SAU/D1	***29/6/07*** INTRAS Instituto de Tráfico y Seguridad Vial VNIVERSITAT DE VALÈNCIA
Report title	**Quality and representativity of the traffic accident data in urban areas State of the Art** ***Final***
Project title:	**SAU**: **S**istemas de análisis de **A**ccidentalidad **U**rbana. *(Urban Accident Analysis System)* (1/4/04-31/6/07).
Report authors	*Dr. Mauricio Chisvert* *Dña Elena López-de-Cózar* *Dña Mª Luisa Ballestar*
Directors of the project	*Dr. Jaime Sanmartín Arce* *Dr. Mauricio Chisvert Perales*
Project type:	Project co-financed by the Directorate-General Energy and Transport (TREN-03-ST-S07.30828) Transport

The current report is part of the “SAU: Sistemas de análisis de accidentalidad urbana (Urban Accident Analysis System)” project as an intermediate report in the framework of the Project co-financed by the Transport and Energy General Directorate (TREN-03-ST-S07.30828) and developed by INTRAS (Institute of Traffic and Road Safety of the University of Valencia).

INDEX

1. Backgrounds

The reduction of the number of road accident fatalities by 50 %, by the year 2010, suggested by the EU, involves the active contribution of all the agents in charge of the road safety in Europe. Even though the accidents that happened in urban areas have a relative smaller severity, it is the place where, for the moment, in absolute terms, the major number of accidents take place in the EU countries, as well as generating serious consequences on the more vulnerable users (pedestrians, cyclists, children, the elderly…).

The current action has as main objective the creation, validation, discussion and spreading, at European level, of the 'best practices' for the collection, processing and analysis of **t**raffic **a**ccident (TA) data in urban areas. The foreseen final result fundamentally consists in the disposal of a European guide of advices or of "best practices" in order to implement / improve the traffic accident collection, analysis and monitoring systems in urban areas.

For that, a compilation of the current "best practices" and on the exchange of experiences between municipalities from several EU countries will be counted on, added to the practical pilot experience that will be carried out as part of this project in several Spanish cities. With the spreading of this guide, the purpose is to contribute to the development of local tools in order to help giving answers and solutions, with more reliable and accurate knowledge, to the problematic of the accident rate in each municipality.

The concrete actions that are developed in the project are the following ones:

1) Bibliographical revision and summary of the "state of the art" on the problem of underreporting, the quality, management and analysis / exploitation of TA data in urban area.
2) Development of an in-depth "case study" and application and evaluation of the best practices in some Spanish municipalities from different sizes.

3) Execution of a survey study with the objective of getting an approximation to the current situation and practice from a representative sample of European cities from different sizes.

4) Organization of a workshop where the results will be discussed, after the fulfilment of the previously exposed objectives.

From all the information obtained from the previous stages, writing and spreading of the Guide of Best Practices throughout the EU.

The document that is delivered hereafter is the Deliverable I: Quality and representativity of the traffic accident data in urban area: State of the art.

In this stage, as it has been pointed out in the project statement, it is expected, through a bibliographical and documentary revision and summary, to get a general perspective of the state of the accident rate urban records in terms of data collection and information quality, as well as the rules in force in relation to the accident rate collection systems at the European level.

The results obtained in this first stage mean a starting point on which the next stages of the SAU study will be developed.

2. General observations

Of all the transport modes, road transport is the more dangerous and the one that claims more human lives. However, road safety is a complex scope of analysis and intervention.

The researches is this field necessarily require the access to all the elements that intervene in traffic, mainly the traffic accident, the mobility and the behaviour of the different users (Fontaine et al, 2003).

Therefore, the accident and trauma databases are considered as one of the essential tools to objectively assess road safety (COM, 2003). This data represent a tool to control and detect problems, to spot priority action areas and to assess the effectiveness of the established intervention measures. That is the reason why most of the developed countries have implemented different accident data collection systems mainly based on the information supplied by police investigation.

The reflection on databases has to be considered in the development of the road safety policy. Several studies have stressed the existence of important difficulties to reduce the number of accidents, and to achieve it, it is necessary to better know their characteristics, their causes and their consequences. So there is the need to develop specific research software on accident rate, in which the access to statistics and data exploitation is provided.

A useful and effective accident rate analysis requires, on the one hand, quality data – accident data as well as risk exposure data (population, km, trips, vehicle fleets, etc.) – and on the other hand, an appropriate analysis methodology.

In the following sections we will focus on the urban accident data collection and management, as well as on the problems that these ones, at present, are showing in the international scope and more particularly in the EU countries scope.

3. Some international directives

In the EU countries, the urban area traffic accident rate involves, as average, more than half of the total of traffic accidents and victims. A great part of this accident rate is concentrated is the large cities, representing a first magnitude problem both from the point of view of public health and the socio-economic cost, and the perspective of the transport management and urban mobility.

In spite of this data, the management of the urban road safety has not been enough considered in the general road safety policies. This situation has been changing for the last years. This way, for example, one of the recommendations of the European Parliament (Session of the 18th of January 2001) urges the European Commission to introduce and stress on the importance of urban road safety management in the road safety programmes of the European Union (specially the measures appointed to create a safe environment for the vulnerable users). Likewise, in the White Paper – European transport policy for 2010 (COM, 2001), it has been focussed on the objective to fundamentally improve the current accident data collection systems in the European Union.

The importance of the accident data is shown up in the fact that the study, the evaluation, the improvement and the optimization of the collection systems represent a priority objective in the international road safety programmes and policies.

This way, the **E**uropean **T**ransport **S**afety **C**ouncil (ETSC), in the document *A Strategic Road Safety Plan for European Union* (ETSC, 1997), points out how one of the key elements of the strategic plan is the strengthening and improvement of the accident data records. The ETSC puts the emphasis on the setting out of database collection systems at the European Union level, example from which are set up the CARE database (*Community data bank on road traffic accidents in Europe)*, and its later updates, CAREPLUS (1996), CAREPLUS 2 (2000) and CADaS (now in development), managed by the DG TREN.

Nevertheless, even though the CARE project makes reference to an international accident database at the European Union level, this database is settled down in the national collection systems, which are also based on the regional and local systems. This way, it has been underlined the necessity to strengthen and improve those records, for their importance at local level as well as at regional, national and international level.

According to us, if what is really expected is to improve the accident record quality, we have to work and act in the same basis of the process, i.e. "at the local level", where the data is collected and generated in the framework of the police interventions. This is the main work philosophy of the SAU project.

On the other hand, in the 2001 report, the European Transport Safety Council (ETSC, 2001) emphasizes on the interest of a coordinated and connected approach between the different databases that the countries have. This kind of approach is already carried out in the American system (US National Automotive Sampling System – NASS) with four levels of collection: the Fatality Analysis Reporting System – FARS, the General Estimates System, the Crashworthiness Data System and the Crash Injury Research and Engineering Network. Moreover, it insists on the main problems of the road safety data collection, on the underreporting of some accidents[1] (for example, slight accidents with a single vehicle) and on the lack of risk exposure data. It recommends creating an information system that is based on the connection between the police and the hospital data, and promotes the introduction of a European information system on road safety open to the public through the web.

From an applied perspective, the important actions are the ones that have been developed for several years since the Transport Research Programme of the OECD (OECD-RTR, *Organisation for Economic Cooperation and Development - Road Transport and Intermodal Linkages Research Programme*). These actions are aimed at strengthening the traffic accident data collection systems, both at international and national, regional and local level.

[1] Later on this problem will be specifically treated.

Among them, we stress on the creation and ongoing update of one of the databases that collects the main traffic accident aggregate data of the Member countries: IRTAD (*International Road Traffic and Accident Database*). Around IRTAD, a series of working and studying expert groups have been set. One of their priority objectives is to define assessment criterions and strategies to improve the traffic accident data quality, which imply the improvement of the data collection and management procedures.

The European Commission too has specifically dealt with the accident record subject. In the assessment of the status of *Promoting Road Safety in the EU The programme for 1997-2001* (COM (97) 131 final) it has been included, in the maximum priority action group, the strengthening of the "*road safety monitoring systems in the EU*". One of the key elements of these monitoring systems is constituted by the accident data collection systems.

The measures and recommendations in this study area are collected in the *Progress Report and Ranking of Actions about EU Road Safety* (COM (2000) 125 final).

On the other hand, in the European Road Safety Action programme for 2003-2010 (COM (2003) 311 final), the European Commission raises a series of measures like the strengthening of road controls, the improvement of the road infrastructure and actions tending to improve users behaviour.

In this programme, major structural deficiencies have been noticed in the traffic accident records at the EU level, and different improvement action lines have been emphasized. In this context, the SafetyNet project launched in May 2004, in the Sixth Framework Programme and having a 4 years length, represents an action with a huge scope of which wider objective is to define and create, from a macro perspective, the needed elements to set up a European Road Safety Observatory (ERSO). In order to achieve that, they have mainly worked in the achievement of essential improvements of the current accident rate information systems, favouring the establishment of information standards that allow a greater homogenisation of the future information about accident rate at the EU level.

In Spain, the Traffic General Directorate (DGT), through its Key Strategic Action Plan for 2005-2008, has joined the European Road Safety Action

Programme proposals, by raising its objectives, strategies and work lines. Both programmes include as basic actions the traffic accident data collection, analysis and dissemination.

