A review of international road transport database files with risk exposure data

João L. Cardoso
Laboratório Nacional de Engenharia Civil
Lisboa, Portugal
joao.cardoso@lnec.pt

George Yannis, Eleonora Papadimitriou
National Technical University of Athens
Athens, Greece
geyannis@central.ntua.gr, nopapadi@central.ntua.gr

***Abstract*— The objective of this paper is to present the results of a critical assessment review of international databases which include risk exposure data (vehicle and passenger kilometres, vehicle fleet, road length, population etc.). In particular, the data collection and analysis as well as the availability and quality of the data were investigated. This work was carried out within the scope of the *SafetyNet* project of the 6th Framework Program for Research, Technological Development and Demonstration of the European Union.**

Existing surveys administered to the international databases file administrators were analysed. Personal interviews with international database provider representatives were carried out in order to collect additional information on the existing risk exposure data, the procedures for its collection, registration and accessibility. The outcomes of this analysis and interviews are presented and discussed in this paper. Selected available risk exposure figures by country and by year are compared amongst the mentioned data files. An overall assessment of the potential for international comparisons was also carried out.

The results show that international databases are useful sources of risk exposure data. However, the availability and quality of the existing data, as well as the objectives and scope of its collection vary significantly amongst databases. It is suggested that the analysed databases may be used in a complementary way.

Significant differences exist in the published figures amongst data files. This is the case for the "most complex" risk exposure measures (e.g. vehicle and passenger kilometres). Therefore, particular caution is recommended when using the information available in road safety analyses.

***Index Terms*— Risk exposure data; international data files; availability; comparability.**

I. INTRODUCTION

Considerable efforts have been made since the early sixties towards setting up International Data Files (IDF) containing detailed data on road accidents and general transport system factors (mainly exposure) for different countries that may be used to evaluate accident risks and to compare the safety performance of different countries and regions.

The interest in international and inter-regional comparisons is not limited to the benchmark of safety performance (namely expressed as the number of accidents or victims divided by a suitable measure of exposure). From a national and regional point of view, provided that the appropriate level of disaggregation is available, these comparisons make it easier to identify less performing areas and overall safety issues; also, they make available a benchmark for what has already been achieved, and therefore sensible targets may be set. Another important aspect of international accident data files is the possibility to get some hindsight to the peculiarities of different national road systems that may affect the international transferability of national best practices and guide their adaptation to other states and regions [1].

With this background, a the review and critical assessment of the current and future potential of the IDF containing risk exposure data (RED) was carried out in the context of ***SafetyNet***, a project of the 6th Framework Program for Research, Technological Development and Demonstration of the European Union [2]. In this paper the main results of the comparative analysis of RED contained in IDF are presented, on the basis of information collected from contacts and interviews to the persons responsible for the IDF, as well as the related publications. The IDF examined are:

- Eurostat
- European Conference of Ministers of Transport (ECMT)
- United Nations Economic Commission for Europe (UNECE)
- International Road Traffic and Accident Database (IRTAD)
- International Road Federation (IRF)

The risk exposure indicators examined are:

- Road length
- Vehicle kilometres
- Person kilometres
- Vehicle fleet
- Population

The discussion is mainly concerned with the following main issues:

- Data availability,
- Definitions used,
- Variables and values considered,
- Collection methods
- Data quality control.

The brief presentation of the data files, is followed by the combined analysis of the selected RED and finally by the synthesis of the current and future potential of the IDF with RED.

II. DESCRIPTION OF THE ANALYSED IDF

A. Eurostat

The EUROSTAT (http://epp.eurostat.cec.eu.int) publishes since 1990 an annual publication, with an overview of

transport and energy statistics for the EU Member States [3]. The objective is to provide the EU with high quality standardized data on transport. Data is collected by means of the common EUROSTAT-UNECE-ECMT questionnaire.

Quality control of published data is ensured by the Members States through their official data providers. Therefore, no data quality control is carried out and no correction factors are applied, as the Member States have to comply with the common definitions.

