
**Konzept zur faktischen Anonymisierung der Einzeldaten der
Statistik der Straßenverkehrsunfälle ab 2002**
SUF STVU 2002ff

Version	Datum	Bearbeiter	Bemerkung
1.0	29.05.2013	FDZ ST	Mit den StaLÄ und dem StBA abgestimmte Version
1.1	29.07.2013	FDZ ST	Redaktionelle Korrekturen in MM-Liste

1. Vorbemerkungen

Mit Inkrafttreten des Bundesstatistikgesetzes¹ im Jahr 1987 wurde, abweichend vom generellen Geheimhaltungsgebot nach §16 Abs. 1 durch §16 Abs. 6 die Möglichkeit eröffnet, dass von den Statistischen Ämtern des Bundes und der Länder „für die Durchführung wissenschaftlicher Vorhaben ... Einzelangaben an Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden [dürfen], wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können und die Empfänger Amtsträger, für den öffentlichen Dienst verpflichtete oder Verpflichtete nach §7 sind“. Diese sogenannte faktische Anonymität erlaubt eine projektbezogene Herausgabe von amtlichen Einzeldaten an einen eingeschränkten Empfängerkreis, wenn der Aufwand einer Reidentifikation den Nutzen der zu gewinnenden Informationen übersteigt. Somit ist für die Beurteilung eine Aufwand-Nutzen-Abwägung durchzuführen, wobei die organisatorischen Hürden des Datenzuganges (Nutzungsverträge, evtl. zivil- und strafrechtliche Konsequenzen) und der Zugangswege² als Aufwand einzubeziehen sind. Das vorliegende Konzept beschreibt die jahresweise Aufbereitung der Statistik der Straßenverkehrsunfälle für den Zugangsweg des Scientific-Use-Files (SUF), d.h. zur direkten Nutzung an einer wissenschaftlichen Einrichtung.

2. Nutzen der Reidentifikation

Das Unverhältnismäßigkeitsgebot des BStatG impliziert, dass eine Verletzung der faktischen Anonymität nur im Falle einer als nutzbringend einzustufenden Reidentifikation gegeben sein kann. Der Nutzen, der durch die Zuordnung eines Merkmals entsteht, hängt stark vom Informationsgehalt und somit von der Sensibilität des zu reidentifizierenden Merkmals ab. Ein Merkmal, das dem Datenangreifer keine verwertbare zusätzliche Information in diesem Sinne bringt, liefert kein plausibles Motiv für einen Datenangriff und gilt daher als unsensibel.

¹ Bundesstatistikgesetz vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 3 des Gesetzes vom 07. September 2007 (BGBl. I S. 2246)

² Für die verschiedenen Zugangswege vgl. <http://www.forschungsdatenzentrum.de/datenzugang.asp>

Über Unfälle, Unfallorte, Unfallzeiten und Unfallhergang veröffentlichen die Polizeidienststellen regelmäßig und ausführlich Presseinformationen³. Dies betrifft insbesondere Unfälle mit Verletzten oder Getöteten, aber auch die Sachschadensunfälle. Durch erfolgreiche Identifikation eines Unfalls im SUF kann daher kein Wissensgewinn und damit auch kein Nutzen unterstellt werden.

Als sensibel (und damit schutzbedürftig) sind hingegen die ebenfalls in der Statistik enthaltenen Individualangaben zu persönlichen Angaben der Unfallbeteiligten, zur Schuldfrage sowie insbesondere die Erlangung von Kenntnis über den Einfluss berauschender Mittel einzustufen, da diese Angaben meist nicht in den vorgenannten Veröffentlichungen zu finden sind und für die Unfallbeteiligten mit teils erheblichen Konsequenzen verbunden sein können.

3. Basismaterial

Ausgangsmaterial der SUFs sind die Einzeldaten jährlichen Ergebnisse der Statistik der Straßenverkehrsunfälle (EVAS 46241). Diese umfassen alle im Berichtsjahr durch die Polizeibehörden erfassten Straßenverkehrsunfälle mit Ausnahme der sonstigen Sachschadensunfälle, also der Unfälle mit ausschließlich Sachschaden und ohne nicht mehr fahrbereite Kraftfahrzeuge sowie ohne Alkoholeinwirkung (Kategorie-5-Unfälle/ „Bagatellunfälle“)⁴. Ohne diese enthalten die SUFs somit jährlich Angaben zu ca. 400'000 Verkehrsunfällen.

³ Beispielhaft sei auf <http://www.presseportal.de/polizeipresse/> , <http://www.berlin.de/polizei/presse-fahndung/presse.html> oder <http://www.sachsen-anhalt.de/index.php?id=polizei> verwiesen.

