

AUTORIN
UND AUTOREN



Dipl.-Ing. Johann Ziegler
ist Wissenschaftlicher
Mitarbeiter der Verkehrs-
unfallforschung an der
TU Dresden GmbH (Vufo).



Dipl.-Ing. Henrik Liers
ist Geschäftsführer der
Verkehrsunfallforschung
an der TU Dresden
GmbH (Vufo).



**Dipl.-Ing.
Albine Chanove**
ist Wissenschaftliche
Mitarbeiterin am Fraun-
hofer-Institut für Verkehrs-
und Infrastruktursysteme
(IVI) in Dresden.



Dr.-Ing. Ludwig Drees
ist Leiter des FAT-Arbeits-
kreises 3 (Biomechanik)
und Unfallforschung und
Spezialist für integrale
Sicherheit bei der
BMW AG in München.

Systematischer Ansatz zur objektiven Bewertung von Verkehrsunfalldatenbanken

Assistiertes, automatisiertes und vernetztes Fahren erfordert im Entwicklungs- sowie Absicherungsprozess Datenquellen. Im FAT-Vorhaben „Objektive Einschätzung der Qualität von Datenbanken zur Verwendung dieser im Forschungs- und Fahrzeugentwicklungsprozess“ (FAT-Schriftenreihe 343) entwickelt die Verkehrsunfallforschung an der TU Dresden GmbH (Vufo) zusammen mit dem Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme (IVI) eine Metadatenbank, die Informationen aus aktuellen, zukünftigen und auf Forschungsfragen abgestimmten Datenbanken ziehen kann.



© Vufo

1	MOTIVATION
2	KONZEPTION DER METADATENBANK
3	RECHERCHE NACH DATENQUELLEN
4	METHODIK DER OBJEKTIVEN BEWERTUNG
5	BEISPIELANWENDUNGEN
5	ZUSAMMENFASSUNG UND AUSBLICK

1 MOTIVATION

Zur Berücksichtigung von Sicherheitsaspekten bei der Fahrzeugentwicklung sind umfangreiche Verkehrsunfalldatenquellen unabdingbar. Weltweit existieren zahlreiche Quellen rund um Straßenverkehrsunfälle, die für die Entwicklung, Absicherung und Feldbeobachtung neuer Sicherheitssysteme herangezogen werden. Es ist jedoch nicht immer offensichtlich, welche Datenquelle für welche Art von Forschungsfrage oder Entwicklungsansatz geeignet ist, da es derzeit keine objektive Inventarisierung und Bewertung dieser Datensammlungen gibt. Neben einer inhaltlichen Bestandsaufnahme sind daher auch die Nutzbarkeit für Forschungs- und Entwicklungsaktivitäten, der enthaltene Detaillierungsgrad sowie methodische und organisatorische Charakteristika von Bedeutung. Der Abgleich der verfügbaren Inhalte mit einem spezifischen Fragenkatalog ermöglicht eine objektive Bewertung der Datenquellen. Eine Metadatenbank, die sich erweitern und anpassen lässt für zukünftige Inhalte, erleichtert Forschern und Forscherinnen, geeignete Datenbanken auch für zukünftige Forschungsfragen zu identifizieren.

2 KONZEPTION DER METADATENBANK

Vom Arbeitskreis 3 der Forschungsvereinigung Automobiltechnik (FAT e.V.) wurde ein Projekt initiiert, in dem eine Metadatenbank entsteht [1], die auf der Datenbankstruktur der interdisziplinären German In-Depth Accident Study (Gidas) [2] beruht, einem der weltweit führenden Unfallforschungsprojekte. Die Metadatenbank enthält zum einen übergeordnete Informationen zu den recherchierten Datenquellen und zum anderen einen inhaltsbezogenen Teil, in dem der verfügbare Umfang inventarisiert wird. Jeder Teil umfasst mehrere Tabellen, die durch Primärschlüssel miteinander verknüpft sind. Die Metadatenbank stellt also nicht nur Informationen zu den in einer Datenquelle vorhandenen Parametern und Inhalten zur Verfügung, sondern auch Metadaten, die direkt die Eignung der Datenquelle für bestimmte Forschungsfragen beziehungsweise die Reliabilität daraus abgeleiteter Aussagen bestimmen. Für die Nutzbarkeit und Bewertbarkeit von Datenquellen sind deren Metadaten zur Feststellung der Repräsentativität elementar.