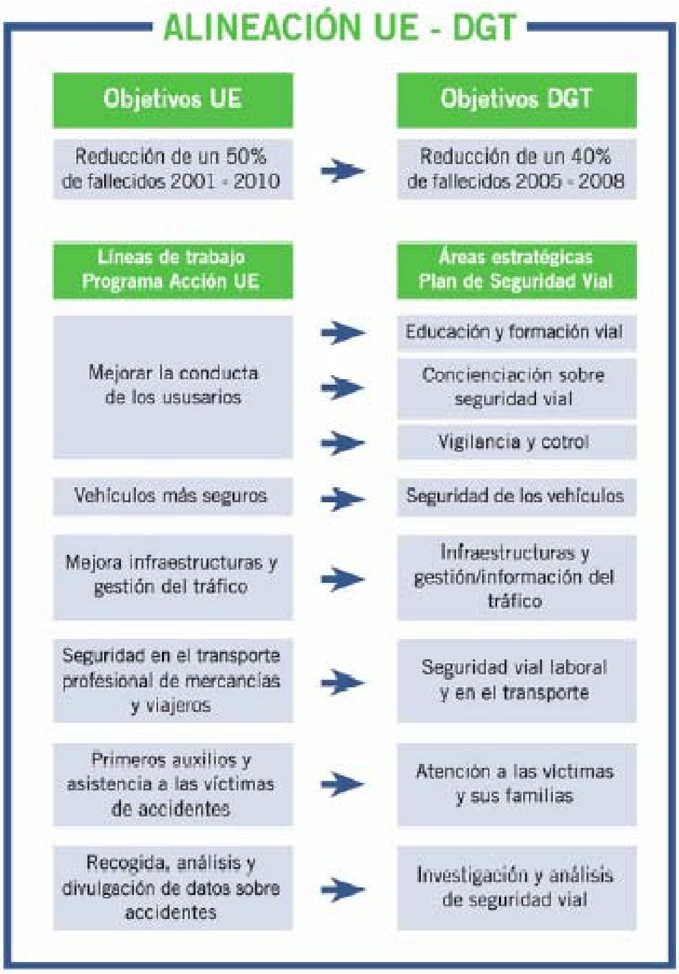


Figure 1: Key Strategic Action Plan for 2005-2008. Traffic General Directorate

All these initiatives from international organizations, related to the study and the establishment of priorities, recommendations and directives in relation with traffic accident data collection systems, will turn out to be more or less effective while they will be properly passed on to the different countries and, fundamentally, to the different administrations in charge – in situ – of the data collection and management. As a last resort each national administration and the different regional and local administrations are the ones in charge of the data collection, management and analysis process, on which all the rest is supported.

In this sense, there are some noticeable differences between the different EU countries as far as the traffic accident data collection systems are concerned. Even more, in each country it is possible to find differences according to the application scope of the record. Most likely, the more noticeable difference in many countries is the one that exists between the motorway traffic accident data records (more homogeneous and standardized) and the urban area ones. For these last ones, there is a greater variability regarding the record exhaustiveness and quality as well as the collected information contents, and the organizational and technical processes used for this task.

In spite of these differences, common elements are predominating in the different collection systems and, consequently, many problems are also common despite the corresponding national, regional and local nuances.

In the next sections a summary of the state of the traffic accident data collection systems are presented, stressing on the special situation of the urban accident rate.

Aiming to have a first approximation with regard to the traffic accident data collection practices and procedures in the European Union, hereafter several general key aspects in different EU countries will be revised:

- ✓ Traffic accident definitions (p. 17)
- ✓ When the traffic accident data is collected for the purposes of statistics (p. 18)
- ✓ Who collects the traffic accident data (p. 26)
- ✓ Data collection tool. (p. 27)
- ✓ European databases (p. 36)
- ✓ National databases (p. 46)

Part of the information gathered in the following sections has been taken out from the reports carried out in the DUMAS WP4 project (1998) and in Chisvert (2000a, 2000b) and from the results obtained from a previous European survey carried out in the SAU project framework[2]. Given the ongoing development of the procedures in the different countries, some of the data available in the mentioned documents have been modified and updated.

[2] In the Deliverable III of this SAU project, the results of this European survey will be described in detail, even though, in this report, some information is brought forward.

3.1. Definitions

The first aspect that has to be taken into account is the definition of the traffic accident, given that - depending on if it is wider or more restrictive - it is going to have influence on the accident rate numbers that are handled, and it will make easy or difficult the data comparison between the different countries.

In practice, the definition from which most of the countries start is the one set by the WHO during the Vienna Convention. This way, the traffic accidents are defined as being the ones that "*happen or start in a street or road open to traffic, with one or more injured person or fatality as a consequence of the accident, and in which there is at least one vehicle in motion. This includes the collisions between vehicles as well as between vehicles and pedestrians, and vehicles and animals or any fixed object. The single accidents, for which there is only one person involved, are also included*".

In its statistics glossary[3], the transport statistics working group (IWG, 2003) defines the traffic accident as "the accident in which there is at least one vehicle in motion involved, in a public road, or in a private one open to the public, in which there is at least one fatality or injured person".

From these reference definitions and keeping the essential elements in most of the cases, national, regional and local criterion diversity and conceptualisations have been developed. This causes divergences between the traffic accident records of several countries, and even between records of different administrations in the same country or region.

3.2. When the traffic accident data is collected

A key aspect is the difference between what is conceptually considered as a traffic accident and which operative criterions are applied – in practice – to decide which accidents are likely to be considered in the official statistical records. For example, the Spanish normative adopts the WHO definition, including as well the damage-only accidents. However, for practical reasons and

[3] The transport statistics glossary was first published in 1994 with the aim of helping the Member countries during transport data collection. It was developed by the ECMT and Eurostat through a common questionnaire.

means availability, the Traffic General Directorate (DGT) only takes into account, for the official statistical records, the accidents with casualties.

In the Table 1 there is a synthetic comparison of the national specifications relating to which accidents are likely to be considered in the traffic accident statistical records in several countries. Commonly, these specifications directly depend on the official traffic accident definitions set in each country and, as we have already mentioned, on some practical criterion type, usually restrictive, regarding which accidents have to be collected for the statistics. It is important to point out that the data is only orientative, given that it can also present regional or local variations, as well as variations through time.

	PUB. R.	MOT.	VICTIM.	OTHER VEH.	SINGLE	DAMAGE
Austria	*	*				
Belgium	*	*				
Denmark	*	*				
Finland	*	*	Slight ones are not included			
France	*	*	*			
Germany	*	*				
Greece	*	*	*			
Italy	*	'Driving'	*			
Norway	*	*				
Portugal	*	*				
Spain	*	*	*			
Sweden	*	*				
The Netherlands	*	*	Slight ones are not included			
United Kingdom	*		*			
	In shady, criterions that are applied. * Compulsory criterion.					

Table 1. Normative criterions to set which traffic accidents have to be included in the statistical records. In shady, the ones that are applied. With the asterisk*, the requisites that have to be compulsorily fulfilled in order to record the accident.
PUB. R.: the accident takes place on the public road; MOT.: vehicle in motion involved; OTHER VEH.: if the accidents in which motor vehicles are not involved (e.g. bicycles) are taken into account; SINGLE: if the single accidents are taken into account. DAMAGE: damage-only accidents.

As we may notice, the two compulsory criterions (with an asterisk in the table) present in most of the countries are:

(1) The accident takes place on a public road and

(2) There is one vehicle in motion involved.

Likewise, in many countries, it is specified that, moreover, there has to be some victim (injured or fatality) in the mentioned accident. This is the case for **Spain**, **France**, **Greece** and **Italy**.

However, in other countries, the severity of the injuries is limited in order to be considered in the official accident record. For example, in **the Netherlands** the accidents considered as very slight are not collected even though the criterions to apply in order to set what exactly is "a slight injury" are not clearly specified. Something similar happens in **Finland**. In the section 3.2.1 (p. 19), the criterions used to define victim severity in different countries are described in detail.

Regarding the type of involved vehicle, there have been some modifications along with time. In all the analysed countries the accidents that are taken into account are the ones where both motor vehicles and non-motor vehicles (e.g. bicycles) are involved. Moreover, all the countries exclude the accidents with casualties and the damage-only accidents where no vehicle in motion is involved (e.g. pedestrian accidents).

As far as the damage-only accidents are concerned (only collected in some countries), the criterions have been changing along with time. This way, for example, in **Germany**, until 1994, the damage-only accidents were collected only when the cause was an offence or when the damages were serious. Since then the criterion is at least one tow-away vehicle as a result of the accident. In other countries like the **Czech Republic**, the police draw up an accident report for damage-only accidents only when the damage is serious, exceeding a certain cost. In **Norway** damage-only accidents were collected up to 1964, in **Belgium** up to 1973, in **Italy** up to 1991, in **Austria** up to 1994 and in **France** up to 2004.

Finally, we want to point out the differences regarding the inclusion of certain "special" accidents. This way, some countries like **Belgium**, **Germany**, **Denmark**, **The Netherlands**, **Austria** and **Portugal** exclude suicide as a traffic accident. **France** extends this limitation by excluding the voluntary actions (voluntary manslaughters, suicides…), and the natural disasters. However, **Spain**, **Italy** and **Luxembourg** do include suicides in their police records (CARE, 2006). On the other hand, in **Hungary**, they have to specify that the accident has to be "fortuitous or unintentional". Finally, all the countries except **Luxembourg** and **Sweden** rule out the natural deaths.

In the **United Kingdom**, there has been an important change in the normative that define the criterions to consider traffic accidents in the statistical records. We want to stress on some of the aspects that differentiate them from the rest of the countries (STATS20, 2005).

The Road Traffic Act of 1988 stated that "*(…) the accidents that have to be communicated to the police are the fatal ones or the ones with injured persons, that take place in public roads, in which at least one motor vehicle is involved and in which, besides, no information exchanges happen between the insurance companies and the involved persons (…)*".

However, nowadays, the criterions have considerably changed.

On the one hand, as for the vehicle type, any involved vehicle is considered, being motorised or non-motorised (including bicycles, ridden horses or vehicles pulled by horses, among others). Moreover, this one has not to be necessarily in motion (e.g. the pedestrian accidents where the pedestrian injures himself with a parked vehicle, or bus passengers that injure themselves while getting in or out of the stopped vehicle). The damage-only accidents and the pedestrians injured without the involvement of any vehicle (e.g. falls on the pavement) are not considered.

On the other hand, as for the delay to notify the accident officially, they have set a limitation that consists in the accident notification within 30 days after it has happened.

Regarding the road type, the definition of public road is extended to the following conditions: Streets or motorways, lanes or special zones (e.g. pedestrian zones with access available for motor vehicles, bus lane, bicycle lane…), public areas with limited access for motor vehicles (cycle or rural paths, or pathways with access available for motor vehicles), and level crossings (only when no train is involved in the accident).

Finally, even though the criterion that there are injured persons and fatalities is maintained, some special cases are excluded: injured persons as a result of a disease immediately prior to the accident, natural deaths or deaths not related with the accident blows, suicides, or witnesses in state of shock but that are not involved.

It is important to mention the fact that the number and type of collected accidents – and consequently the accident rate statistics – is also going to depend on the normative interpretation and application that, in practice, the persons in charge of the data collection will carry out, both at the organisational and individual level. A clear example is the one of a single bicycle accident. Some police officers probably do not consider these accidents as necessarily recordable, not even if the consequences are serious, if they happen in roads open to motor vehicles or fulfil the criterions set to be collected.

As a conclusion, the type and number of accidents collected by the police institutions are influenced, among other factors, by: (1) the traffic accident definition assumed by the national/regional/local administration; (2) The normative criterions regarding which accidents have to be collected to be statistically managed and treated and (3) the real practice carried out by the persons in charge of the records, which are mainly fixed by different aspects like the availability of enough means to carry out the collection task systematically, the internal normative, the motivation and training of the agents, etc.

