Description of data-sources used in SafetyCube

Deliverable 3.1

SafetyCube

Description of data-sources used in SafetyCube

Work package 3, Deliverable #1

Please refer to this report as follows:
Hagström, L., Thomson, R., Hermitte, T., Weijermars, W., Bos, N., Talbot, R., Thomas, P., Dupont, E., Martensen, H., Bauer, R., Hours, M., Høye, E., Jänsch, M., Murkovic, A., Niewöhner, W., Papadimitriou, E., Pérez, C., Phan, V., Usami, D., Vázquez-de-Prada, J., (2015), Description of data-sources used in SafetyCube, Deliverable 3.1 of the H2020 project SafetyCube.

Grant agreement No 633485 - SafetyCube - H2020-MG-2014-2015/ H2020-MG-2014_TwoStages

Project Coordinator:
Professor Pete Thomas, Loughborough Design School, Loughborough University, Ashby Road, Loughborough, LE11 3TU, UK

Project Start date: 01/05/2015

Duration: 36 months

Organisation name of lead contractor for this deliverable:
SAFER, SE.

Report Author(s):
Hagström, L., Thomson, R., (SAFER), Sweden
Hermitte, T., (LAB), France
Weijermars, W., Bos, N., (SWOV), Netherlands
Talbot, R., Thomas, P., (LOUGH), United Kingdom
Dupont, E., Martensen, H., Nuyttens, N. (BRSI), Belgium
Bauer, R., (KFV), Austria
Hours, M., (Ifsttar), France
Høye, E., (TØI), Norway
Jänsch, M., (MHH), Germany
Murkovic, A., (AVP), Slovenia
Niewöhner, W., (Dekra), Germany
Papadimitriou, E., (NTUA), Greece
Pérez, C., (ASPB), Spain
Phan, V., (CEESAR), France
Usami, D., (CTL), Italy
Vázquez-de-Prada, J., (Cidaut), Spain

Due date of	**31/10/2015**	**Submission date:**	**29/10/2015**

deliverable:			

Project co-funded by the by the Horizon 2020 Framework Programme of the European Union
Version: Draft2
Dissemination Level: PU Public

Co-funded by the Horizon 2020
Framework Programme of the European Union

Table of contents

Executive summary

Safety CaUsation, Benefits and Efficiency (SafetyCube) is a European Commission supported Horizon 2020 project with the objective of developing an innovative road safety Decision Support System (DSS) that will enable policy-makers and stakeholders to select and implement the most appropriate strategies, measures and cost-effective approaches to reduce casualties of all road user types and all severities.

This deliverable describes the available data in the form of an inventory of databases that can be used for analyses within the project. Two general types of data are available: one describing the involvement of different components for the road safety (vehicles, infrastructure, and the road user) and one describing the injury outcomes of a crash. These two database categories are available to the partners of SafetyCube and gathered in two excel tables. One table contains traffic databases (accident and naturalistic driving studies) and the second table contains injury databases. The tables contain information on 58 and 35 variables, respectively. The key information describing the databases that was needed for the inventory were items such as:

- Type of data collected (crashes, injuries, etc.)
- Documentation of the variables
- Sampling criteria for the data collected
- SafetyCube partners with access to the data
- Extent of data access (raw data vs. summary tables)

The tables contain 36 traffic accident databases, five naturalistic driving studies or field-tests and 22 injury databases where of four were coded in both sheets.

1 Introduction

This introduction describes the project SafetyCube and its aims. It also intends to give an overview of the purpose of this deliverable.

1.1 SAFETYCUBE

Safety CaUsation, Benefits and Efficiency (SafetyCube) is a European Commission supported Horizon 2020 project with the objective of developing an innovative road safety Decision Support System (DSS) that will enable policy-makers and stakeholders to select and implement the most appropriate strategies, measures and cost-effective approaches to reduce casualties of all road user types and all severities.
SafetyCube aims to:

1. develop new analysis methods for (a) Priority setting, (b) Evaluating the effectiveness of measures (c) Monitoring serious injuries and assessing their socio-economic costs (d) Cost-benefit analysis taking account of human and material costs
2. apply these methods to safety data to identify the key accident causation mechanisms, risk factors and the most cost-effective measures for fatally and seriously injured casualties
3. develop an operational framework to ensure the project facilities can be accessed and updated beyond the completion of SafetyCube
4. enhance the European Road Safety Observatory and work with road safety stakeholders to ensure the results of the project can be implemented as widely as possible

The core of the project is a comprehensive analysis of accident risks and the effectiveness and cost-benefit of safety measures focusing on road users, infrastructure, vehicles and injuries framed within a systems approach with road safety stakeholders at the national level, EU and beyond having involvement at all stages.