All aggregate data is freely available on the internet (http://europa.eu.int/comm/dgs/energy_transport/figures/pocketbook/).

B. ECMT

The ECMT (www.cemt.org) publishes accident statistics since 1975. Between 1975 and 1984 these statistics were included in the Transport Statistics Yearbook; since 1985 accident statistics are presented in a separate publication: the annual Road Accident Statistics Yearbook [4]. These publications are intended for supporting political decision-making concerning European transport policies. The ECMT road accident data file and the transport statistics database contain data on accidents, victims and exposure that provide road accident related indicators (especially rates).

C. UNECE

The United Nations publishes since 1955, through UNECE (www.unece.org), an annual publication containing statistics on the road traffic system activity in Europe and North America [5]. Data on accidents and victims are presented, with data on road length, traffic volumes, number of registered vehicles and population. There are 55 countries in the UNECE data file.

The accident data in the database is concerned with injury accidents only. The data is collected from replies by member countries to the Eurostat-ECMT-UNECE common questionnaire and from official national and international sources.

D. IRTAD

IRTAD (http://www.bast.de/htdocs/fachthemen/irtad) was established by the Steering Committee of the OECD Road Transport Research Programme, to provide a framework for the integrated collection of international aggregated data on accidents, injuries and exposure on a continuous basis.

This database is mainly research-oriented and its development was adjusted to the following objectives: scientific cooperation; collection of harmonized and timely aggregate accident and exposure data; improvement of data available for research and policy planning; harmonization of definitions; and identification of special safety issues deserving further research [6].

IRTAD was established in 1989. Annual aggregated data are collected for every year since 1970, on several safety related issues, namely accidents, casualties, exposure and safety belt wearing rates. Currently, IRTAD has 50 member institutes and data are collected for all OECD countries, except Mexico and Slovenia (29 countries). Data are collected continuously, using electronic forms. Access (on-line or through diskette) to the database is only possible for members of IRTAD; however, a brief overview is available to the public on the internet.

Quality control of input data is performed, especially in what concerns recorded definitions and mathematical correctness. Corrective factors are applied to data that does not comply with the IRTAD standardized definitions.

E. IRF

The IRF (www.irfnet.org) is a non-governmental, not-for-profit international organization established in 1948 to promote development and maintenance of better and safe roads and road networks. Members include both private and public organizations, including some government agencies, from several countries worldwide.

Development of the IRF database started in 1958, and the first data tables were first published in 1964, concerning 20 countries. Data are collected annually, using paper and electronic forms. Aggregated data for 84 countries (up to year 2004) are presented in the 2006 data tables [7]. On-line access to the data is provided to IRF members only.

No validation is performed on the provided data, since these are national official data. However, when needed, national representatives provide corrections to data previously sent.

III. COMPARATIVE ANALYSIS OF THE COLLECTED RED

A. Road length

As regards the collection of data on road length, the common questionnaire for EUROSTAT, UNECE and ECMT divides the roads in two major classes: motorways and other roads. The “other roads” class is further divided in three administrative classes, resulting in a total of four road classes. There is no specific reference to type of road environment (urban or rural), as a classification criteria.

Tables published by EUROSTAT and UNECE present the road length for each of the mentioned four road classes [3]-[5]. Published tables by ECMT do not address specifically road length [5]. However, an additional road data disaggregation is used in the common questionnaire specifically in what concerns accidents and victims, which indicates that additional data is being collected. Accident data are divided in four road classes [5]:

- Motorways;
- Roads in built-up area;
- Roads outside built-up area;
- unknown type of road.

The IRF database contains information on road length by class of road (Motorway; Main Highways or National Roads; Secondary or Regional Roads; and Other Roads), type of operation (public or private), type of surface (paved and unpaved) and condition (good, fair or poor). The IRF database includes data on road density (in km per km^2), as well.

The IRTAD database contains data on road length, according to four road classes: motorways; A-level rural roads; other rural roads; and roads inside urban areas. A-level roads are the primary national road network [8].