3.2.1. How is set the severity of the casualties?

In the Table 1 we may notice that the presence of injuries or casualties is a key aspect to set which accidents are statistically recorded. Nevertheless, terms commonly used in the accident statistics like "fatality", "victim", "slight victim" and "serious victim" are differently applied in several countries.

This way, for example, in **Spain**, a serious victim is the injured person that needs to be hospitalized for, at least, one day. In other countries, the hospitalization is not needed to talk about serious traffic accident casualties, by using criterions strictly based on the type and severity of the injuries.

There is not any internationally accepted classification of the severity levels. The classification of the World Health Organization set up in the Vienna Convention has an orientation purpose, but turns out to be quite complex when applied, and as a consequence, few countries are using it in practice. This definition states what follows (OMS, 1968):

"It is considered that a person has been hurt or injured because of a traffic accident when he/she did not die as a consequence of it but shows slight or serious wounds or injuries. The serious injuries are considered to be the fractures, the concussions, the internal injuries, the lacerations or serious cuts, the states of shock that need medical treatment in general and any other injury that involves an hospitalization. The slight injuries are considered to be the secondary injuries like sprains, bruises and scratches."

The transport statistics working group considers as a traffic accident casualty "every single person that dies or is injured in a traffic accident. A traffic accident fatality is every single person that dies in the accident or within the 30 days following it." (IWG, 2003)

In practice, as we already mentioned before, the different countries have adopted their own definitions even though they share many common elements. In the Table 2 the definitions of several OECD countries are gathered. Just like it happened for the traffic accident definitions, the information has an orientation purpose and may present regional and local variations, or time variations.

COUNTRY	SLIGHT INJURY	SERIOUS INJURY	FATALITY	CORRECTION FACTOR
Germany	All the injuries not considered as being serious	In-patient (> 24 hours) or death after 30 days.	3 days (<1978) 30 days (>1977)	1
Austria	Injury that, not being serious, requires medical treatment.	Injury that implies a health deterioration and a work disability (> 24 hours), in-patient (>7 days) or death after 30 days.	3 days (<1992) 30 days (>1991)	1.12 (<1992) 1 (>1991)
Belgium	All the injuries not considered as being serious	In-patient (> 24 hours) or death after 30 days.	30 days	1
Cyprus	All the injuries not considered as being serious	Disability (> 25 hours)		
Denmark	All the injuries not considered as being serious, excluding the slightest ones	According to the type and the severity of the injury (similar to the WHO). Death after 30 days.	30 days	1
Slovenia	Slight injuries	In-patient (> 24 hours)		
Spain	All the injuries not considered as being serious	In-patient (> 24 hours) or death after 30 days.	24 hours*	*
Finland	No difference between slight and serious injuries		30 days	1
France	Injury that requires medical treatment or hospitalization of less than 6 days	In-patient for more than 6 days	6 days (<2005) 30 days (>2004)	1.09 (<1994) 1.057 (1994-2004) 1 (>2004)
United Kingdom	All the injured persons not considered as being serious, and the ones that are not injured but require a treatment for being in state of shock	In-patient (> 24 hours), serious injuries out-patients or death after 30 days.	30 days	1
Greece	Injury that, not being serious, requires medical treatment.	In-patient (> 24 hours) or death after 30 days. According to the type and severity of the injury (similar to the WHO).	24 hours (<1996) 30 days (>1995)	1.15 (<1996) 1 (>1995)
The Netherlands	All the injuries not considered as being serious (the slightest ones are excluded)	In-patient (> 24 hours) or death after 30 days.	30 days	1

* In Spain, the fatalities within 30 days are established by the DGT by applying a correction factor that is calculated through the periodic follow up of a sample of injured persons (see the more detailed data in the following pages).

COUNTRY	SLIGHT INJURY	SERIOUS INJURY	FATALITY	CORRECTION FACTOR
Hungary	All the injuries not considered as being serious	In-patient or injuries that imply a treatment for more than 8 days	30 days	1
Italy	No difference between slight and serious injuries		7 days (<1999) 30 days (>1998)	1.08 (<1999) 1 (>1998)
Latvia	All the injuries not considered as being serious	In-patient (> 24 hours)		
Luxembourg	All the injuries not considered as being serious, that only need out-patient treatment	In-patient (> 24 hours)	30 days	1
Malta	Slight injuries	Serious injuries		
Norway	All the injuries not considered as being serious and that do not need to be hospitalized	In-patient (> 24 hours), injuries that imply serious or permanent disability, or death after 30 days.	30 days	1
Poland	In-patient for less than 7 days or slight injuries	In-patient (> 7 days) or serious injuries		
Portugal	Injury that, not being serious, requires medical treatment.	In-patient (> 24 hours)	24 hours	1.3 (<1998) 1.14 (>1997)
Sweden	All the injuries not considered as being serious	In-patient	30 days	1

Table 2. Definition of traffic accident slight/serious injury and fatality in several countries. The CORRECTION FACTOR refers to the one that is applied in the international statistics to the countries that do not supply 30 days data. In shady (unknown data for this country). Source: Care Glossary (2006), WP-1 SafetyNet (2006) and statistical documentation of each country.

Generally, all the European countries are adopting the definition suggested by the European Union with the aim of harmonizing criterions.

In this line, during the Council of Ministers on the 13th of October 2004, the president of the French Republic requested that the traffic accident casualty census be even more accurate. In order to give an answer to this request, a new definition of the injury severity was proposed and came into force on the 1st of January 2005. From that date on, in **France**, accident *fatality* refers to "the victim that dies at the accident or within the 30 following days as a consequence of the accident" (previously they considered as a fatality the victim that died within the first 6 days after the accident). *Seriously injured person* is "the injured person that has been an in-patient for more than 6 days". *Slightly injured person* is "the injured person that has been an in-patient for less than 6 days or has needed medical care". *Unhurt* is "the person involved in the accident and that has not needed any medical care".

The applied definition diversity raises the need to set up common criterions that harmonize the data and allow comparisons. In this sense, to the countries that do not apply the death within 30 days definition, a correction factor that depends on the criterion applied in each case is applied. This was the case of **France** until 2004 (+5,7%), of **Italy** until 1998 (+7,8%), of **Greece** until 1995 (+18%) or of **Austria** (+12% only in 1991). The only countries that are still collecting the death within 24 hours are **Portugal** (+14%) and **Spain**.

In **Spain**, the DGT periodically carries out studies from which are derived a series of disaggregated correction factors by user type and scope of the accident occurrence (motorway or urban area). These factors are applied to the number of deaths within 24 hours, which is the time that the Spanish police take now (Table 3).

Up to 1992	K(30)=K*1.3	
1993-1996:	Outside urban area	Inside urban area
Driver	K(30)=K+SI*2.46%	K(30)=K+SI*1.02%
Passenger	K(30)=K+SI*2.29%	K(30)=K+SI*0.94%
Pedestrian	K(30)=K+SI*7.22%	K(30)=K+SI*3.87%
1997-2000:	Outside urban area	Inside urban area
Driver	K(30)=K+SI*2.44%	K(30)=K+SI*1.93%
Passenger	K(30)=K+SI*2.17%	K(30)=K+SI*1.80%
Pedestrian	K(30)=K+SI*4.76%	K(30)=K+SI*5.71%
2001 onwards:	Outside urban area	Inside urban area
Driver	K(30)=K+SI*2.41%	K(30)=K+SI*2.17%
Passenger	K(30)=K+SI*2.24%	K(30)=K+SI*2.15%
Pedestrian	K(30)=K+SI*6.17%	K(30)=K+SI*4.34%

Table 3. Correction coefficient applied in Spain to estimate the deaths within 30 days, according to the area and the user type. K= number of person killed; SI= number of persons seriously injured

This type of "universal" criterions poses many problems, given the existing differences between the countries and the regions in relation to the health system and the emergency services. For example, in a study carried out in Catalonia with 1986 data a correction factor of +20,2% was set facing the -30% applied at that time by the international organizations (Costa and Arnau, 1989).

In other cases, like in **Finland**, the fatality data collected by the police is complementary to the death cause records (based on the death certificates), so it is estimated that the traffic death rate data collected in the official statistics have a coverage of 100% (Statistics Finland, 2006).

The problem is more complex in the case of the injured persons. It is easy to control that, even though there is a certain agreement in accepting the hospitalization criterion as defining the condition of seriously injured, there is a high heterogeneity of definitions. For example, **The Netherlands** and **Denmark** do not take into account a good part of the injuries defined as being slight and, consequently recordable for the WHO. Something similar happens in **Finland** where, since 1978, the less serious injuries (bruises, scratches…) that did not need medical treatment were left out of the official statistics. In **Luxembourg**, the difference between seriously injured persons and slightly injured persons is fixed from several criterions like the place of the injury, the hospitalization duration, and the duration of the sick leave or incapacity to work.

In view of the lack of criterions applied internationally, the **OECD** sets a definition to be applied to the IRTAD data and that is based on the hospitalization criterion. This way, are considered as serious the traffic accident casualties that needed hospitalization (> 24 hours). In this case, the greater problem is to get reliable information in the police records, given the recurrent difficulties that the police has to get hospital information.

3.3. Who collects the traffic accident data

In all the countries considered, the police are the ones responsible to collect and record the statistical accident data. Nevertheless, there are differences in relation with, for example, the police force in charge of this task, or regarding the existence or not of specialized traffic forces and/or units. Hereafter the situation in several European countries is described synthetically and as an example.

In **Spain** the road accidents are assumed by the Traffic Guardia Civil – specialized forces - , or the autonomous polices in the case of Catalonia and the

Basque Country. In the case of accident in urban area the local polices are generally responsible of the actions in relation with the accident and its data collection. In this case, there might be, or not, specialized traffic units, according to, usually, the size of the city and the available means.

In **France**, the National Police, dependent on the Ministry of the Interior, is responsible for collection accident information in the large cities, while the Gendarmerie, dependent on the Ministry of Defence, is responsible to intervene in the accidents in smaller cities and in the motorways. So, the same police force collects motorway and urban accidents.

In **Italy**, the accident data is collected by three different police forces: "Polizia Stradale" (specialized), "Carabinieri" and the Municipal Police. In this case, unlike what happens in Spain, the local authorities are not generally responsible to collect traffic accident data and they rarely carry out this task.

In **the Netherlands**, there is a police force specialized in motorway traffic at the national level, and is responsible of the main network. The other areas (secondary roads and urban areas) are autonomously organized in each of the 25 regions that divide the country.