1.1.1 Work Package 3

Work package 3 of SafetyCube will assist the other operational work packages of SafetyCube (Work Packages 4, 5 and 6 and in some cases also WP 7) to identify, analyse and develope road safety information in a manner that is consistent for the project and facilitates integration of diverse road safety issues. The methods in Work Package 3 will be used to develop the content of a road safety Decision Support System that can be used both at a national and a European level.

1.2 PURPOSE OF THIS DELIVERABLE

The objective of this deliverable was to identify and describe the available data sources for the analysis in Work Packages 3 to 7. The analyses will cover road user behavior, vehicle, infrastructure, and injuries. The different Work Packages may have different topics but many analyses will be conducted from the same databases. This deliverable provides a searchable inventory of available databases.

2 Method

This first chapter describes the method of how the database inventory tables were developed and populated.

All analysts in SafetyCube were asked to identify what type of information they anticipate using in the project. This was used to identify the attributes of databases that should be documented so this information could be shared with other partners. Microsoft Excel was used to collate the data.

The key information that was requested for each data-base were items such as:

- Type of data collected (crashes, injuries, etc.)
- Documentation of the variables
- Sampling criteria for the data collected
- SafetyCube partners with access to the data
- Extent of data access (raw data vs. summary tables)

Partners provided information for each database including hyperlinks to any documentation. When a database was added by more than one partner (i.e. multi-user international databases, open access data, etc.), the resulting contributions were merged.

3 Result

The resulting inventory of databases is presented in this chapter. This includes the number of databases available to SafetyCube, what kind of data they include and to which work packages they could be useful.

The contents of accident- and injury data-bases differs strongly. Therefore the inventory process resulted in two tables, one for traffic accidents and naturalistic driving studies and one for injury databases. Some data sources provided information on both. There were 36 traffic accident databases, five naturalistic driving studies or field-tests and 22 injury databases. Four data-bases contained detailed information on injury as well as accident and were consequently coded in both sheets.

The databases cover international as well as national and regional data. Twenty of the traffic accident databases include information collected on-the-spot in-time, where the investigation takes place at the scene of the crash usually while the involved vehicles are still in situ, and 18 of the databases include statistical data that allows nationally representative estimates. All of Europe is covered in this data even if the most data comes from Germany, Spain, France, Italy, Netherlands, Sweden and United Kingdom. All vehicle types are covered in several of the databases. For the injury databases the majority (13/22) are road accident databases with a focus on injuries and the rest are mainly national hospital injury databases. The databases available include information that can be useful for all work packages of SafetyCube.

3.1 DATABASES FOR WP4 USE

For WP4 – Road user behaviour analysis - there are 31 databases that can be used for different analyses. Twenty of the databases for WP4 use include information on human factors and causes of accidents but it also includes five databases with naturalistic driving studies and exposure data and one survey on traffic behaviour.

3.2 DATABASES FOR WP5 USE

For WP5 – Infrastructure safety analysis there are 32 databases including information on infrastructure. 13 of them are in depth accident databases, 15 are statistical databases, two are naturalistic driving studies and three are a descriptive accident databases.

3.3 DATABASES FOR WP6 USE

WP6 – Vehicle safety analysis have access to 27 databases that can be used for different analyses. 18 of these include information on technology in the vehicle. Eleven are descriptive statistics databases, two are naturalistic driving studies and one is a vehicle damage describing database. The databases cover all types of vehicles.

3.4 DATABASES FOR WP7 USE

There are 21 injury databases, of which four also appear in the traffic data-base table. These 17 data-bases form the main source of information for WP7. Additionally 13 of the other traffic-data-bases also contain injury information and might be of additional use.

4 Intended usage of the Database tables

This chapter describes how the tables are intended to be used.

The datasheets are intended to be used when searching for a particular type of data. The data can be filtered on one or multiple variables to show which databases contain the desired information. Suggested variables to filter on are Type of investigation (On the spot in time, Exposure Data, Statistical Data etc.), Vehicle types covered (Car, Bicycles, Trucks etc.) or Data content (Information of injuries, cause of accidents, vehicle technology, infrastructure etc.).

As much of the data has a sensitive nature, most data is not freely available and can only be accessed by members of the original consortium. For this reason analyses of these databases have to be conducted the partner organisation with access. The results can then be presented within SafetyCube.

5 Variables

This section describes all the variables that are coded for each database in the two inventory tables.

