Anonymised
Road Carriage (RC) micro-data

User Manual

Table of contents

1. Introduction

1.1. Background information

Eurostat compiles various data collected by Member States. For some data collections, data received by Eurostat are already in tabular form, for the others they are in the form of micro-data, *i.e.* sets of records containing information on individual respondents. Micro-data can be divided into business micro-data (from enterprises) and social micro-data (from individuals or households).

At the EU level, Regulation 557/2013 defines the rules of access to confidential data for scientific purposes. In order to make the micro-data available for research purposes the following conditions have to be fulfilled:

- The survey has to be listed in Article 7 and/or Article 8 of the Regulation 557/2013;
- The anonymisation criteria need to be proposed by Eurostat and then agreed by the countries providing the data.

1.2. Carriage of goods by road

According to the provisions of Regulation (EC) No 202/2010 amending Regulation (EC) No 6/2003 concerning the dissemination of statistics on the carriage of goods by road, researchers and the scientific community in general should have access for scientific purposes to the data transmitted to the Commission (Eurostat) under Regulation (EC) No 70/2012 of 18 January 2012 (recast, replacing Regulation No 1172/98), according to the principle laid down in Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics.

According to the provisions of Article 1 of Regulation (EC) 70/2012 of 18 January 2012 on statistical returns in respect of the carriage of goods by road, each Member State compiles statistics on the carriage of goods by road by means of goods road vehicles which are registered in that Member State and on the journeys made by such vehicle. In particular collected data relate to the following areas: vehicle, journey and goods.

1.3. Aim of this report

The extended report [1] provides a detailed anonymisation methodology proposed for the microdata on road carriage (RC) of goods collected by Eurostat (for a summary report, see [2]). This report first describes the structure of the RC micro-data. Then re-identification and plausible intruder scenarios related to RC data are discussed. Based on the intruder scenarios, the questionnaire sent to Member States and the input of some expert users, the analytical importance and the sensitivity of RC variables were investigated. Four possible anonymisation schemes were subsequently proposed, which suppress some variables, reduce detail in others and provide different levels of geographic detail.

The present report explains which method was selected to anonymise regular RC 2011 micro-data and how this method was applied on RC variables. It aims to help the user for a correct use and interpretation of anonymised RC data.

2. The Road Carriage anonymised micro-datasets

This section introduces the reader to the RC anonymised micro-datasets. Firstly, the anonymisation plan selected is cited and referenced. Secondly, the structure of the anonymised micro-datasets is described. Thirdly, the reader is provided with the definition of all RC variables along with the anonymisation rules applied on them.

2.1. Anonymisation plan

Document [1] (or [2] for a summary) describes the anonymisation strategy applied for the micro-data on road carriage (RC) of goods collected by Eurostat. The objective of the proposed anonymisation is to enable release of RC data in a manner compliant with Council Regulation No. 223/2009 and Commission Regulation 557/2013.

Based on the analysis of disclosure scenarios, computational work and feedback from the Member States after various meetings (RFT-WG, CGST, Expert Group on Confidentiality), the proposed scheme 3 [1] has been selected.

Please note that the anonymisation described is primarily designed for off-line release of data to researchers having signed a non-disclosure agreement. It is not intended to generate public-use files.

2.2. Structure

Anonymised individual records show the same structure as the original micro-data (for a description, see documentation on the Road Carriage data [3]). They consist of 3 linked datasets (figure 1) covering:

- A1 : vehicle-related variables
- A2 : journey-related variables
- A3 : goods-related variables