In summary, EUROSTAT, IRTAD and IRF consider four road classes in their published tables. IRTAD road classes differentiate urban roads from rural roads. Similarities can be found between the road classifications of IRF and EUROSTAT/ECMT/UNECE, as both do not take in account the type of road environment (urban or rural) and include operational (motorway/non motorway) and administrative (main/secondary) criteria.

A comparison of 2000 and 2001 data from EUROSTAT and IRF is presented in Figures 1 to 2. Comparison is presented by means of ratios, where the denominator is EUROSTAT data and the numerator is the other IDF data. No major differences between the main highway length data were detected for the 25 EU countries. The only detected case is probably due to a typing error in the IRF database. In what concerns the data on motorway length, several cases of significant differences were identified (10 cases). The differences vary between -7% and +28%.

EUROSTAT and UNECE motorway length data show considerable agreement, except for one single country. Considerable disagreement is observed in main and secondary highway length data. This, in part, is due to differences in terminology, concerning the concepts of main, secondary, state, provincial and local roads.

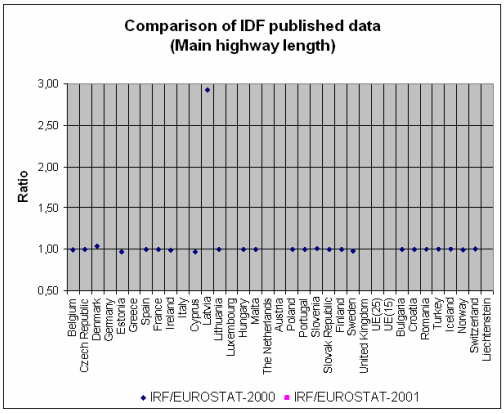

Fig. 1 - Comparison of main highway length data in EUROSTAT and IRF

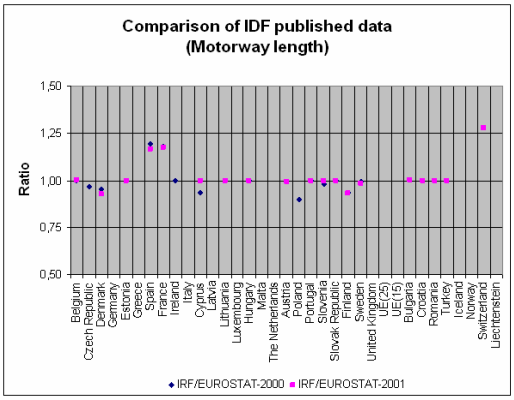

Fig. 2 - Comparison of motorway length data in EUROSTAT and IRF

B. Vehicle kilometres

As regards the collection of data on travelled distance, the common questionnaire for EUROSTAT, UNECE and ECMT is designed to collect data on four vehicle classes (motorcycles, passenger cars, buses, lorries and road tractors), irrespective of the road class. However, published UNECE tables with yearly data provide travelled distance for five vehicle classes: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; and lorries and road tractors. In the last available publication from UNECE, data for 31 countries are provided. However, several countries do not provide data for two wheeled vehicles.

Yearly data for travelled distance are provided in the IRF database for four vehicle classes: motorcycles and mopeds; passenger cars; buses and coaches; and lorries and vans (IRF).

The IRTAD database contains data on travelled distance, according to the four road classes and to the six vehicle types considered: mopeds and mofas (mopeds with maximum speed of 30 km/h); motorcycles and scooters; passenger cars and station wagons; goods motor vehicles; buses; and other motor vehicles.

In summary, availability and disaggregation of travelled distance by road vehicles vary among the analysed IDF. UNECE, ECMT and IRF have information disaggregated by vehicle class; IRTAD has information disaggregated by road class and vehicle class; the most recent EUROSTAT tables do not contain any information regarding vehicle×kilometres travelled.

A comparison between 2000 and 2001 data from EUROSTAT, IRF, ECMT and IRTAD is presented in Figure 3. Considerable differences are detected: +/-80%.