5.1 TRAFFIC DATABASES

The variables coded for each database in the traffic accident datasheet are:

Database name – short name of the database
Full name – full name of the database
Link homepage – link to database or project website
Owner – name of owners of database
Investigation teams – name of organisations performing the investigations
Principal focus - Each accident data collection are created to answer to some research questions.
- Descriptive = oriented to descriptive statistics (not in-depth) like CARE, IRTAD databases
- Active = oriented to the accidental mechanism, to understand how the injury accidents occur (concerns pre-crash phase)
- Passive = oriented to injury mechanism, related to the protection and to understand how injuries occur (concerns crash phase)
- Tertiary = refers to all the measures taken and the resources mobilized following an accident in order to assist and care for the injured (concerns post-crash phase)
- Exposure = Exposure data (driven km, population, fleet, etc.)
- NDS = Naturalistic Driving Study

Main objective – describing the main objective of the data. For example: national statistics, detailed reconstructions, personal injury, crashworthiness, etc.
Type of investigation - what kind of investigations were performed to collect the data in the database
- On the spot in time – accident investigations on the accident scene when the involved vehicles are still there
- Retrospective – accident scene visited by investigators but after involved vehicle were moved
- Police report analysis – more technical analysis based on police reports, for example reconstruction
- Exposure data – Naturalistic driving studies or other collection of exposure data
- Statistical data – national statistic gathering

Vehicle types covered – what kind of vehicles are covered in the database
- Car
- Bicycles – also a note if e-bikes are included
- Powered two wheelers
- Pedestrians
- Truck
- Bus
- Other – train, farm equipment etc.

Years of data collection – which years the data was collected
Selection criteria – The criteria for collecting the data (e.g. vehicle involved, injury criteria etc.)
Sampling details – short description of statistical sampling plan or if it's census

Status – in process or finished
Number of accidents per year – an average over the years
Total number of vehicles up to now or expected
Number of vehicles year model ≥ 2000 – Number of vehicles of model year 2000 or newer in the database
Access to aggregated data
- Availability level – public, consortium, restricted
- Access procedures for external use – if possible for externals to get access to the data describe the procedure
- SafetyCube partner(s) that have access – list partners that have access to aggregated data of the database
- Software – software needed to get access to the data
- Available export format – access, cdv, etc.
- Language

Access to raw data
- Raw data available – Yes/No
- Availability level – public, consortium, restricted
- Access procedures for external use – if possible for externals to get access to the data describe the procedure
- SafetyCube partner(s) that have access – list partners that have access to aggregated data of the database
- Software – software needed to get access to the data
- Available export format – access, cdv, etc.
- Language

Region covered – which regions does the data cover
- Region – Part of country, country, part of Europe, global etc.
- Comment – more details to region covered if needed

Data content – What the database contains
- Total number of variables
- Pictures available – 1=yes, 0=no
- Scaled sketch available – 1=yes, 0=no
- Cause of injuries – information on the cause of injuries, 1=yes, 0=no
- Injuries – description of injuries to the road users, 1=yes, 0=no
- Injury severity – AIS coding of the injuries, 1=yes, 0=no
- Long term injury follow up – 1=yes, 0=no
- Human factors / cause of accidents - Use of a model base on human error such as HFF or DREAM, 1=yes, 0=no
- Vehicle technology – information on the presence and/or use of safety systems or vehicle technology on board (e.g. ESC, Airbags, ACC, LKW, GPS, etc.), 1=yes, 0=no
- Accident situation – Accident configuration and/or vehicle accidental situation. Most of time coded by pictograms (GFSO, LAB, CADAS accident type or DaCoTA accident type), 1=yes, 0=no
- Road user – Information regarding road user as height, weight, seat, belted etc, 1=yes, 0=no
- Infrastructure – Information like road characteristics, speed limit etc, 1=yes, 0=no
- Reconstruction – Computerised reconstruction of the accidents in a program like PC crash or similar, 1=yes, 0=no

Online documentation hyperlinks – please enter link if available
- Methodology – link to description of used methodology
- Coding convention – link to description of any coding convention
- Statistical sampling plan – link to description of statistical sampling plan
- Questionnaire – link to any questionnaire used in the investigations
- Glossary of terms – Link to any list with glossary of terms

Language – in what language(s) are the online documentation available.

Any other information – Any other information that can be of importance when selecting the database for analysis. Also a note if information is also included in another database (e.g. CARE).

5.2 INJURY DATABASES

The variables coded for each database in the injury datasheet are:

Database name – short name of the database
Full name – full name of the database
Link homepage – link to database or project website
Owner – name of owners of database
Main objective – describing the main objective of the data. For example: national statistics, injury prevention etc.
Data source – Police, Hospital, Accident & emergency, ambulance.
Injury population – Slight, Serious, Fatal
Selection criteria – The criteria for collecting the data (e.g. vehicle involved, injury criteria etc.)
Regional coverage – Regional, national, part of Europe, global etc.
Years available– which years the data was collected
Selection of traffic casualties possible? – Yes, No, It is only traffic casualties
Traffic Patients/ year – Number of traffic patients per year
Access to aggregated data
- Availability level – public, consortium, restricted
- Access procedures for external use – if possible for externals to get access to the data describe the procedure
- SafetyCube partner(s) that has access – list partners that has access to aggregated data of the database
- Software – software needed to get access to the data
- Available export format – access, cdv, etc.
- Language

