is associated with significant costs, users must make an appropriate data source selection for their use cases. The results generated by the matching process provide users with a basis for objective evaluation of data sources. The result matrix thereby represents the suitability of individual data sources for all or concrete, user-specific topics.

For the implementation of future questions, the meta database contains a form where users can implement their question(s) and identify the parameters necessary to answer them. After performing the script-based matching process, the result matrix is generated, or rather an existing one updated.

6 SUMMARY AND OUTLOOK

The project aims to research, inventory and objectively evaluate diverse accident data sources from different countries. For this purpose, a meta database is designed and filled with selected data sources. The objective evaluation is based on a questionnaire with current research questions and a matching process, which matches the content of the questions with the availability of variables per data source. The result is a matrix that shows the availability of necessary variables per data source for each research question.

The current question catalogue and the evaluation of the suitability of the researched data sources is mainly based on research questions from research and development departments of automotive companies associated in the FAT. For a broader evaluation, the questionnaire could be expanded to include questions from legislative bodies, authorities, associations, universities, or consumer protection organizations. Useful additions could be made, for example with weather or traffic flow data, as well as research of data sources with observations of driving behavior. The meta database is a useful tool for making the increasingly data-driven vehicle development process more efficient by greatly reducing the research effort required to find suitable data sources to answer the relevant development questions.

REFERENCES

- [1] 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
- [2] Gidas: German In-Depth Accident Study (Gidas): Online: www.gidas.org, access: August, 10, 2021
- [3] Destatis: Verkehr. Verkehrsunfälle. Fachserie 8, Reihe 7, Federal Statistical Office, Wiesbaden, 2021

THANKS

The research project “Objective assessment of the quality of databases for the use of these in the research and vehicle development process” is self-funded by the German Research Association for Automotive Technology (FAT) and carried out at Traffic Accident Research at TU Dresden GmbH (Vufo) and the Fraunhofer Institute for Transportation and Infrastructure Systems (IVI) and was initiated by the FAT Working Group 3 “Biomechanics and Accident Research” under the leadership of Dr.-Ing. Michael Düring (Volkswagen AG) and Dipl.-Ing. (FH) Michael Wagner (Continental AG). The authors would like to thank FAT for the funding and all project partners involved.