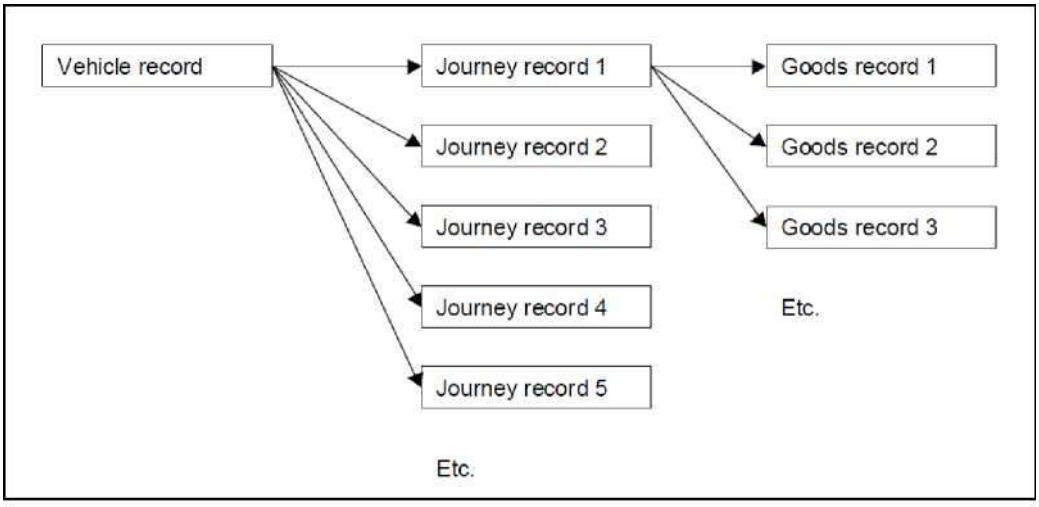

Figure 1. Link between datasets.

The key variables are Year, Quarter, QuestN (vehicle identifier), JournN (journey identifier), and GoodsN (good identifier). These variables allow linking the A1, A2 and A3 records.

The vehicle-related variables for a vehicle form a vehicle record, the journey-related variables for a journey form a journey record, and the goods-related variables for a certain load of goods form a goods record. A vehicle record may be linked to several journey records (journeys made by the vehicle within the survey period, normally one week within the quarter), but each journey record is linked to a single vehicle. Similarly, a journey record may be linked to several goods records (several different transport operation may be effected in the same journey), but each goods record is linked to a single journey (see figure 1 above for an illustration).

2.3. Released variables

Reporting countries provide Eurostat with quarterly micro-data on vehicles selected for the sample, journeys carried out by these vehicles and goods transported during these journeys between regions.

List of reporting countries:

EU Member States	EFTA countries	Third country
Belgium (BE)	Liechtenstein (LI)(2)	United Kingdom (UK)
Bulgaria (BG)	Norway (NO)	
Czech Republic (CZ)	Switzerland (CH)	
Denmark (DK)		
Germany (DE)		
Estonia (EE)		
Greece (GR)		
Spain (ES)		
France (FR)		
Ireland (IE)		
Italy (IT)		
Croatia (HR)		
Cyprus (CY)		
Latvia (LV)		
Lithuania (LT)		
Luxembourg (LU)		
Hungary (HU)		
Malta (MT) (1)		
Netherlands (NL)		
Austria (AT)		
Poland (PL)		
Portugal (PT)		
Romania (RO)		
Slovenia (SI)		
Slovak Republic (SK)		
Finland (FI)		
Sweden (SE)		

- (1) See paragraph 3 of article 1 of Regulation (EU) N° 70/2012 of the European Parliament and of the Council of 18 January 2012 on statistical returns in respect of the carriage of goods by road (recast)
- (2) See Decision of the EEA Joint Committee No 175/2013 of 8 October 2013

The data availability tables per country are consultable under this link:
https://circabc.europa.eu/ui/group/0c7a12bf-2645-4509-9339-a266f3e1e44d/library/09be2be6-defd-4cb4-818c-134ded0107ca?p=1

In the tables below, the symbol "-" in the column "Anonymisation rules" means that the variable is released as in the original micro-data, without alteration. For a detailed description of the RC variables, see documentation on the RC data [3].