⁴ Die aus technischen Gründen in einzelnen Ländern vorliegenden Individualinformationen zu den Unfällen der Kategorie 5 sind lt. Straßenverkehrsunfallstatistikgesetz §2 Abs. 2 nicht Bestandteil der Statistik, werden daher nicht geprüft bzw. plausibilisiert und können nicht zur externen Nutzung bereitgestellt werden.

4. Anonymisierungsmaßnahmen

Zum Erreichen der faktischen Anonymität und unter Berücksichtigung der vorgenannten Schutzinteressen werden die die Einzelangaben wie folgt anonymisiert:

- Die Angaben werden auf den Merkmalsträger „Unfall“ zusammengefasst.
- Im Datensatz verbleiben lediglich die in Anlage 1 aufgeführten Merkmale, alle anderen Angaben werden ersatzlos gelöscht.
- Unfälle der Kategorie 6 („sonstige Sachschadensunfälle unter dem Einfluss berauschender Mittel“) unterscheiden sich inhaltlich von den nicht enthaltenen Unfällen der Kategorie 5 nur durch das besonders sensible Merkmal der Alkoholeinwirkung. Es besteht kein Bezug zu den ansonsten im SUF verbleibenden schweren Unfällen der Kategorien 1 - 4. Zum Schutz der Beteiligten werden die Unfälle der Kategorie 6 deshalb ersatzlos gelöscht
- Die Paginiernummern werden durch eine systemfreie Unfallnummer ersetzt.
- Die Art der Unfallbeteiligungen und die Höhe des entstandenen Schadens stellen für die Wissenschaft sehr interessante Merkmale dar. Um hier trotz der Beschränkung auf die Unfallebene auswertbare Informationen bereitzustellen, werden
 - die Sachschadenssummen der einzelnen Beteiligten zu einer Gesamtsumme je Unfall aufsummiert (BSACHSCHADENSUM) und
 - die Anzahl der Unfallbeteiligten in verschiedenen Beteiligungskategorien eingefügt (UANZVERKBET1 – UANZVERKBET99)
- Die Angabe der vorläufigen Unfallursache wird auf die nicht personenbezogenen Angaben beschränkt (UURSACHE1 und UURSACHE2). Beteiligtenbezogene Ursachen (wie z.B. Alkoholeinfluss, Geschwindigkeits-, Vorfahrts- oder Abstandsverstoß usw.) werden nicht ausgewiesen.

-
- Um die Information zum Einfluss berauschender Mittel zu unterdrücken, wird das Merkmal UALKOHOLEINW entfernt.
 - Um die Ableitung der Schuld der Beteiligten durch eine zu detaillierte Beschreibung des Unfallherganges zu vermeiden, werden die Angaben zur Untergliederung des Unfalltyps nicht bereitgestellt. Es verbleibt die Kategorisierung entsprechend der 7 Basistypen (UTYP1).
 - Bei den Alleinunfällen (Unfälle mit nur einem Beteiligten) ergibt sich die Schuldfrage automatisch. Um den Informationsgewinn aus der Identifikation eines solchen Unfalls zu mindern, werden die Sachschadenshöhen bei diesen Unfällen wie folgt recodiert:
 - bei Schadenshöhen unter 1'000 EUR auf den Wert 500,
 - bei Schadenshöhen zwischen 1'000 und 4'999 EUR auf den Wert 3'000,
 - bei Schadenshöhen zwischen 5'000 und 9'999 EUR auf den Wert 7'500 sowie
 - bei Schadenshöhen von 10'000 EUR oder mehr auf den Wert 10'000.

Für die Nutzung der durch die beschriebenen Maßnahmen erzeugten Scientific-Use-Files ist von interessierten Wissenschaftlern für jedes konkrete Projekt ein Nutzungsantrag an eines der Forschungsdatenzentren des Bundes oder der Länder zu stellen. Dieser wird im konkreten Einzelfall durch einen Nutzungsvertrag zwischen einem Statistischen Amt und der entsprechenden wissenschaftlichen Einrichtung begleitet, in welchem u.a. ein Gebot zur Geheimhaltung der Einzelangaben sowie ein Verbot der gezielten Deanonymisierung enthalten sind. Für Verstöße gegen diese Regelungen sind finanzielle Vertragsstrafen vorgesehen. Zusätzlich ist der Datenzugang nur Personen gestattet, die vorab auf statistische Geheimhaltung entsprechend §1 des Verpflichtungsgesetzes⁵ verpflichtet wurden. Aus dieser Rechtsnorm ergeben sich im Falle eines Verstoßes ebenfalls Konsequenzen. Außerdem ist der Datenzugang mit Kosten⁶ verbunden.