In **Germany**, as well as the police in charge of investigating the accidents in each region, in some large cities there are traffic accident specialized units (VUD Verkehrsunfalldiens).

In short, the usual situation is the coexistence of several police forces in charge of the accident data collection task, sometimes with different action scope and sometimes with the same one (e.g. **Italy**). In this context, it is useful to set standardized procedures to collect, manage and exploit or analyse the accident rate data.

With regard to this last point, most of the countries have set up standardized procedures to uniformly collect accident rate data, through the design and use of a statistical questionnaire or common accident form that has to be used with uniform criterions of completion. However, as for the accident rate management and analysis, the situation is quite insufficient, given that many countries have developed neither tools nor standardized procedures to manage and exploit accident data, approach that could be really useful, more particularly in the case of the local administrations.

3.4. Data collection tools

The **accident statistical questionnaire** is the structured protocol used by the different police forces for the standardized collection of traffic accident information, information obtained from the investigation or reconstruction of each particular accident. Generally, most of the countries that have developed traffic accident data collection systems have a common accident statistical questionnaire used by the different administrations or police forces in charge of collecting data, both in the motorway and in urban area. This allows a homogeneous data collection through standardized procedures, as far as the considered information type and the data collection and coding process are concerned.

In the Table 4, we can see the name given to some of the accident questionnaires officially used in each country.

Country	National statistical questionnaire
Belgium	Formulaire d'analyse des Accidents de la Circulation avec tués ou blessés (FAC)
Spain	Cuestionario estadístico de Accidentes con Víctimas
France	Bulletin d'Analyse d'Accident Corporel de la Circulation (BAAC)
Greece	Road Accident Data Collection Form - DOTA (ΔΕΛΤΙΟ ΟΔΙΚΟΥ ΤΡΟΧΑΙΟΥ ΑΤΥΧΗΜΑΤΟΣ)
Hungary	Questionnaire No. 1009: Road traffic accidents involving personal injury (Személysérüléses közúti közlekedési balesetek)
Italy	Incidenti Stradali
Luxembourg	Procès-verbaux
Norway	Road Accident Report Form (Rapport om veitrafikkuhell)
Poland	Karta zdarzenia drogowego
United Kingdom	STATS19 (Gran Bretaña)
Czech Republic	Statistic form on road accident registration (Statistický formulář pro registraci dopravních nehod)
Sweden	Informationsunderlag Vägtrafikolycka

Table 4: National statistical questionnaires to collect traffic accidents.

However, in many cases, this national questionnaire/protocol coexists with other questionnaires, generally at the local level. This would be the case of many town councils where both the official statistical report, at the national level, and local versions – used locally – that usually consider some particular aspects of the urban accident rate in detail or that might be particularly useful in other procedures in relation with the accident, like the police traffic accident reports or the technical reports.

In **Spain**, there is an official statistical questionnaire (knows as accident report), from the Traffic General Directorate (DGT). Sometimes, this accident report coexists with other ones from different administrations, at the local level, of which data is used by the administration that collects the data. Anyway, in the case of the accidents with casualties, filling in and sending the DGT questionnaire to the centralised body is always compulsory, independently from the fact that other questionnaires are filled in for a local purpose.

For example (Frantzeskakis et al, 2000), in the **United Kingdom**, some police forces do collect additional information for their own use, but because of difficulties in collecting this type of data, this in not a requirement of the STATS19 form. In **Greece**, except for the data collection form aimed at the creation of the national data file, the police forces fill-in two additional forms, one for their own purposes and one for the Ministry of Environment, Planning and Public Works. In **Austria**, the Ministry of Home Affairs runs an additional aggregated database on fatal accidents featuring accident causes and circumstances; each fatal accident is reported by the police to the Ministry via electronic link.

In **France**, the police forces electronically fill in the traffic accident analysis bulletin (BAAC) from the police reports (records that are collected for the judicial proceedings). This bulletin serves as a base for the traffic accident exhaustive epidemiological studies.

In **Luxembourg**, up to 1964 in addition to the reports, the police agents had to fill in a statistical bulletin for each accident and send it to STATEC (organization in charge of the statistical data management and exploitation). Since then, they have been sending a copy of the report (up to 2000 to STATEC and since then directly to the Ministry of Transports). Nowadays, the Ministry is dealing with the entry of the report in the national database and STATEC cooperates with the statistical exploitation.

Each national accident statistical questionnaire comes together with the corresponding instructions for its completion and, less frequently, for transmission of the data (electronic or paper) to the national record or to the database.

For example, in the **United Kingdom**, the national questionnaire, known as STATS19, comes with a Completion manual called STATS20 (Dft, 2005), while the STATS21 manual describes how the data has to be managed and validated before being sent to the national record or database. In **Norway**, the *Rettledning til utfylling av rapport om veitrafikkuhell* (Instructions for filling out road traffic accident reports) is used as a standard for coding and revising the data. In **Spain**, there is the *Manual de Normas para cumplimentar el Cuestionario Estadístico de Accidentes de Circulación con Víctimas* (Manual of Rules to fill out the Statistical Questionnaire of Road accidents with casualties) (the latest version is from 1993).

In each country, the statistical questionnaire includes, at least, the basic variables concerning the accident. However, the values in which such variables are expressed present important differences from one country to another.

The basic accident data, *common* or shared by most of the countries' statistical questionnaires, may be grouped in three large blocks:

- *Accident information*
    - Time and place: Year, month, weekday, time and place.
    - Accident type and manoeuvre type.
    - Road and environmental circumstances: Road type, road category, weather conditions, light conditions, road surface type, road surface condition and traffic condition.
- *Vehicle*: Vehicle type, age and nationality.
- *Victims*: Age, gender, user type, alcohol consumption, use fo the seatbelt, position in the vehicle, driving license, nationality and injury severity.

In the Table 5, basic information regarding accident statistical questionnaires of several EU countries is illustratively shown. There are noteworthy differences among the accident statistical questionnaires of the different European countries as far as the number of variables, values and pages are concerned. This way, there are cases like Italy at one end, with 15 variables, and the Spanish one at the other end with 78 variables.

	N. of Variables	N. of Values	Values per variables	Pages
Germany	30	149	5	3
France	68	348	5.1	1
Italy	15	197	13.1	2
The Netherlands	38	148	3.9	2
Belgium	61	239	3.9	4
Luxembourg	26	153	5.9	2
United Kingdom	50	255	5.1	4
Ireland	29	171	5.9	1
Denmark	45	202	4.5	1
Greece	20	168	8.4	1
Spain	78	277	3.6	1
Poland		166		2
Portugal	22	82	3.7	3

Table 5. Some basic characteristics of the national accident reports in twelve EU countries (orientative data). Source: DUMAS project (WP-4) and own writing. The data refers to the situation in 1991, so there might have been some changes. In shady: unknown data for this country.

Even sharing the pointed information types, each national record has its own structure and characteristics (format, completion procedure, answering alternatives…) (Frantzeskakis et al, 2000).

In many countries there are at least two sub-files (one for the accidents and another one for the casualties), while in other countries more sub-files are also included (vehicles and/or road information). In **Norway**, for example, they use four different files (accidents, vehicles, involved persons and a textual description of the accident). In **Spain**, there are three (accidents, vehicles and involved persons).

On the other hand, in some countries like **United Kingdom** and **Denmark**, the national traffic accident databases are linked by computer with other external national databases (vehicle registration file, driver license file, road network record, etc.), giving more detailed and accurate information that has to be collected at the moment of the accident). In the case of **Sweden**, the accident data collection system is fed by the police data as well as the hospital data, so the information is integrated in a unique national record system (STRADA system).

The design of the accident statistical questionnaire, and the procedure set to fill it in, is a fundamental aspect that highly influences the quality of the collected data, especially if we take into account the complex and difficult circumstances in which the information collection procedure is carried out. So, aspects like the specification of the type of data that has to be collected, the easiness of its handling, the items clarity, its alternatives and the data coding, among others, are aspects that may influence the quality of the data collected finally, even though they are not always taken into account at the time of designing the accident statistical questionnaire.

It is very important to bear in mind that the accident statistical questionnaire only represents one among many other documents that are generated when traffic accidents occur – traffic accident report, technical reports, etc. – with which it share certain type of information, but in a more summed up and coded way, in a way that it allows its statistical treatment keeping the anonymity, thanks to prior entry of them in a computer database.

Undoubtedly, for many police officers, the task of filling in the statistical questionnaire is added to the accident investigation and to the different technical and legal documents that derived from such investigation, involving real work overloads sometimes. This implies that the task is perceived as unnecessary and even "annoying" by the persons in charge of carrying it out, given that it takes time and means to other more urgent, priority and immediate tasks than the accident management (traffic control, vehicle removals, attention to the victims…).

These problems are often the consequence of the inadequacy of the procedures in force that use to imply the achievement of several transcriptions of the same data, multiplying the work to be done.

This way, in practice, in the first place the main traffic accident data is collected on the spot, data that is usually used to create the report and the first procedures. This data uses to usually be descriptive narrations on the accident (first transcription). Then, the report information is used to fill in the accident report (second transcription), of which data is next entered in the database (third transcription). In the cases where there is also a differentiated local accident report, we have up to four transcription processes of the same information.

3.5. International databases

3.5.1. IRTAD (International Road Traffic and Accident Database)

The International Road Traffic and Accident Database (IRTAD) was established in 1989 by the Steering Committee of the OECD Road Transport Research Programme. The database was initially hosted by BASt, in Germany. Since 2004 the hosting has changed to the Joint OECD/ECMT Transport Research Centre and since 2006 the Centre is acting as information centralized server. Central to the operation, development and use of the database is the International Traffic Safety Data and Analysis Group (now made up of 53 public and private organizations from countries belonging or not to OECD/ECMT).

The IRTAD database is a very useful tool for international comparisons and it includes aggregated data on accidents, injuries, and relevant exposure data (population, car park, network length, vehicle kilometre,…) from some OCDE countries: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, United Kingdom, United States.