Weitere Parameter widmen sich den Methoden der Erhebung von Daten und den Möglichkeiten, auf sie zuzugreifen. Für die Ausleitung von Expositionsgrößen oder Basiszahlen sind länderspezifische Kennzahlen hinterlegt, beispielsweise die Anzahlen von Verkehrsunfällen pro Jahr sowie Angaben zu Demografie, Fahrzeugflotte und Infrastruktur. Die Ablage inhaltlicher Aspekte erfolgt in thematisch sortierten Tabellen auf Unfall-, Beteiligten-, Personen- und Verletzungsebene. Die entwickelte Metadatenbank in der aktuellen Version enthält 15 Tabellen mit 237 Variablen. Alle Entitäten der Datenbank sind in einem englischsprachigen Codebuch beschrieben.

3 RECHERCHE NACH DATENQUELLEN

Es werden sowohl nationale Quellen als auch tieferegehende Unfall-datenquellen (In-Depth) recherchiert. Erstere basieren auf den Daten der polizeilichen Unfallerehebungen, enthalten meist sehr hohe Unfallzahlen und bieten Anwendern die Möglichkeit der makroskopischen Sicht auf das Unfallgeschehen. Für das Unfallgeschehen mit Personenschaden stellen sie mit Ausnahme des Dunkelfelds nicht gemeldeter Unfälle eine Vollerhebung dar und gelten als repräsentativ für das jeweilige Land. Schwerpunkte der polizeilichen Erfassung sind die Erhebung unfallbezogener Daten zur Beweissicherung und Beobachtung des allgemeinen Unfallgeschehens. In-Depth-Datenerhebungen verfolgen dagegen einen anderen, meist interdisziplinär geprägten und wissenschaftlich motivierten Ansatz. Sie zeichnen sich im Vergleich zu nationalen Datenquellen durch eine deutlich geringere Fallanzahl, aber einen signifikant höheren Detaillierungsgrad aus. Dies ermöglicht mikroskopische Analysen des Unfallgeschehens. Im laufenden Projekt wird neben der Konzeption und Umsetzung der Metadatenbank auch die detaillierte Recherche für die in **TABELLE 1** aufgeführten Länder durchgeführt. Dabei wird unterschieden zwischen vollständig recherchierten Datenquellen und solchen, deren Existenz bekannt ist, zu denen jedoch bisher keine weiteren Angaben extrahiert werden.

4 METHODIK DER OBJEKTIVEN BEWERTUNG

Ein zentrales Anliegen des Projektes ist die objektivierte Bewertung der Qualität der zahlreich existierenden Datenquellen. Zur Bewertung der Metrik wird – auch mit Blick auf eine praxistaugliche Anwendung der Metadatenbank – die Eignung der Datenquellen für die Beantwortung aktueller und zukünftiger Forschungsfragen überprüft. Für deren Zusammenstellung wurden Mitgliedsunternehmen der FAT aufgerufen, für sie relevante Fragestellungen aus dem Bereich der Konzeption, Entwicklung und Bewertung von Verkehrssicherheitsmaßnahmen zusammenzutragen und den Forschern zur Verfügung zu stellen. Auf diesem Weg entstand ein interdisziplinärer Fragenkatalog mit mehr als 190 Fragen, von denen 120 inhaltlich geprägte Forschungsfragen und die restlichen Fragen zu den Charakteristika und Metadaten der recherchierten Datenquellen sind.