⁵ nutzungsberechtigt sind ausschließlich „Amtsträger, für den öD besonders Verpflichtete oder Verpflichtete nach §16 Absatz 7 BStatG“, welcher auf §1 des Verpflichtungsgesetzes vom 2. März 1974 (BGBl. I S 469 Artikel 42), zuletzt geändert durch das Gesetz vom 15. August 1974 (BGBl. I S. 1942), verweist.

⁶ entsprechend der jeweils geltenden Entgeltregelungen der FDZ

Fazit

Durch die beschriebene Begrenzung des Informationsumfanges der Statistik der Straßenverkehrsunfälle (insbesondere durch die Beschränkung auf Angaben der Unfallebene, den Wegfall der Informationen zur persönlichen Schuld bzw. Art der Beteiligung von Einzelpersonen) ist der Nutzen einer erfolgreichen Reidentifikation einzelner Unfallbeteiligter erheblich reduziert. Die enthaltenen Informationen können unter Berücksichtigung der Veröffentlichungspraxis größtenteils als ohnehin bekannt vorausgesetzt werden.

Gleichzeitig stellen die Umsetzung der organisatorischen Hürden (Antragsverfahren, Nutzungsvertrag, Entgelt) sowie die bei erfolgreicher Reidentifikation drohenden rechtlichen Konsequenzen einen im Verhältnis dazu unverhältnismäßig großen Aufwand dar.

Die entsprechend dieses Konzeptes anonymisierten Dateien sind daher als faktisch anonym einzustufen.

Anlage 1

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002			
Variable	Typ¹⁾	Länge	Etikett
UIDENT	Char	8	[UIDENT] systemfreie Unfallnummer
ULAND	Char	2	[ULAND] Bundesland 01 Schleswig-Holstein 02 Hamburg 03 Niedersachsen 04 Bremen 05 Nordrhein-Westfalen 06 Hessen 07 Rheinland-Pfalz 08 Baden-Württemberg 09 Bayern 10 Saarland 11 Berlin 12 Brandenburg 13 Mecklenburg-Vorpommern 14 Sachsen 15 Sachsen-Anhalt 16 Thüringen
UREGBEZ	Char	1	[UREGBEZ] Regierungsbezirk
UKREIS	Char	2	[UKREIS] Kreis
UGEMEINDE	Char	3	[UGEMEINDE] Gemeinde
UAGS	Char	8	[UAGS] amtlicher Gemeindeschlüssel
UJAHR	Char	4	[UJAHR] Jahr
UMONAT	Char	2	[UMONAT] Monat
UTAG	Char	2	[UTAG] Tag
USTUNDE	Char	2	[USTUNDE] Zeitpunkt des Unfalls (Stundenangabe)
UMINUTE	Char	2	[UMINUTE] Zeitpunkt des Unfalls (Minutenangabe)
UANZBET	Num		[UANZBET] Anzahl der Unfallbeteiligten

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002

Variable	Typ¹⁾	Länge	Etikett
UANZGETOE	Num		[UANZGETOE] Anzahl der Getöteten (verstorben bis 30 Tage nach dem Unfall)
UANZSCHWER	Num		[UANZSCHWER] Anzahl der Schwerverletzten (mind. 24h stationäre Behandlung)
UANZLEICHT	Num		[UANZLEICHT] Anzahl der Leichtverletzten
UANZVERKBET1	Num		[UANZVERKBET1] Anzahl der beteiligten Krafträder mit Versicherungskennzeichen (Schl-Nr: 01 - 04)
UANZVERKBET2	Num		[UANZVERKBET2] Anzahl der beteiligten Krafträder mit amtlichem Kennzeichen (Schl-Nr: 11 - 13, 15)
UANZVERKBET3	Num		[UANZVERKBET3] Anzahl der beteiligten Personenkraftwagen und Wohnmobile (Schl-Nr: 21, 22)
UANZVERKBET4	Num		[UANZVERKBET4] Anzahl der beteiligten Busse (Schl-Nr: 31 - 35)
UANZVERKBET5	Num		[UANZVERKBET5] Anzahl der beteiligten Liefer- und Lastkraftwagen (Schl-Nr: 40 - 46, 48, 51 - 55, 57, 58)
UANZVERKBET6	Num		[UANZVERKBET6] Anzahl der beteiligten übrigen, sonstigen oder unbekanntem Fahrzeuge, Fuhrwerke (Schl-Nr: 59, 91, 92)
UANZVERKBET7	Num		[UANZVERKBET7] Anzahl der beteiligten Schienenfahrzeuge (Schl-Nr: 61, 62)
UANZVERKBET8	Num		[UANZVERKBET8] Anzahl der beteiligten Fahrräder/ Pedelecs (Schl-Nr: 71, 72)
UANZVERKBET9	Num		[UANZVERKBET9] Anzahl der beteiligten Fußgänger, Tierführer, Tiertreiber mit und ohne Sport- oder Spielgerät oder Handwagen/-karren (Schl-Nr: 81 - 84)
UANZVERKBET10	Num		[UANZVERKBET10] Anzahl der beteiligten anderen Personen (Schl-Nr: 93)
UANZVERKBET99	Num		[UANZVERKBET99] Anzahl der Beteiligten mit unbekannter Art der Beteiligung (Schl-Nr: 00, 10, 50, 80)