2.3.1. Vehicle-related variables (A1)

Variable	Description	*Anonymisation rules*
RCount	Reporting country.	*Dropped.*
Year	Year of the dataset.	-
Quarter	Quarter of the dataset.	-
QuestN	Questionnaire identifier. This identifies the vehicle record within the dataset.	*As some countries use common vehicle identifiers, the deletion of the variable RCount implied some non-unique identifiers (indeed, RCount x QuestN was unique but not QuestN alone). Then, QuestN was replaced by a random (unique) number.*
A1_1	(optional) Possibility of using vehicle for combined transport. The values are {Y - Yes; N - No}.	*Dropped.*
A1_3	Age of the road motor vehicle (lorry or road tractor). Years from first registration (numeric).	*Top-coded so that ages beyond 15 are replaced by ">15".*
A1_6	(optional) NACE Revision 2 activity class of the vehicle operator. This activity is coded at NACE 4-digit level.	*Dropped.*
A1_8_1	Total vehicle-kilometres during the survey period - loaded. Number of kilometres travelled by the vehicle in the survey period (loaded).	-
A1_8_2	(optional) Total vehicle-kilometres during the survey period - empty. Number of kilometres travelled by the vehicle in the survey period without load.	-
A1_9	Vehicle weighting to be used to get full results from individual data, if the collection of data is carried out from a sample. This is a grossing-up factor.	-
Stratum	Stratum identifier of sample in which vehicle appears.	*Dropped.*
A2_link	Number of linked A2 records. Number of journey records linked to this vehicle record.	-

2.3.2. Journey-related variables (A2)

Variable	Description	*Anonymisation rules*
RCount	Reporting country.	*Dropped.*
Year	Year of the dataset.	-
Quarter	Quarter of the dataset.	-
QuestN	Questionnaire identifier. This identifies the vehicle record within the dataset.	*As some countries use common vehicle identifiers, the deletion of the variable RCount implied some non-unique identifiers (indeed, RCount x QuestN was unique but not QuestN alone). Then, QuestN was replaced by a random (unique) number.*
JournN	Journey identifier. This identifies the journey within the journeys performed by the same vehicle.	-
A1_2	(optional) Axle configuration. Although the axle configuration is a vehicle property, the configuration can change for each journey. Hence, it is included in the journey-related variables.	*Dropped.*
A1_4	Maximum permissible laden weight. Multiples of 100 kg. This may change from journey to journey (depending on destination, etc.).	-
A1_5	Load capacity. Multiples of 100 kg. This may change from journey to journey.	-
A1_7	Type of transport. This variable specifies whether the journey is for hire or reward or on own account. This variable might seem connected to be vehicle (vehicle for hire or own vehicle), but in fact it may change for each journey: some journeys may be for hire and some on own account. The values are {1-Hire or reward; 2-Own account; 9-Not recorded}.	-
A2_1	Type of journey. Laden journey involving one single basic transport operation; or laden journey involving several transport operations, but not considered as a collection or distribution round; laden journey of the collection or distribution round type; unladen journey. The values are {1-Laden journey involving one single basic transport operation; 2-Laden journey involving several transport operations, but not considered as a collection or distribution round; 3-Laden journey of the collection or distribution round type; 4-Unladen journeys}.	-

A2_2	Weight of goods. Gross weight in 100kg.	-
A2_3	Place of loading (of the goods road motor vehicle, for a laden journey). For Member States, this is coded using level 3 of the Nomenclature of Territorial Units for Statistics (NUTS). For EEA countries not in the EU, this is coded using a list of administrative regions supplied by the country. For third countries, the 2- alpha ISO-3166 codes are used.	*(1) Recoded from NUTS3 to NUTS2.* *(2) Value of A2.3 was replaced by "SUPP" when A2.3=A3.5 AND FREQ(A3.1xA3.5)<3.*
A2_4	Place of unloading (of the goods road motor vehicle, for a laden journey). Same coding as A2.3.	*(1) Recoded from NUTS3 to NUTS2.* *(2) Value of A2.4 was replaced by "SUPP" when A2.4=A3.6 AND FREQ(A3.1xA3.6)<3.*
A2_5	Distance travelled: actual distance excluding the distance covered by the goods road motor vehicle while being transported by another means of transport. Numeric in km.	-
A2_6	Tonne.km effected during the journey. Numeric.	-
A2_7	Countries crossed in transit (not more than 5; if more than 5, the country reports the 2 first and the 3 last crossed countries). These are countries different from the country of loading and the country of unloading. Country names are encoded as two alpha characters (NUTS top level), hence this variable is encoded as 10 alpha characters.	*Suppression of small countries from A2.7 for those journey records in which at least one of the countries crossed is very small (Andorra, Liechtenstein, Malta, Monaco, San Marino, and Vatican City). The aim is to avoid publishing trajectories with too much accuracy. Indeed, stops in small countries would provide accurate information on track routes.*
A2_8	(optional) Place of loading, if any, of the goods road motor vehicle on another means of transport. Same coding as A2.3.	*Dropped.*
A2_9	(optional) Place of unloading, if any, of the goods road motor vehicle from another means of transport. Same coding as A2.3.	*Dropped.*
A2.10	(optional) Degree of loading of vehicle (in terms of maximum volume or space used during the journey). The values are {0-Unladen journey; 1-Not fully loaded (less than 90%); 2-Fully loaded (at least 90%); 9- Unknown}.	-
A3_link	Number of linked A3 records. Number of goods records linked to this journey record.	-