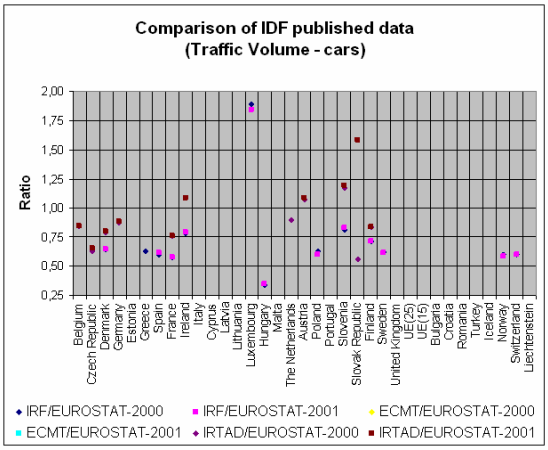

Fig. 3 - Comparison of distance travelled by passenger cars in EUROSTAT, IRF, ECMT and IRTAD

C. Person kilometres

As regards the collection of data on transport activity, the common questionnaire for EUROSTAT, UNECE and ECMT is designed to collect data on passenger travelled distance for three vehicle classes (motorcycles, passenger cars and buses) and on goods haulage distance for lorries and tractors. No disaggregation by road class is provided.

UNECE tables do provide data on passenger travelled distance by vehicle type (motorcycles, passenger cars and buses) and goods haulage distance by type of traffic - national, international, loaded and unloaded, cross trade and road cabotage (UNECE).

Yearly data for passenger kilometres are provided in the IRF database for road based private and public transport. Concerning the amount of surface goods transport, the IRF data tables contain the total transported tonnage (in tons) and the total hauled road distance (in tons×km).

IRTAD contains information about passenger kilometres for passenger cars and public transportation by year, but not for all countries.

In summary, as observed in the previous section, availability and disaggregation of passenger travelled distance and of goods haulage distance by road vehicles vary among the analysed IDF. EUROSTAT, ECMT and IRF have information disaggregated by vehicle class, even though the classes do not overlap completely; UNECE and IRTAD tables do not contain any information regarding ton×kilometres travelled; UNECE does not have data on passenger×kilometres travelled.

A comparison of 2000 and 2001 data from EUROSTAT, IRF, UNECE and ECMT is presented in Figures 4 and 5. Availability of this data in EUROSTAT is good, for passenger cars and for buses; the same cannot be said for two wheeled vehicles. On the IRF and UNECE data bases these data are missing for several countries. No ECMT data for 2001 was available at the time of this research.

In what concerns the distance travelled by passengers of private transport, no major differences were detected between the two IDF. Data on passenger distance travelled by means of private transport does not differ very much with the IDF (-10%, +12%), except for a pair of cases that appear to be caused by wrong data input. Data on passenger distance travelled by means of public transport show large variation, according to the originating IDF; there seems to be considerable differences in the way the number of passenger×kilometres is considered in each IDF, in spite of the fact that the same data form is used by EUROSTAT and UNECE to collect the data.

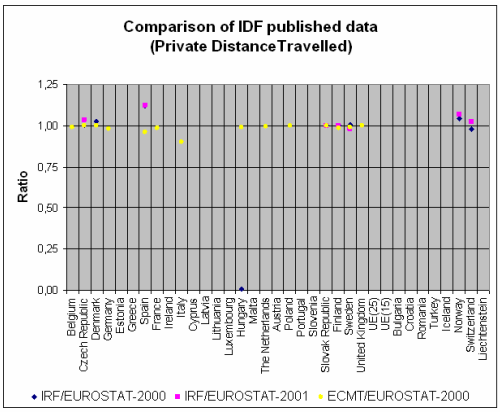

Fig. 4 - Comparison of distance travelled on private vehicles in EUROSTAT, IRF and ECMT