Land	Anzahl der Datenquellen		Anzahl der vollständig recherchierten Datenquellen	
	National	In-Depth	National	In-Depth
Dänemark	2	-	2	-
Deutschland	3	9	2	1
Frankreich	1	1	1	-
Griechenland	1	1	1	1
Schweden	1	2	-	1
Tschechische Republik	1	1	1	1
USA	4	3	4	2
Andere	17	15	-	-
Gesamt	30	32	11	6

TABELLE 1 Anzahl und Art der recherchierten Datenquellen nach Ländern (© Vufo)

Die Forschungsfragen werden semantisch analysiert, um eine Zuordnung der mit der Fragestellung verknüpften Inhalte zu den Parametern der Metadatenbank zu bewirken. Für jede Frage sind somit die Anforderungen an die Datenquelleninhalte definiert. Zur vereinfachten prozessualen Behandlung werden diese Anforderungen an die Existenz entsprechender Inhalte in Binärcodes auf Parameter Ebene abgelegt. Anschließend gleicht ein automatisiert durchführbarer Matching-Prozess die frageseitigen Anforderungen mit der Existenz von Datenquelleninhalten ab und speichert das Ergebnis in einer Ergebnismatrix ab. Der Inhalt dieser Matrix gibt schließlich für jede Datenquelle den Anteil an Variablen an, die mit den zur Beantwortung der jeweiligen Forschungsfrage notwendigen Variablen übereinstimmen. **BILD 1** zeigt im Quervergleich von fünf ausgewählten Datenquellen, wie viele der 120 Forschungsfragen vollständig beziehungsweise zu welchem Prozentsatz beantwortet werden können. Mit der Datenquelle aus „EU-Land I“ können beispielsweise 115 der im Katalog inkludierten Fragen vollständig beantwortet werden.

Ergänzend ist anzumerken, dass keine Einschätzung zur Vollständigkeit und Plausibilität der tatsächlichen (Unfall-)Daten stattfindet.

Diese Qualitätsprüfung sowie die Entscheidung, ob eine Datenquelle auch bei einem geringeren Prozentwert an beantworteten Fragen und damit fehlenden Informationen für eine Analyse herangezogen wird, obliegen den Anwendern.

5 BEISPIELANWENDUNGEN

In der Konzeptionsphase neuer Sicherheitssysteme, Verbraucherschutz- und legislatorischer Vorgaben gehören retrospektive und zunehmend auch prospektive Analysen von Verkehrsunfalldatenquellen zu den Standardwerkzeugen. Dabei finden oft auch Quervergleiche zwischen den Maßnahmen und Situationen verschiedener Länder und Regionen statt, die einen globalen Blick auf das jeweilige Thema bieten und gegebenenfalls auch zur Identifikation bereits bewährter Maßnahmen beitragen. Zudem agieren die meisten Unternehmen der Automobilindustrie in vielen Ländern und auf verschiedenen Kontinenten, was eine Analyse der Datenquellen aus diesen verschiedenen Staaten beziehungsweise Wirtschaftsräumen erfordert.

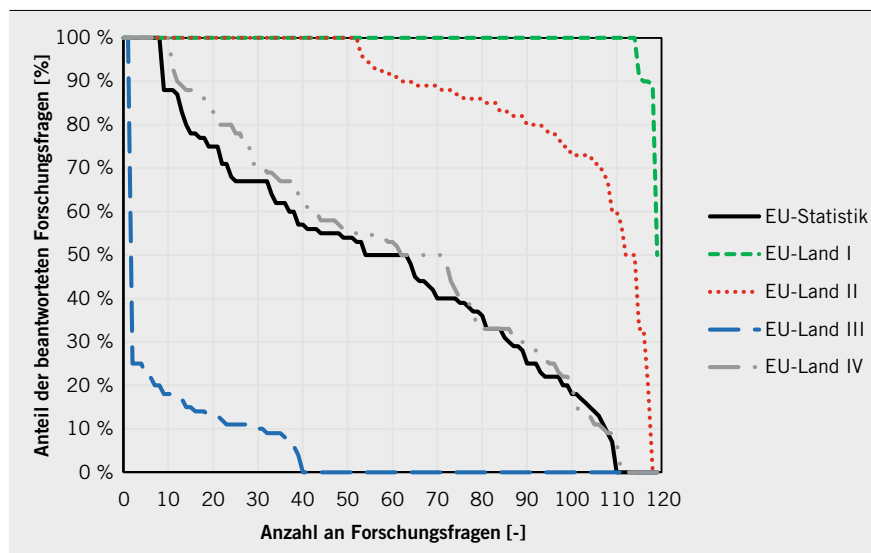


BILD 1 Visualisierung der Nutzbarkeit von fünf ausgewählten Datenquellen aus jeweils einem EU-Land für die Beantwortung der in der Metadatenbank hinterlegten Forschungsfragen (© Vufo)