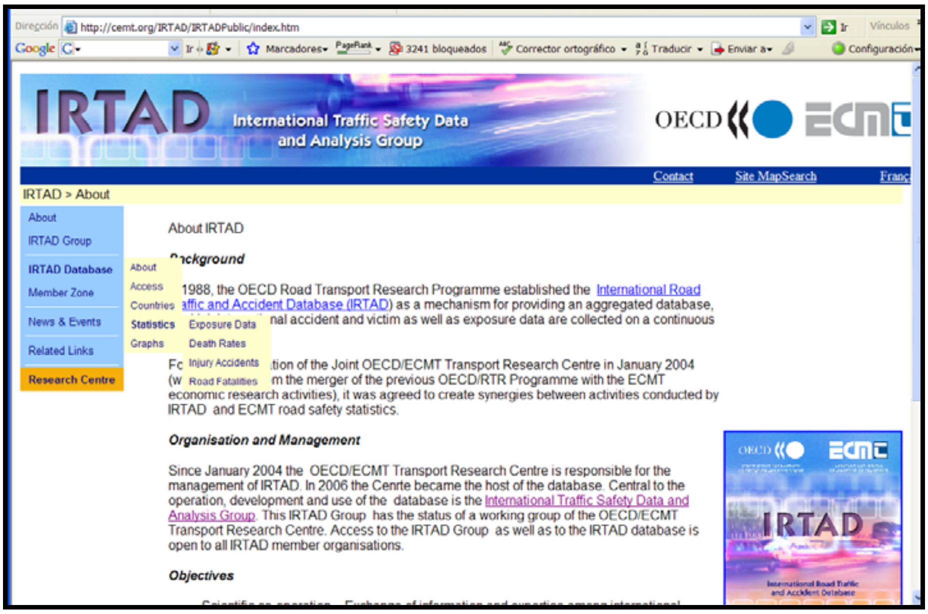


Figure 2: IRTAD webpage (**http://cemt.org/IRTAD/**)

3.5.2. IRF (International Road Federation)

The International Road Federation (IRF) is a non-governmental, non-profit organization. It was established in 1948 to promote development and maintenance of better and safe roads and road networks. Development of the IRF database started in 1958, and aggregate data for 84 countries worldwide are collected annually (up to year 2002). On-line access to the data is provided to IRF members only.

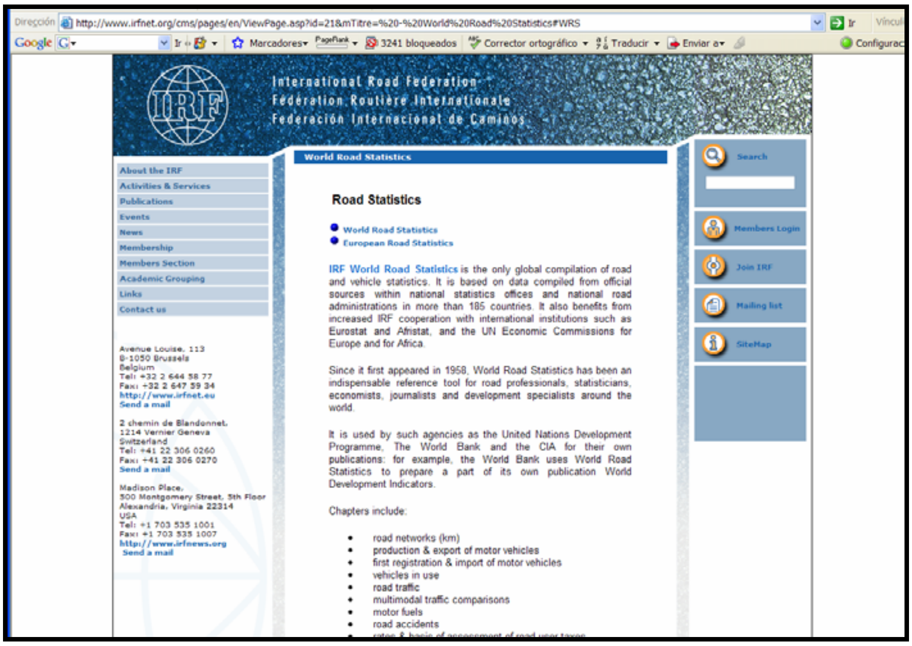


Figure 3: IRF webpage (**http://www.irfnet.org/**)

3.5.3. UNECE (United Nations Economic Commission for Europe)

The UNECE publishes since 1955 an annual publication containing statistics on the road traffic system activity in Europe and North America. Data on accidents and casualties are presented, with data on risk exposure like road length, traffic volumes, number of registered vehicles and population. There are 55 countries in the UNECE data file.

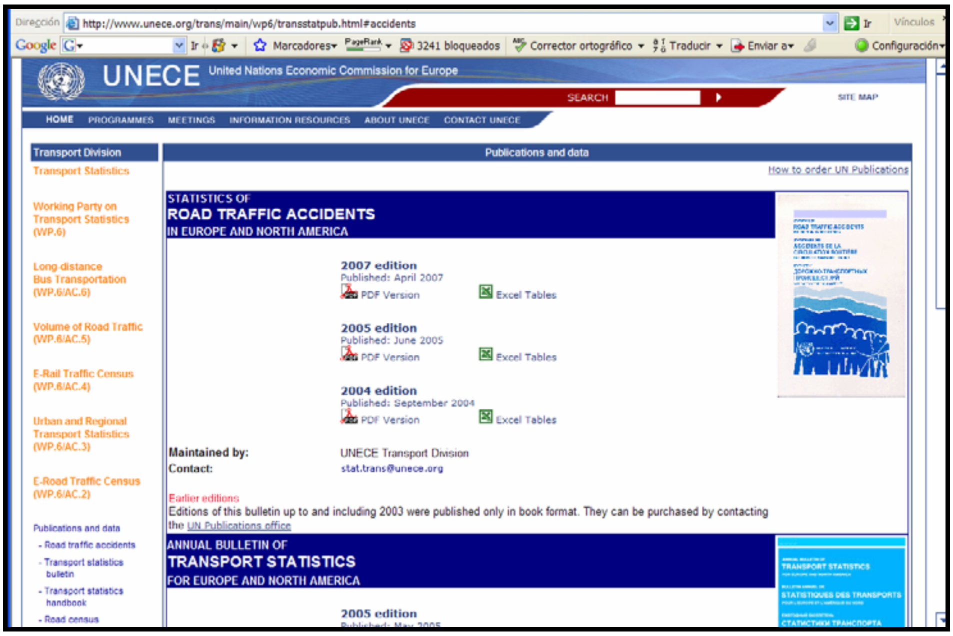


Figure 4: UNECE webpage (**http://www.unece.org/trans/welcome.html**)

3.6. European databases

At the European level, several road accident databases have been set up with the objective of evaluating road safety and define priorities for the development of actions in this field. This process of creating databases started in 2003 (ETSC, 2001) and refers to all transport modes.

Hereafter several features of the main European databases on road accident are described (for more details, please look up the European Road Safety Observatory webpage: http://www.erso.eu). Some are general, providing with an exhaustive record of accident rate in European countries and some others focus on specific accident groups (children, motorcycles or trucks) and on methodologies on obtaining specialized information (accident reconstructions or in-depth investigations).

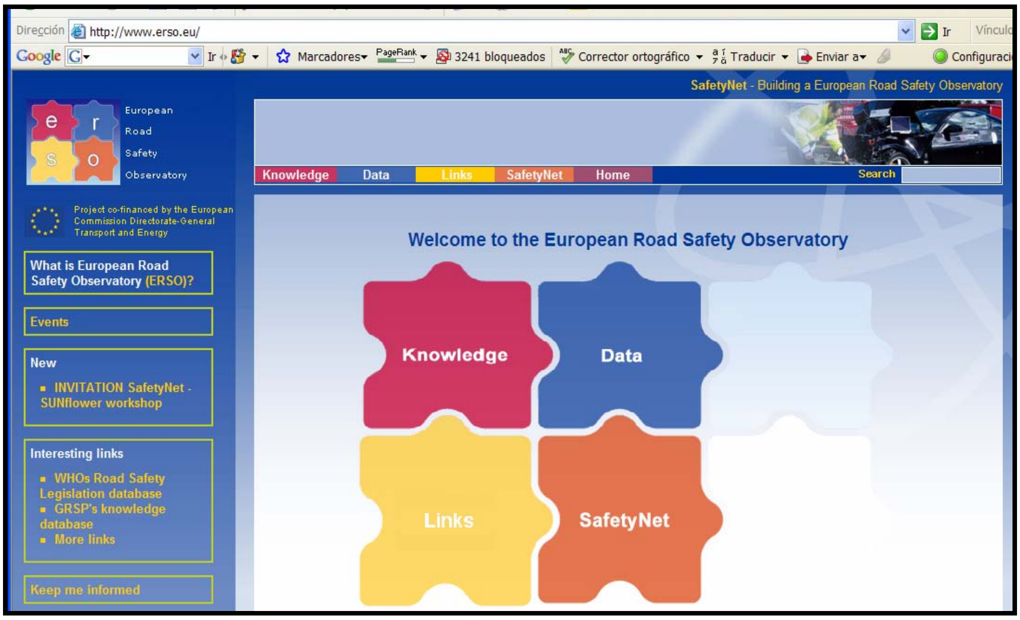


Figure 5: European Road Safety Observatory webpage (**http://www.erso.eu**)

3.6.1. CARE (Community database on Accidents on the Roads in Europe)

Database with disaggregated data on the whole road accidents with casualties that happened in the European Union. The 17 Member States annually send their data file for integration in the CARE database: Austria, Belgium, Denmark, Finland, France, Greece, Ireland, Italy, Luxembourg, The Netherlands, Portugal, Spain, Sweden, United Kingdom and, since 2005, Estonia, Poland and Hungary.

Detailed data is available only to the Directorate-General Energy and Transport (DG TREN) and few selected bodies of each Member State. Statistical information (aggregated) is available on the CARE website and studies based on CARE data are accessible through CARE or other websites such as SafetyNet: http://www.erso.eu).

The level of detail, the sub-record and the number of variables vary a lot from one country to another. The data files of each country are integrated in the CARE database using their original structure and their own definitions. The CARE PLUS project (phase 1 and phase 2) allows to facilitate the bridges between the different national databases, supplying a framework of transformation rules to ensure the comparability of the variables and values.