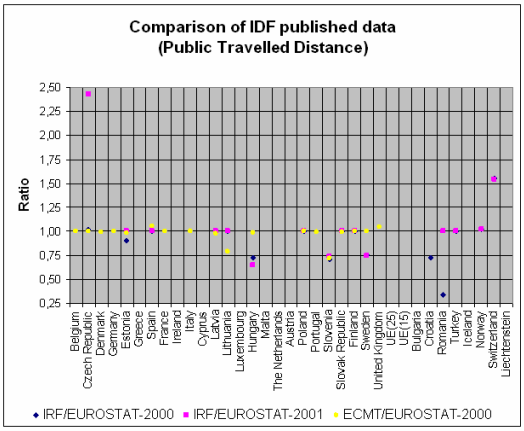

Fig. 5 - Comparison of distance travelled on public vehicles in EUROSTAT, IRF and ECMT

D. Vehicle fleet

As regards the collection of data on the number of registered vehicles, the common questionnaire for EUROSTAT, UNECE and ECMT has nine road vehicle classes: mopeds; motorcycles; passenger cars; motor coaches, buses and trolley buses; trams; lorries; road tractors; semi-trailers; and trailers.

Motorcycles are further divided in two classes according to the engine size. Passenger cars are further divided by age (four classes), type of energy and engine size (10 classes), unloaded weight (four classes). Buses are divided by age (four classes) and type of energy (six classes). Lorries are divided by age (four classes), load capacity (eight classes), motor energy (six classes) and kind of transport (two classes). Road tractors are divided by age (four classes), type of energy (three classes) and kind of transport (two classes). Semi-trailers are divided by load capacity (five classes) and kind of transport (two classes). Trailers are divided by load capacity (five classes) and kind of transport (two classes).

However, yearly data tables for registered vehicles produced by UNECE database allow for eight vehicle classes, only: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; lorries; road tractors; semi-trailers; and trailers. In the last available publication, data for 43 countries are provided. Some countries do not provide data for mopeds.

The EUROSTAT tables for the yearly data on registered vehicles comprise five vehicle classes: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; and goods vehicles.

The ECMT tables contain five vehicle classes: mopeds; motorcycles; passenger cars; buses, coaches and trolley buses; lorries and road tractors.

The yearly number of vehicles is provided in the IRF database for five vehicle classes: passenger cars; buses and coaches; lorries and vans; road tractors; and motorcycles and mopeds.

Yearly IRTAD tables provide vehicle registration data. Six vehicle types are considered: mopeds and mofas (mopeds with maximum speed of 30 km/h); motorcycles and scooters; passenger cars and station wagons; goods motor vehicles; buses; and other motor vehicles.

In summary, data availability on the number of registered vehicles is good in the analysed IDF. Two wheeled vehicles are separated from the rest of vehicles in all databases. Most IDF separate cars from buses, and both these vehicle classes from lorries and from road tractors. The only exception is the IRTAD database, which considers only goods vehicles (aggregating both lorries and road tractors in the same category).

A comparison between 2000 and 2001 data from EUROSTAT, IRF and UNECE is presented in Figures 6 to 7. Availability of this data is good in all IDF, for passenger cars, buses and lorries. As regards two wheeled vehicles, only IRF and UNECE do provide data for a considerable number of countries; for these vehicles it was possible to collected data on two countries in EUROSTAT.

Differences between the number of cars in each IDF are minor (largely within the +/-5% band). The registered numbers of buses, trucks and lorries present differences that vary considerably. Only very few cases of comparable numbers of two wheeled vehicles were detected and they are not presented here.