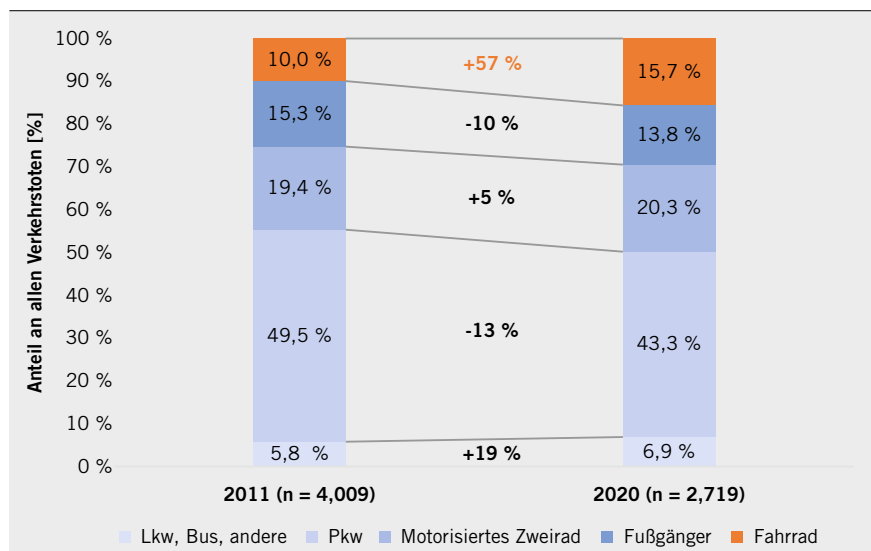


BILD 2 Verteilung der Arten von Verkehrsbeteiligten in Unfällen mit Todesopfern innerhalb Deutschlands (2011 und 2020) (© Vufo)

Land/Region	Anzahl an Datenquellen	Art der Datenquelle	Unfälle mit Radfahrern	Helmnutzung	Anprallrichtung Radfahrer/Fahrzeug	Einzelverletzungen
Weltweit	1	In-Depth	Ja	Ja	Ja	Nein
EU-Statistik	1	National	Ja	Ja	Ja	Nein
EU-Land I	1	In-Depth	Ja	Ja	Ja	Ja
EU-Land II	1	In-Depth	Ja	Ja	Ja	Ja
EU-Land III	1	In-Depth	Ja	Ja	Nein	Nein
EU-Land IV	1	National	Ja	Ja	Ja	Nein
EU-Land V	1	National	Ja	Ja	Nein	Nein
Nordamerika	4	National und In-Depth	Ja	Ja	3 Ja 1 Nein	1 Ja 3 Nein

TABELLE 2 Ergebnis für ausgewählte Parameter zur Radfahrersicherheit in verschiedenen Datenquellen (© Vufo)

An einem konkreten Beispiel soll eine ausgewählte Anwendungsmöglichkeit skizziert werden. Mit Blick auf das deutsche Unfallgeschehen und die durch die Covid-19-Pandemie beschleunigte Zunahme des Radverkehrs soll der Fokus auf ebendieser Verkehrsbeteiligungsart liegen. **BILD 2** verdeutlicht auf Basis von Zahlen des Statistischen Bundesamts (Destatis) den steigenden Anteil getöteter Radfahrer in der letzten Dekade [3]. Interessante Aspekte für Analysen zur Radfahrersicherheit sind speziell zur Verletzungsentstehung die Anprallkonstellation sowie zu potenziellen Schutzmaßnahmen die Fahrradhelmnutzung. Mithilfe einfacher Abfragen können in der Metadatenbank effizient alle Datenquellen mit den relevanten Informationen gefiltert werden. Im Beispiel in **TABELLE 2** ist ersichtlich, dass elf der enthaltenen Datenquellen Informationen zu Unfällen mit Radfahrern und dem Aspekt der Helmnutzung bereitstellen. Für weiterführende Analysen zur Anprallkonstellation verbleiben noch fünf Datenquellen; tiefergehende Untersuchungen auf Basis von Einzelverletzungen sind dagegen nur mit drei Datenquellen möglich.