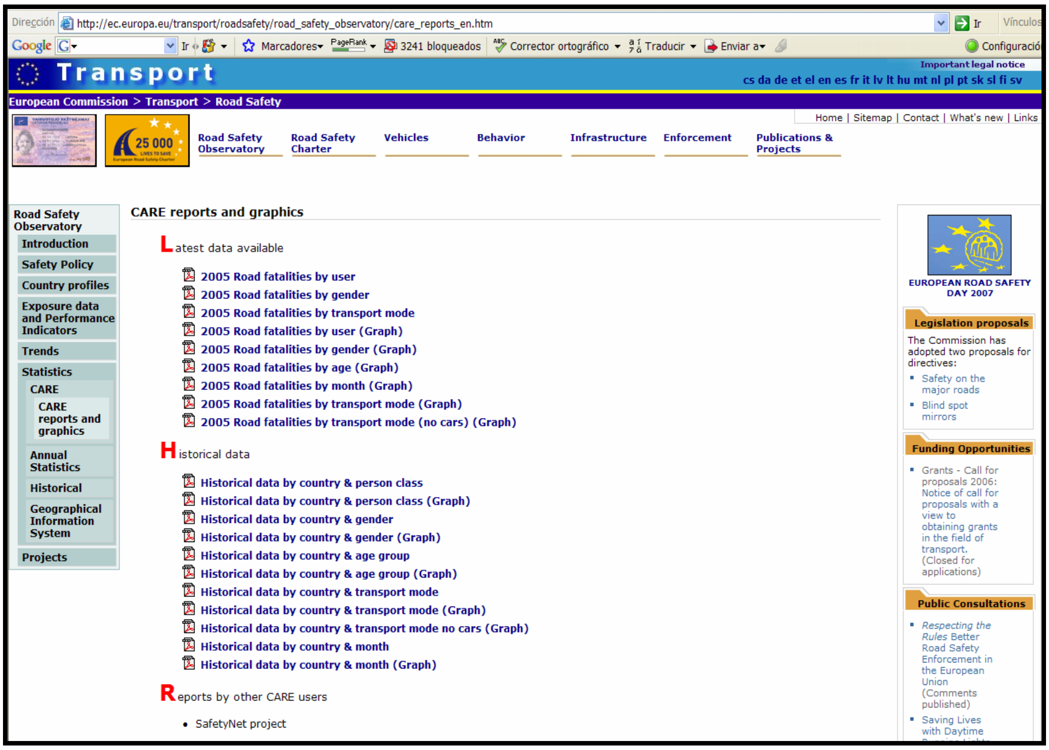


Figure 6: CARE website (http://ec.europa.eu/transport/roadsafety)

3.6.2. CHILD (Child Injury Led Design)

This project started in 2002; it was set up by members of 6 different countries (Germany, Spain, Sweden, United Kingdom, France and Italy). Together with the data from the Child Restraint Systems (CREST), it tries to go into detail in the achievement of child accident data, from in-depth investigations, and to analyse the relationship between the use of child retention systems with the severity and the type of injuries.

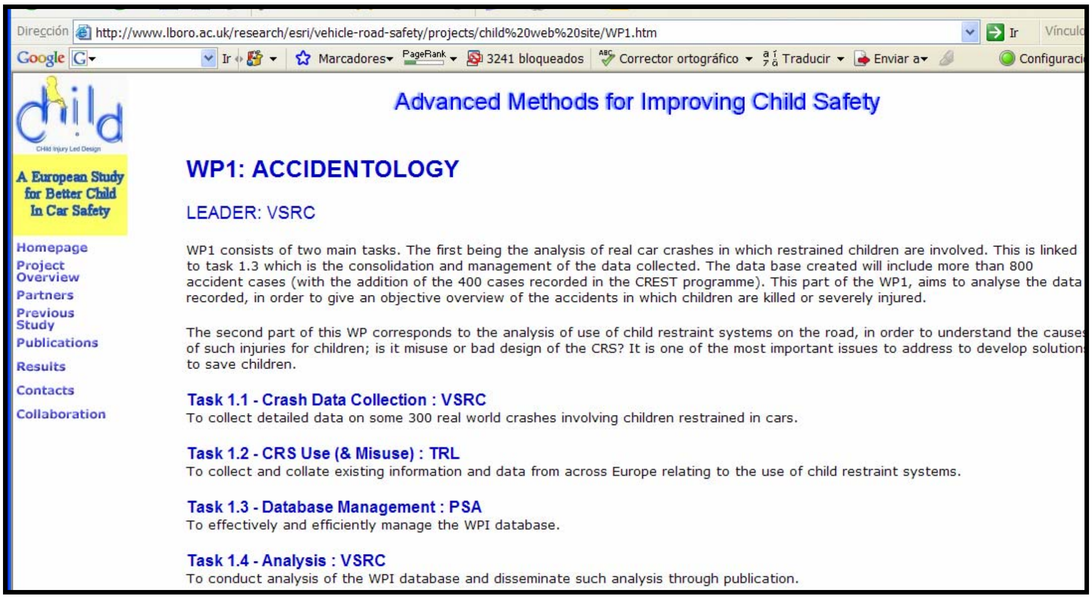


Figure 7: CHILD website (http://www.childincarsafety.org)

3.6.3. EACS (European Accident Causation Survey)

This project was launched by the European Automobile manufacturers Association (ACEA) and the European Commission in 1996. The project aim is to collect accurate information on the causes of road accidents. It focuses on the pre-crash phase, particularly on vehicle factors and safety systems (e.g. ESP); less depth data are reported on injuries.

The collected information is encoded - after an in-depth investigation of the accidents - in the DAMAGE database. Specific data coding methodology were defined in order to harmonize data and ensure consistency across countries.

3.6.4. ECBOS (Enhanced Coach and Bus Occupant Safety)

The ECBOS project took place under the 5th Framework Programme. Based on the accident databases of each involved country (Austria, United Kingdom, Germany, Italy and Spain), a statistical analysis of all bus accidents was performed. From the results, in-depth studies were carried out on several accidents. The data from the Governmental database together with the data from the in-depth studies were integrated into a general bus accident database.

Figure 8: ECBOS website (**http://www.vsi.tugraz.at/ecbos/**)

3.6.5. ECMT (European Conference of the Ministers of Transport)

The European Conference of the Ministers of Transport (ECMT) 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. 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 and casualties, and on risk exposure related data.

Nowadays, this institution is part of the Joint OECD/ECMT Transport Research Centre and collaborates in the development and management of the IRTAD international database.

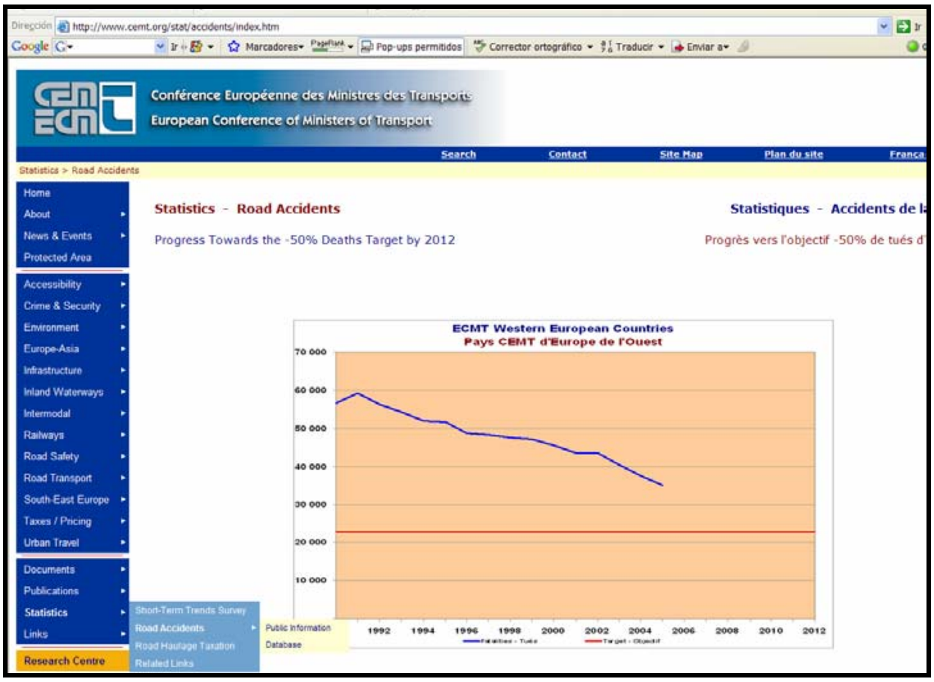


Figure 9: ECMT website (**http://www.cemt.org**)

3.6.6. ETAC (European Truck Accident Causation Study)

The objective of the project is to study the causes of heavy vehicles accidents. The project was initiated by the European Commission and the International Road Transport Union (IRU) in 2004. Researchers from France, Slovenia, Germany, Spain, Hungary, The Netherlands and Italy deeply study more or less 600 accidents with trucks and create a database on the main causes of such accidents.

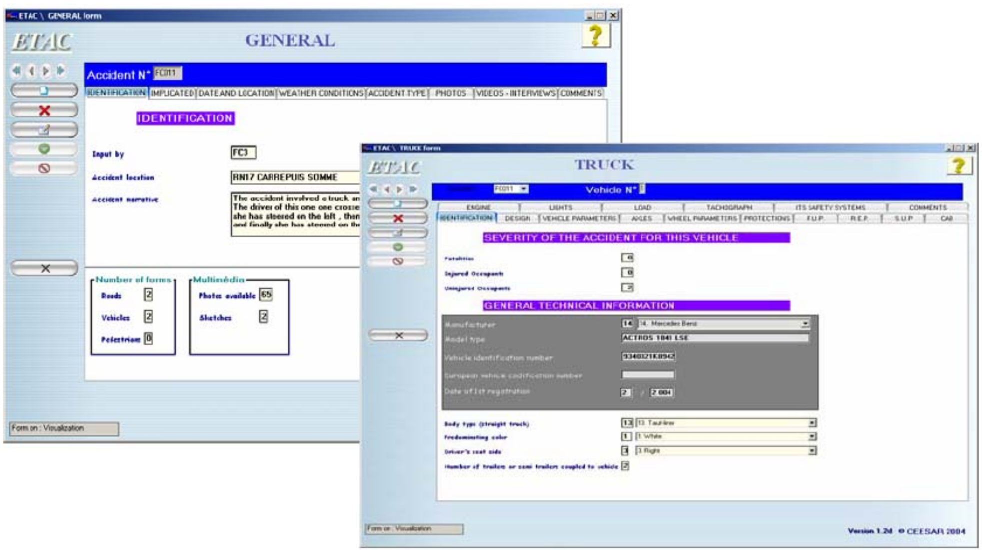


Figure 10: Screens of the ETAC database (**http://www.iru.org**)

3.6.7. *Eurostat (Statistical Office of the European Communities)*

The Eurostat publishes an overview of transport statistics for the EU Member States and European Free Trade Association (EFTA) countries (Norway, Liechtenstein, Iceland and Switzerland). Data is collected by means of supporting legal acts and voluntary questionnaires to be filled by Member States.