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002

Variable	Typ ¹⁾	Länge	Etikett
UART	Char	1	<p>[UART] Unfallart</p> <ul style="list-style-type: none"> 1 Zusammenstoß mit anfahrendem/ anhaltendem/ ruhendem Fahrzeug 2 Zusammenstoß mit vorausfahrendem/ wartendem Fahrzeug 3 Zusammenstoß mit seitlich in gleicher Richtung fahrendem Fahrzeug 4 Zusammenstoß mit entgegenkommendem Fahrzeug 5 Zusammenstoß mit einbiegendem/ kreuzendem Fahrzeug 6 Zusammenstoß zwischen Fahrzeug und Fußgänger 7 Aufprall auf Fahrbahnhindernis 8 Abkommen von Fahrbahn nach rechts 9 Abkommen von Fahrbahn nach links 0 Unfall anderer Art
UCHARUNFST1	Char	1	<p>[UCHARUNFST1] Charakteristik der Unfallstelle (1. Kennziffer)</p> <ul style="list-style-type: none"> 1 Kreuzung 2 Einmündung 3 Grundstücksein-/ausfahrt 4 Steigung 5 Gefälle 6 Kurve
UCHARUNFST2	Char	1	<p>[UCHARUNFST2] Charakteristik der Unfallstelle (2. Kennziffer)</p> <ul style="list-style-type: none"> 1 Kreuzung 2 Einmündung 3 Grundstücksein-/ausfahrt 4 Steigung 5 Gefälle 6 Kurve
UCHARUNFST3	Char	1	<p>[UCHARUNFST3] Charakteristik der Unfallstelle (3. Kennziffer)</p> <ul style="list-style-type: none"> 1 Kreuzung 2 Einmündung 3 Grundstücksein-/ausfahrt 4 Steigung 5 Gefälle 6 Kurve

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002

Variable	Typ ¹⁾	Länge	Etikett
UBESONUNFST1	Char	1	[UBESONUNFST1] Besonderheit der Unfallstelle (1. Kennziffer) 2 Schienengleicher Wegübergang 3 Fußgängerüberweg (Zebrastreifen) 4 Fußgängerfurt 5 Haltestelle 6 Arbeitsstelle 7 Verkehrsberuhigter Bereich
UBESONUNFST2	Char	1	[UBESONUNFST2] Besonderheit der Unfallstelle (2. Kennziffer) 2 Schienengleicher Wegübergang 3 Fußgängerüberweg (Zebrastreifen) 4 Fußgängerfurt 5 Haltestelle 6 Arbeitsstelle 7 Verkehrsberuhigter Bereich
UBESONUNFST3	Char	1	[UBESONUNFST3] Besonderheit der Unfallstelle (3. Kennziffer) 2 Schienengleicher Wegübergang 3 Fußgängerüberweg (Zebrastreifen) 4 Fußgängerfurt 5 Haltestelle 6 Arbeitsstelle 7 Verkehrsberuhigter Bereich
ULICHTZANL	Char	1	[ULICHTZANL] Lichtzeichenanlage (falls vorhanden) 8 Lichtzeichenanlage in Betrieb 9 Lichtzeichenanlage außer Betrieb
UFAHRBKfZ	Char	1	[UFAHRBKfZ] mindesten 1 Kfz nicht mehr fahrbereit 1 Mindestens 1 Kfz nicht mehr fahrbereit leer Alle beteiligten Kfz noch fahrbereit