Da der Zugang zu Einzel- beziehungsweise Rohdaten oft limitiert und die Datennutzung mit erheblichen Kosten assoziiert ist, müssen Anwender und Anwenderinnen eine geeignete Datenquellenauswahl für ihre Anwendungsfälle treffen. Die durch den Matching-Prozess erzeugten Ergebnisse bieten ihnen eine Grundlage zur objektiven Bewertung der Datenquellen. Die Ergebnismatrix stellt dabei die Eignung einzelner Datenquellen für alle oder konkrete, anwenderspezifische Themen dar.

Für die Implementierung zukünftiger Fragen enthält die Metadatenbank ein Formular, in dem Anwender und Anwenderinnen ihre Frage(n) einbringen und die für die Beantwortung notwendigen Parameter identifizieren können. Nach der Durchführung des skriptbasierten Matching-Prozesses wird die Ergebnismatrix erzeugt beziehungsweise eine vorhandene aktualisiert.

5 ZUSAMMENFASSUNG UND AUSBLICK

Das Forschungsvorhaben hat das Ziel, diverse Unfalldatenquellen aus verschiedenen Ländern zu recherchieren, zu inventarisieren und objektiv zu bewerten. Zu diesem Zweck wird eine Metadatenbank konzipiert und mit ausgewählten Datenquellen gefüllt. Die objektive Bewertung beruht auf einem Fragenkatalog mit aktuellen Forschungsfragen und einem Matching-Prozess, der den Inhalt der Fragen mit der Verfügbarkeit von Variablen je Datenquelle abgleicht. Das Ergebnis ist eine Matrix, die für jede Forschungs-

frage die Verfügbarkeit notwendiger Variablen pro Datenquelle anzeigt. Der aktuelle Fragenkatalog und die darauf basierende Bewertung zur Eignung der recherchierten Datenquellen stützen sich maßgeblich auf Forschungsfragen aus Forschungs- und Entwicklungsabteilungen von in der FAT vereinigten Automobilunternehmen. Für eine breiter aufgestellte Bewertung kann der Fragenkatalog um Fragen von Gesetzgebungsorganen, Behörden, Verbänden, Hochschulen oder Verbraucherschutzorganisationen erweitert werden. Denkbar sind sinnvolle Ergänzungen, beispielsweise Daten mit Wetter- oder Verkehrsflussinformationen sowie Datenquellen mit Fahrverhaltensbeobachtungen. Die Metadatenbank ist ein nützliches Werkzeug, um den zunehmend datengetriebenen Fahrzeugentwicklungsprozess effizienter zu gestalten, indem die Rechercheaufwände für geeignete Datenquellen zur Beantwortung der relevanten Entwicklungsfragen stark reduziert werden.

LITERATURHINWEISE

- [1] Ziegler, J.; Liers, H.; Chanove, A.; Pohle, M.: Objective assessment of database quality for use in the automotive research and development process. FAT-Schriftenreihe 343, Berlin, 2021
- [2] Gidas: German In-Depth Accident Study (Gidas): Online: www.gidas.org, aufgerufen: 10. August 2021
- [3] Destatis: Verkehr. Verkehrsunfälle. Fachserie 8, Reihe 7, Statistisches Bundesamt, Wiesbaden, 2021

DANKE

Das Forschungsprojekt „Objektive Einschätzung der Qualität von Datenbanken zur Verwendung dieser im Forschungs- und Fahrzeugentwicklungsprozess“ wird von der Forschungsvereinigung Automobiltechnik e. V. (FAT) aus Eigenmitteln gefördert und von der Verkehrsunfallforschung an der TU Dresden GmbH (Vufo) und dem Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme (IVI) durchgeführt und wurde vom FAT-Arbeitskreis 3 „Biomechanik und Unfallforschung“ unter der Leitung von Dr.-Ing. Michael Düring (Volkswagen AG) und Dipl.-Ing. (FH) Michael Wagner (Continental AG) initiiert. Die Autoren und die Autorin danken der FAT für die Förderung und allen beteiligten Projektpartnern.



READ THE ENGLISH E-MAGAZINE

Test now for 30 days free of charge: www.atz-worldwide.com