Access to raw data
- Raw data available – Yes/No
- Availability level – public, consortium, restricted
- Access procedures for external use – if possible for externals to get access to the data describe the procedure
- SafetyCube partner(s) that have access – list partners that have access to aggregated data of the database
- Software – software needed to get access to the data
- Available export format – access, cdv, etc.
- Language

Region covered – which regions does the data cover
- Region – Part of country, country, part of Europe, global etc.
- Comment – more details to region covered if needed

Data content (most detailed) – What the database contains
- Total number of variables
- Diagnoses (injuries) available – Yes, No
- Coding of injuries – AIS, ICD10 etc.
- Road user type available – Yes, No
- Linking possible to police data – Yes, No

Online documentation
- Link – please enter link if available
- Language

Any other information – Any other information that can be of importance when selecting the database for analysis.

The Appendix A provides snapshots of parts of both datasheets. The Excel file is available on the SafetyCube project website.

6 Discussion/Conclusions

This chapter contains the conclusions of this deliverable.

The result of the database inventory table indicates that there are sufficient databases to cover all areas of interest for analyses within SafetyCube. The databases include all different kind of data needed; statistical, in-depth, naturalistic driving studies (NDS, field tests) etc. Even though, in the case of NDS, more reliable and representative data is a wish by most stakeholders.

This deliverable is a central resource for the project. It allows the different partners to identify information sources for analyses and find which partners have access to them. This document provides a focal point for discussing new and ongoing work items and is a platform for coordinating actions of different partners.

Appendix A Screenshots of the database inventory tables

Table 1 A screenshot of the traffic accident database inventory table

Database name	Full Name	Link homepage	Owner	Investigation Teams	Principle focus (active, passive, both etc)	Main Obje
					Descriptive / Active / Passive / Terciary / Exposure / NDS	
CARE	Community Road Accident Database	http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm	EU	Not Applicable. Compilation of national databases	Descriptive statistics	Statistics at th
CHILD	Advanced methods for improved Child Safety	http://cordis.europa.eu/project/rcn/63634_en.html	Consortium (Renault)	DE, ES, FR, IT, SE, UK	Passive	In depth severe c
DACOTA	Data Collection Transfer & Analysis	http://www.dacota-project.eu	Consortium (VSRC Loughborough)	CIDAUT Spain CTL University of Rome, Italy DGT Spain GIE RE PSA RENAULT France Hellenic Institute of Transport Greece IBSR Brussels IDIADA Spain IFSTTAR France INSIA Spain Medical University of Hanover Germany Universidad de Navarra Spain NTUA Greece SAFER Chalmers University of Technology, Sweden SWOV The Netherlands TSRC Loughborough University, UK	Active/passive	Establishing a Pan-Eu Investigation Netwo methodol
DEKRA accident database		www.dekra.de	DEKRA	DEKRA experts who get the task from police to reconstruct a single accident	Active / (Passive)	
DEKRA damage				DEKRA experts, who get the task to calculate		

…

…

Table 2 A screebshot of the injury database table

Database name	Full Name	Owner	Link homepage	Main Objective	Data source	Injury Population	Selection criteria	Regional coverage	Years available	selection of traffic casualties possible?	Tr Pat / y
					Police, Hospital, Acident&Emergency, Ambulance, IntensiveCare, Forensic	Slight, Serious, Fatalities					
Ambulance	Ambulance Trip Register	RIVM/AZN	http://www.ambulancezorg.nl/engels	Assess performance of Ambulances	Ambulance	All		National	2009+	Yes	5(
BronlinkedLMR	BRON-MAIS	SWOV	http://www.swov.nl/UK/Research/Cijfers/Toelichting-gegevensbronnen/geregistreerde-aantallen.html	improve severity indication of police by AIS	Police	All	Linked casualty to patient	National	1993 - 2009	Yes	1(
CARE	Community Road Accident Database	EU	http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm	Statistics at the EU level	Police	All	All national accident databases in the EU		1991 - 2004	it is only traffic casualties	
CMBD_AH Spain	National Hospital Discharge Database	Ministry of Health	http://www.msssi.gob.es/estadEstudios/estadisticas/cmbdhome.ht	Financial management	Hospital	Serious	All public hospitals and many private	National	2000	Yes	2(
DUHAT	Dades d'Urgències Hospitalàries per Accident de Trànsit (Hospital Emergency	ASPB	http://www.aspb.cat/quefem/documents_lesions_transit.htm	Injury Surveillance	Emergency Department	All	7 main hospitals in Barcelona	Municipality	1997	it is only traffic casualties	1(

…

…