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002

Variable	Typ ¹⁾	Länge	Etikett
UGESCHWBEGR	Char	3	[UGESCHWBEGR] Geschwindigkeitsbegrenzung
			005 5 km/h
			010 10 km/h
			015 15 km/h
			020 20 km/h
			025 25 km/h
			030 30 km/h
			040 40 km/h
			050 50 km/h
			060 60 km/h
			070 70 km/h
			080 80 km/h
			090 90 km/h
			100 100 km/h
			110 110 km/h
			120 120 km/h
130 130 km/h			
Z07 Fußgängerzone/ verkehrsberuhigter Bereich (Zone Schrittgeschwindigkeit)			
Z20 Zone 20 km/h			
Z30 Zone 30 km/h			
ULICHTVERH	Char	1	[ULICHTVERH] Lichtverhältnisse
			0 Tageslicht
			1 Dämmerung
			2 Dunkelheit
USTRZUST1	Char	1	[USTRZUST1] Straßenzustand (1. Kennziffer)
			0 trocken
			1 naß/ feucht
			2 winterglatt
			5 schlüpfrig (durch Öl, Dung, Laub usw.)
USTRZUST2	Char	1	[USTRZUST2] Straßenzustand (2. Kennziffer)
			0 trocken
			1 naß/ feucht
			2 winterglatt
			5 schlüpfrig (durch Öl, Dung, Laub usw.)

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002

Variable	Typ ¹⁾	Länge	Etikett
UAUFPRHIND	Char	1	[UAUFPRHIND] Aufprall auf Hindernis 0 Baum 1 Mast 2 Widerlager 3 Schutzplanke 4 sonstiges Hindernis 5 kein Aufprall
UURSACHE1	Char	2	[UURSACHE1] vorläufig festgestellte Unfallursache (1. Kennziffer) 01 Straßenverhältnisse (z.B. Verunreinigungen, Schnee, Regen, Schlamm) 02 Straßenzustand (z.B. Spurrillen, mangelh. Beleuchtung/ Beschilderung) 03 Witterungseinflüsse (z.B. Wind, Starkregen, Nebel, blendende Sonne) 04 Hindernisse (z.B. Tiere, Arbeitsstellen, Gegenstände) 05 sonstige Ursachen [leer] Ursachen durch Unfallbeteiligte
UURSACHE2	Char	2	[UURSACHE2] vorläufig festgestellte Unfallursache (2. Kennziffer) 01 Straßenverhältnisse (z.B. Verunreinigungen, Schnee, Regen, Schlamm) 02 Straßenzustand (z.B. Spurrillen, mangelh. Beleuchtung/ Beschilderung) 03 Witterungseinflüsse (z.B. Wind, Starkregen, Nebel, blendende Sonne) 04 Hindernisse (z.B. Tiere, Arbeitsstellen, Gegenstände) 05 sonstige Ursachen [leer] Ursachen durch Unfallbeteiligte
UORTSLAGE	Char	1	[UORTSLAGE] Ortslage 1 innerorts 2 außerorts
UKATEGORIE	Char	1	[UKATEGORIE] Unfallkategorie 1 Unfall mit Getöteten 2 Unfall mit Schwerverletzten 3 Unfall mit Leichtverletzten 4 schwerwiegender Unfall mit Sachschaden (mind. ein Kfz nicht fahrbereit)

Scientific-Use-Files der Straßenverkehrsunfallstatistik ab 2002

Variable	Typ ¹⁾	Länge	Etikett
UTYP1	Char	1	[UTYP1] Unfalltyp 1 Fahr Unfall 2 Abbiegeunfall 3 Einbiegen-/ Kreuzen-Unfall 4 Überschreitenunfall 5 Unfall durch ruhenden Verkehr 6 Unfall im Längsverkehr 7 sonstiger Unfall
USTRKLASSE	Char	1	[USTRKLASSE] Straßenklasse 1 Autobahn 2 Bundesstraße 3 Landesstraße 4 Kreisstraße 5 Gemeinde- oder sonstige Straße
UWOCHENTAG	Char	1	[UWOCHENTAG] Wochentag 1 Sonntag 2 Montag 3 Dienstag 4 Mittwoch 5 Donnerstag 6 Freitag 7 Samstag
BSACHSCHADENSUM	Num		[BSACHSCHADENSUM] Summe des entstandenen Sachschadens bei allen Beteiligten in EUR (vorläufige Schadenshöhe am Unfallort bestimmt) bei Alleinunfällen recodiert in: 500 Sachschaden unter 1'000 EUR 3000 Sachschaden zwischen 1'000 und 4'999 EUR 7500 Sachschaden zwischen 5'000 und 9'999 EUR 10000 Sachschaden 10'000 EUR oder mehr

¹⁾ Num = numerische Variable, Char = alphanumerische Variable (String)