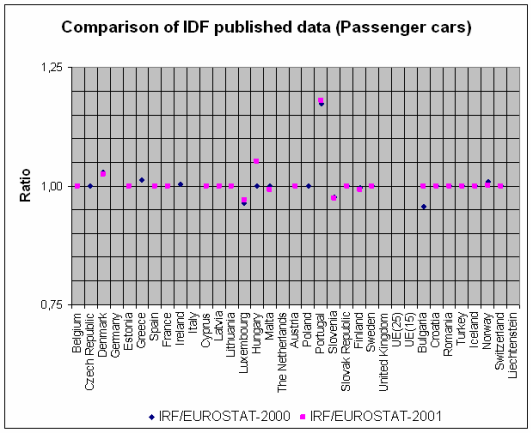

Fig. 6 - Comparison of number of cars in EUROSTAT and IRF

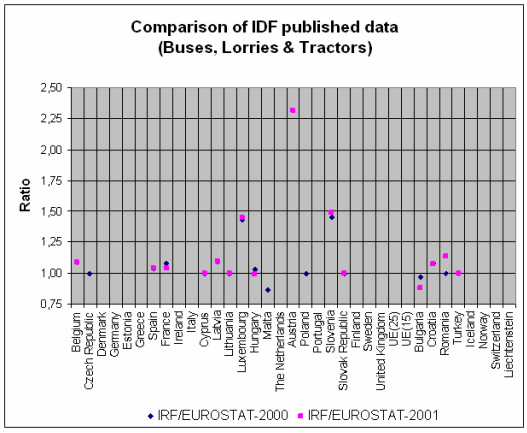

Fig. 7 - Comparison of number of buses, lorries and tractors in EUROSTAT and IRF

E. Population

The UNECE/EUROSTAT/ECMT databases contain data on resident population, according to eight age classes: less than 6 years old; 6 to 9 years; 10 to 14 years; 15 to 17 years; 18 to 20 years; 21 to 24 years; 25 to 64 years; and 65 or more years old.

In the last available UNECE publication, data for 55 countries are provided. Overall there are no missing data; however, some countries do not provide the data in agreement with the standardized age group classification.

The IRF database does contain information on each country's total population, for the years since 1994. No disaggregation of population by age group is provided.

In IRTAD, population data is divided in twenty age groups: 0 to 5 years old; 6 to 9; 10 to 14; 15 years old; 16; 17; 18; 19; 20; 21 to 24; 25 to 64; 25 to 34; 35 to 44; 45 to 54; 55 to 59; 60 to 64; 65 to 69; 70 to 74; 75 to 79; and 80 years or more.

In summary, the importance of population as an overall accident risk indicator at the national level is recognized in all analysed IDF. Nevertheless, disaggregation of published data by age group varies with the considered IDF: IRF and EUROSTAT tables do not provide classification by age group. In the case of EUROSTAT, however, use of other EUROSTAT statistical tables, not directly related with transport, may overcome the absence of this information.

A comparison of 2000 and 2001 data from EUROSTAT, ECMT, IRATD and IRF showed no major differences in the figures published.

IV. Synthesis

In the framework of the present analysis, a detailed comparison of RED published by each IDF was carried out, in terms of availability and quality, and several interesting results and conclusions were obtained:

- Exposure data available in each IDF are in a much more aggregate form than the RED collected at national level,
- Accordingly, the more disaggregate national exposure data are not exploited within the context of IDF,
- Significant differences are observed among IDF in the published figures for each exposure measure; these differences are more important for the more "sophisticated" exposure measures (i.e. vehicle and passenger kilometres),
- These differences are partly due to the different national sources and definitions used,
- However, another reason may concern insufficient data quality control within each IDF.

Some of the analysed IDF use common definitions and there is, to some extend, overlapping in the collected data and the corresponding published tables. This indicates that there is scope for combining the data collection procedures in a common questionnaire. That has been already achieved to a great extent with the EUROSTAT-ECMT-UNECE common questionnaire.

Important RED are not collected in some IDF: that is the case for fuel sales (which may be used to estimate the amount of vehicle kilometres) and, especially the number of active driving licenses in each country. In addition, not all relevant

disaggregated secondary variables are collected by all countries: this is the case for two wheeled vehicles, especially mopeds and bicycles, for which few countries consistently provide data.

As regards quality control, IRF acknowledges that they do rely on the quality control systems used by their data providers, as in most cases they are using official data. UNECE does not have internal quality control. EUROSTAT and ECMT have some routines for internally checking the data provided, especially in what concerns the coherence between partial and total values and with the values published on other databases. IRTAD checks the correctness of received data and, especially with new members, may resort to follow up actions to ensure correct use of agreed definitions.