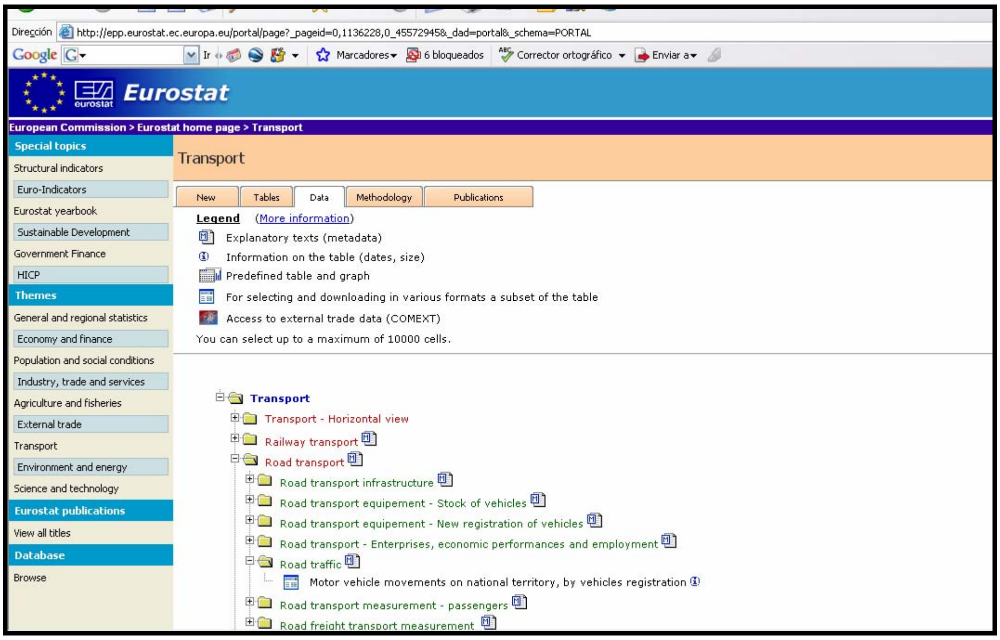


Figure 11: Eurostat website (**http://epp.eurostat.ec.europa.eu**)

3.6.8. *MAIDS (Motorcycle Accident In-depth Study)*

The purpose of the study was the identification of the causation factors of motorcycle accidents. The project focuses on injury prevention, motorcycle improvements, and a better understanding of the human factor.

The study looked at 921 motorcycles crashes in France, Germany, Italy, The Netherlands and Spain. The research focussed on the reconstruction of each crash, witness interviews, involved vehicles inspection, and medical records of injured riders and passengers analysis to identify all the factors that contributed to the crash and its outcome.

Figure 12: MAIDS website (**http://maids.acembike.org/**)

3.6.9. *PENDANT (Pan-European Co-ordinated Accident and Injury Database)*

The project provides accident and injury data to support EU vehicle and road safety policy making by developing two new European data systems. The first will be collected in eight European countries and will contain in-depth crash and injury data. The second information system is using hospital injury data collected in three EU countries.

This project continues the work of the STAIRS project (Standardization of Accident and Injury Registration Systems) on European harmonization of in-depth data on accidents and casualties. It is expected that this database will coordinate with CARE, given that both systems provide complementary information on accidents.

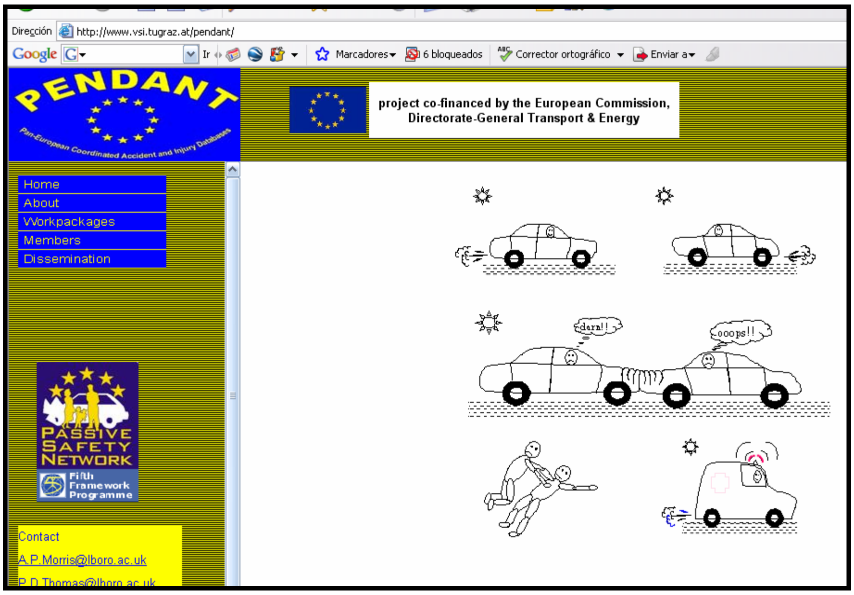


Figure 13: PENDANT website (**http://www.vsi.tugraz.at/pendant/**)

3.6.10. RISER (Roadside Infrastructure for Safer European Roads)

RISER is a European road safety project co-financed by the European Commission through its Competitive and Sustainable Growth Programme. The project took place from 2002 to 2005. The objective of RISER is to provide reports and guidelines for motorway safety professionals to design and operate safer roadside infrastructure. The aim is to minimise the consequences of single vehicle collision.

The statistical databases of different European countries (Austria, Finland, France, Spain, Sweden, The Netherlands and United Kingdom) were investigated to understand single vehicle collisions. Additionally, data was collected from existing databases for an in-depth study of single vehicle accidents.

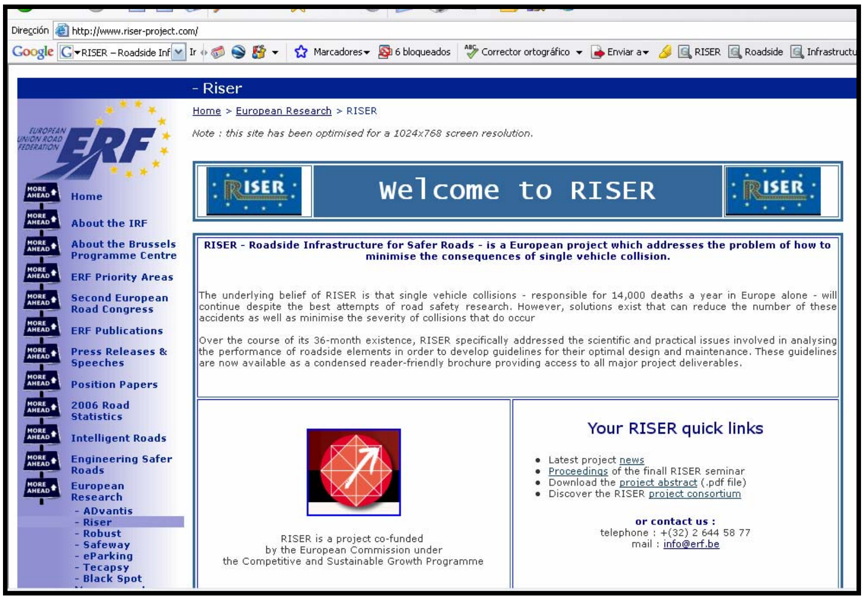


Figure 14: RISER website (http://www.riser-project.com/)

3.6.11. SafetyNet

SafetyNet is an Integrated Project funded by DG-TREN of the European Commission. The objective of the project is to build the framework of a European Road Safety Observatory, which will be the primary focus for road safety data and knowledge. Among other activities, it will extend the CARE database to incorporate the new EU Member States and will develop two new databases: an in-depth fatal accident database and an in-depth accident causation database (WP 5).

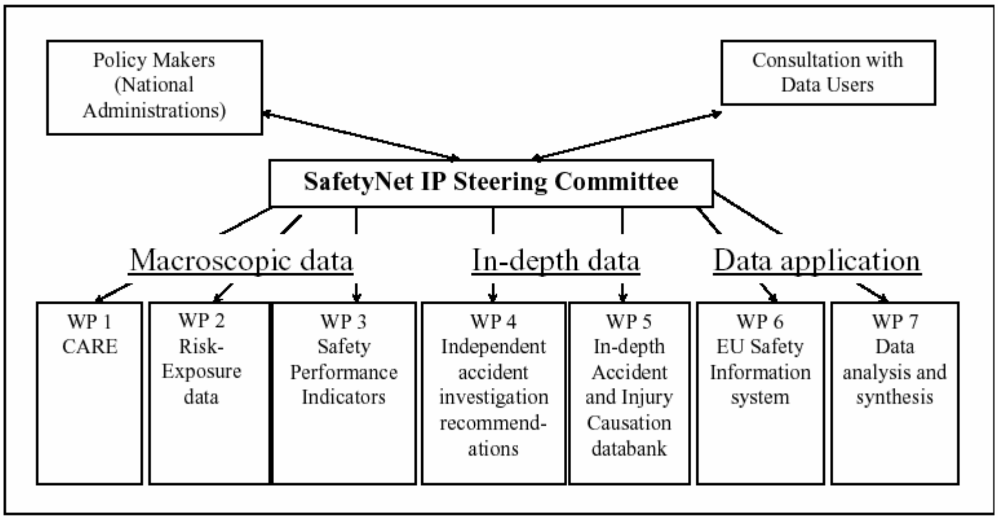


Figure 15: SafetyNet's technical and management structure (http://www.erso.eu/safetynet/content/safetynet.htm)

➤ **In-depth fatal accident database**

The information collected comes from the National police forces data (or insurance company data for Italy) from different European countries: Germany, Italy, France, Sweden, Finland, The Netherlands and United Kingdom. From a representative sample of fatal accidents collected in the participating countries, information has been extracted on environmental factors, in relationship with the road, the vehicle and the driver, as well as variables related with injuries.

➤ **In-depth accident causation database**

The database will contain over 400 variables on the accident circumstances, with a strong focus on understanding the causal factors leading to the accident. For that, specific methodologies of accident investigation and reconstruction were used in six European countries.

3.7. The national accident databases

The statistical data collected through the accident reports has the objective to consolidate several databases both at the national and at the regional and local scale. The stored data represents the basic matter for the statistical study of the accident rate with aims both of diagnosis and of evaluation and control. The quality of the collected information is fundamental in the road safety policies.

The national database management shows differentiated profiles in the different European countries. This way, the administrations in charge of the data management are varied: National Institutes or Offices of Statistics, Ministry of the Interior, Ministry of Transport, Ministry of Justice, etc.