2.3.3. Goods-related variables (A3)

Variable	Description	*Anonymisation rules*
RCount	Reporting country.	*Dropped.*
Year	Year of the dataset.	-
Quarter	Quarter of the dataset.	-
QuestN	Questionnaire identifier. This identifies the vehicle record within the dataset.	*As some countries use common vehicle identifiers, the deletion of the variable RCount implied some non-unique identifiers (indeed, RCount x QuestN was unique but not QuestN alone). Then, QuestN was replaced by a random (unique) number.*
JournN	Journey identifier. This identifies the journey within the journeys performed by the same vehicle.	-
GoodsN	Goods operation identifier. This identifies the goods operation within the journey.	-
A3_1	Type of goods, according to the groups referring to an appropriate classification. Coding according to the division level (the highest one) of the NST 2007 classification (see Annex E of Regulation 1304/2007 [9, 13]). This classification consists of 20 flat categories numbered from 01 to 20, describing the type of goods.	-
A3_2	Weight of goods. Gross weight in multiples of 100 kg.	-
A3_3	Classification of dangerous goods. Coding according to the classification in Annex E of Regulation 1172/98. This classification consists of 13 categories of dangerous goods, labelled 1, 2, 3, 4.1, 4.2, 4.3, 5.1, 5.2, 6.1, 6.2, 7, 8, 9.	*Recoded to a Boolean variable: 0 if the good is not classified as dangerous and 1 if it is classified as dangerous.*
A3_4	(optional) Type of cargo. Coding according to the classification in Annex F of Regulation 1172/98. This classification consists of 10 categories labelled from 0 to 9 describing the packaging of goods, if any. For example, 0 stands for "Liquid bulk goods (no cargo unit)", 1 stands for "Solid bulk goods (no cargo unit)", 2 stands for "Large freight containers", etc.	-
A3_5	Place of loading of the goods. Same coding as A2.3.	*(1) Recoded from NUTS3 to NUTS2.* *(2) Value of A3.5 was replaced by "SUPP" when FREQ(A3.1xA3.5)<3.*
A3_6	Place of unloading of the goods. Same coding as A2.3.	*(1) Recoded from NUTS3 to NUTS2.* *(2) Value of A3.6 was replaced by "SUPP" when FREQ(A3.1xA3.6)<3.*

A3 7	Distance travelled, actual distance excluding the distance covered with the goods road motor vehicle while being transported by another means of transport. Numeric, in km.	

2.4 Data use restrictions

The anonymised data sets may contain confidential information or information the use of which is restricted. The user of the data must therefore comply with restrictions related to dissemination of tables described in the Regulation 6/2003 of December 2002, Art. 3.

2.5 Use of software for opening the anonymised files

As the anonymised files contain millions of records, specialised software that can cope with such high number of records (lines in a spreadsheet) should be used for opening the comma separated value (csv) files. Otherwise, e.g. using MS Excel, lines will be truncated and records for several quarters may seem to be missing from the anonymised data files.

References

[1] J. Domingo-Ferrer, *Anonymisation of Road Carriage (RC) Microdata (Proposed Methodology, version 2.1),* extended report submitted to Eurostat, February 2013.

[2] J. Domingo-Ferrer, *Anonymisation of Road Carriage (RC) Microdata (Summary description of the methodology, version 2.1)*, summary report submitted to Eurostat, February 2013.

[3] Eurostat, *Road Freight Transport Methodology: Reference Manual for the implementation of Council Regulation No 70:2012 on statistics on the carriage of goods by road,* Eurostat Manuals and Guidelines, 2016 edition.