The presented comparison of two years' data from EUROSTAT, UNECE, ECMT, IRTAD and IRF, highlighted the fact that differences in definitions may exist as regards some disaggregated basic variables such as motorway length, heavy vehicle and two wheeled vehicle fleets, and the distance travelled by public transport users.

Table 1 summarizes the road safety risk rates published by the various IDF, by combining accident and fatality data with the available RED.

Table 1 – Risk indicators in international data files

Risk indicator		International data file (IDF)				
		EUROSTAT	ECMT	UNECE	IRTAD	IRF
Accidents per inhabitant					●	
Accidents per vehicle-.km	General				●	●
	Build-up				●	
	Road class				●	
Fatalities per inhabitants	General	●	●	●	●	
	Age group	●		●	●	
	Age group and sex	●		●		
Fatalities per vehicles			●		●	
Fatalities per road user by type			●		●	
Fatalities per vehicle×km	General				●	●
	I/O build-up area				●	
	By road class				●	
Injured per inhabitants	General	●			●	
	Age group	●			●	
	Age group and sex	●				
Injuries per licensed drivers						
Injuries per vehicles					●	
Injuries per vehicle×km	General				●	
	vehicle×km build-up area				●	●
	By road class				●	

Summarizing, these data files are useful and accessible aggregate data sources, resulting from several decades of important data collection efforts. However, they have different objectives; they collect diverse data in different forms and structure, in some cases by different national sources, and are maintained by organizations with different scopes and policies.

It is interesting to notice that the RED available in each IDF are in a much more aggregate form that the RED collected at national level, as reported by the countries. Additionally, it is not always known whether IDF receive more (disaggregate) data than they publish.

It should be noted that data availability in different IDF does not always imply comparability. Apart from the intrinsic comparability issues due to the national collection methods, as discussed above, other issues may further compromise the comparability of RED [2]. In the framework of the present analysis, it was demonstrated that differences in the published exposure estimates are observed among IDF, these differences being more significant for the more "sophisticated" exposure measures (i.e. vehicle and passenger kilometres).

These differences may be attributed to the fact that some of the exposure estimates in the IDF may be based on crude national estimates, whereas the actual data source may not always be known. Additionally, another reason may concern insufficient data quality control, which may be either not carried out at all, or limited to the correction of only obvious mistakes by checking the totals and comparing with other IDF.

Despite the limitations discussed above, the considerable effort made during the last decades for gathering and exploitation of road safety related data is clearly reflected in these IDF. The fact that there are various IDF for RED at European level is positive for the road accident statistics users, because they can choose from a variety of information. The objectives and scopes of these data files, as well as the quantity and quality of available data contained inside the IDF, differ among the various data providers, making them to function complementarily in most of the cases. Consequently, particular caution is required from the data users, in order to optimally use the available information in reliable road safety analyses.

V. ACKNOWLEDGMENT

This paper is based on the work carried out in the ***SafetyNet*** research project, partly financed by the European Commission.

REFERENCES

[1] National Technical University of Athens (1996). Overview of international road accident data files. DGT-EC.

[2] SafetyNet (2005). State of the Art report on risk and exposure data. NTUA, Athens.

[3] European Commission (2006). EU energy and transport in figures. EC, Luxembourg.

[4] ECMT (2006). Trends in the Transport sector – 1970-2004. ECMT, Paris.

[5] UNECE (2006). Annual bulletin of transport statistics for Europe and North America. UNECE, Geneva.

[6] Bruhning, E. (1995). Five years of IRTAD: Review and perspectives. Seminar on international road traffic and accident data bases. OECD, Helsinki, 11-13 September 1995.

[7] IRF (2006). World Road Statistics - 2005. IRF, Geneva.

[8] IRTAD (1998). Definitions and data availability. OECD - RTR Road Transport Research Programme. BASt Germany, June 1998.