Just to mention some illustrative examples of the different situations, in **Spain**, the organization in charge of the management of the accident data collected by the police is the Research and Training General Sub-directorate of the Traffic General Directorate (that depends on the Ministry of the Interior). In **The Netherlands**, the national database is managed by the AVV Transport Research Centre, of the Ministry of Transport. In the **United Kingdom**, there are two accident databases, one for Great Britain (STATS19 since 1948), managed by the Ministry of Transport, and another one for Northern Ireland. In **Austria**, the accident databases are sent by the federal police to the Austrian Office of Statistics (OSTAT). In **Denmark**, the Office of Statistical Information is in charge of the data. As for **France**, the Service of Road and Motorway Technical Studies (SETRA: *Service d'Etudes Techniques des Routes et Autoroutes*), that belongs to the General Directorate of Road Safety and Circulacion (DSCR: *Direction de la Sécurité et de la Circulation Routières*), is in charge of managing the national database.

As far as the data transmission is concerned, for the last years there has been a trend to enter the data locally and send it to the national authority in electronic format.

For example, in **Denmark** and in **Italy**, a number of local police forces fill in a computerised accident data collection form and then forward this information by electronic means (mainly on a disk). A similar procedure is used in **France**, **Luxembourg** and the **United Kingdom**. In **Germany**, the development of an electronic data collection system allows the immediate sending of the information to the national databases (see section 5.1.3 about the EUSKa system, p. 108). In **Finland** the PATJA information system is used. In **Belgium**, since 2006, a new integrated system has been launched, system that makes the accident data entry and sending (electronically) easier for the federal and local polices (which already were working, since 2003, with independent computer systems) (see section 5.1.5 about Pol Office, p. 113). In **Norway**, up to 2001, filled in and send by post the "Road Accident Report" to the Norwegian Statistical Centre, organization in charge of coding and statistically analysing the data. Since then, the information is directly entered in the police traffic accident database and is electronically sent to be statistically treated. In **Poland**, the police enter the information collected through the accident forms in the SEWIK central database. Once the validation procedures are done, they are sent to the central administration to be statistically treated.

In **Spain**, in the local administrations, the procedure of sending it in paper is still in use. Nevertheless, many municipalities have already started with the data electronic sending systems. This way, for example, in 2005, the Servei Català de Trànsit (Catalonia) developed a system to entry the data on-line

(SIDAT 1 and 2 projects), that has been simultaneous to a full reformulation of the information contents of the autonomous accident statistical report. At the level of the entire Spanish state, the DGT is raising, in the ARENA2 (2005-07) framework, the reformulation of the current national accident data system (called ARENA), with a special focus on the improvement of the collection systems in urban areas.

In addition to the official databases, based on the police information, and from which the information that is finally collected at the European level in CARE is extracted, we may stress on some other national databases, in which the collection, the management and the aim of the databases are defined by law (for a better knowledge of these databases, see the European Observatory webpage or the SafetyNet Deliverable D4.2). In Table 6, the organizations in charge of managing and exploiting some national databases are pointed out.

Country	Official Databases	Other databases
Austria	Statistics Austria: http://www.statistik.at Austrian Road Safety Board: http://www.kfv.at	
Belgium	Institut National de Statistique/ SPF Economie: http://statbel.fgov.be/port/mob_fr.asp Institut Belge pour la Sécurité Routière: http://www.bivv.be	
Denmark	Danmarks Statistik: http://www.statbank.dk	
Finland	Statistics Finland: http://www.stat.fi Central organization for traffic safety: http://www.liikenneturva.fi Finnish Road Administration: http://www.tiehallinto.fi	VALT (database on fatal road traffic accidents and database accidents compensated by insurance): http://www.liikennevakuutuskeskus.fi
France	National Interministerial Road Safety Observatory: http://www.securiteroutiere.equipement.gouv.fr	
Germany	Statistisches Bundesamt (STBA): http://www.destatis.de Der Polizeipräsident in Berlin: http://www.berlin.de/polizei/verkehr/statistik.html	German In-Depth Accident Study (GIDAS): http://gidas.bast.de
Greece	National Statistical Service Of Greece (Nssg): http://www.statistics.gr	
Hungary	Hungarian Statistical Office: http://portal.ksh.hu	
Italy	National Institute of Statistics (ISTAT): http://www.istat.it	
Luxembourg	Ministère des Transports: http://www.gouvernement.lu STATEC: http://www.statistiques.public.lu	
The Netherlands	Transport Research centre (AVV): http://www.rws-avv.nl Road Crash registration (BRON): http://www.swov.nl	Sports Utility Vehicle Study http://www.verkeerenwaterstaat.nl AAHTWO (Accident Analysis Heavy Trucks II): http://www.automotive.tno.nl/docs/btw_01_2001.pdf
Norway	Statistics Norway: http://www.ssb.no	
Poland	National Road Safety Council: http://www.krbrd.gov.pl/	
Portugal	Direcção-Geral de Viação (DGV): http://www.dgv.pt/dgv/index.asp	
Spain	Base de datos de accidentes con víctimas (DGT): http://www.dgt.es	
Sweden	Swedish road administration: http://www.vv.se Swedish Institute for Transport and Communications Analysis: http://www.sika-institute.se	
United Kingdom	Department for Transport: http://www.dft.gov.uk	

Table 6: Organizations in charge of managing and exploiting the national databases.

4. Accident data quality

The data quality is a delicate problem for those who use statistical information. In the case of traffic accident data, the problem gets a special importance given that the statistical exploitation results of such data represent the basis to assign resources and establish the road safety policies and actions.

Many studies carried out in several countries confirmed the existence of significant quality problems for accident data, problems that fundamentally make reference to the representativity of the recorded data in connection with the accidents that really happen and to the reliability of the collected information.

The term traffic accident "data quality" is generically used to make reference to a different set of problems that has to be clarified, mainly because it is not a very accurate term.

This way, the expression is very often used as a synonym of accuracy or reliability. However, this is just one aspect of the quality. In relation to the accident records, quality, from a broader perspective, may refer to two main dimensions: to what extent all the accidents are recorded, and to what extent the recorded data on those accidents is complete, precise and reliable. Some authors like Pfefer, Raub and Lucke (1998) add another dimension that makes reference to the level of accessibility and delay of the data in order to be able to be used by the users.

Even though these wide perspectives of the term "data quality" seem to us to be more than acceptable, for practical purposes and aiming to clarify, we will use the term "underreporting" when referring to the percentage of traffic accidents that are not collected in the record system, while we will use the term "quality" in a more delimited way to refer to different aspects of the accident data that is recorded, like the information accuracy or reliability and missing data.

Both underreporting and quality problems are, as a last resort, the result of several types of errors that take place in the data collection and management process. So, hereafter a typology of the main types or categories of errors that may appear in the traffic accident records are presented, as well as the deficiencies or limitations that derive from those errors.

4.1. Types of errors

At the general level, the distinction between two wide possible categories of data to which it is possible to refer has been made: sample or census.

The *sample data* is a subgroup of data of our interested population, selected by means of scientific procedures. The sampling is generally used when getting data from the whole population turns out to be excessively expensive.

We talk about *census* when the data collected comes from the whole interest population. From that perspective, traffic accident (with casualties) data is considered as census data, aside that, for practical purposes, the information of all the accidents is not notified. This is due to particular malfunctions of the record system, and not to the existence of a pre-established sampling.

On the other hand, we distinguish two types of errors: sampling error and non-sampling error. The first category is only applicable to sample data. The second one is applicable to both sample data and census data. Focussing on the second category, it is possible to distinguish several types of errors that will determine to what extent our data is far from the truth. These are:

- Non-response error
- Measurement or response error
- Encoding or recording error
- Non-coverage error

4.1.1. Non-response error.

It happens when a certain information, or data, is not collected from part of the population members (in this case, of certain accidents, involved vehicles, or traffic accident casualties).

We may have two situations:

(1) When there is no information regarding some population members – we just know that they exist – we are talking about unit non-response, or no information or data. For example, it would be the hypothetical case that we would know that an accident took place in a particular place because there are rests of broken glasses and other damages but there are neither vehicles and involved persons nor witnesses. Even though strictly speaking this would not be a unit non-response, given that we know the place of the accident at least, practically speaking we could consider it as a case of this type because there is no formal record of that accident.

(2) When only some pieces of record information or data are missing. In this case, we talk about item non-response.

Aside from the non-response type, we always talk about missing data.

Sometimes, missing data happens at random. On the contrary, when missing data turns up for a part of the interest population that recurrently differs in some particular aspect, we can talk about a plausible non-response bias.

4.1.2. Measurement or response error

In this case, some data obtained regarding one or several population components are wrong. In the case we are interested in (traffic accidents), this can be due to, among other reasons, the inherent difficulties of the data collection process, the complexity and difficulty to obtain some data, the inappropriate design of the accident form, or the inaccuracy, intentional or not, of the information given by the involved persons and witnesses of the accident. An example is the mistakes made when valuing the severity of the injuries or the psychophysical conditions of the accident casualties, or when deciding on the accident causing factors.

4.1.3. Encoding or recording error

In this case we do know the correct data, but the mistake is made when encoding or recording it in any record system (questionnaire, electronic record, etc.). In this case the question is to what extent the final record reliably collects the information we know. A typical example would be the errors that happen when entering the data in the computer.

We would also include in this section the errors generated during the information processing, as for example during the recording, purge (detection of inconsistencies), and incorrect transformation of the data.

4.1.4. Non-coverage error

It refers to the population components that are not included in the record. These population components are not recorded because they have not been localized, because there is not any proof that they exist or in the case when they are known, but for any reason it has been decided that they must not be recorded. It would be, for example, the only-damage accidents in **Spain**, which do not appear in the traffic accident official statistical records.

4.2. Limitations of the traffic accident data

The previously listed errors, in the concrete case of the traffic accident records, give rise to the three main types of deficiencies or limitations in which we will focus our interest:

- Underreporting
- Under-recording or missing data
- Bias and errors

4.2.1. Underreporting

It makes reference to the accidents or traffic casualties that, for different reasons, are not collected in the records. So, it is connected with the level at which the records truthfully represent the figures in relation to the real number of traffic accidents and casualties. This problem is the direct consequence of the non-coverage errors and of the non-response error.

When valuing the level of representativity of the accident records, it is necessary to clarify that, in principle, it has to be considered in connection with the target population. This population, as we already saw, is being determined by the definitions in force in each state or region regarding which accidents have to be statistically taken into account. This implies that when the definition of the record target population will be more delimited, this will favour an increase of the potential problems of representativity in connection with the traffic accidents that really happen. In these cases, to the accident records that, by definition, should appear in the statistics but that do not for several reasons, we have to add the ones that are not recorded for not being included in the established definition.

Eventually, we may differentiate between two types of representativity: the traffic accidents and casualties that are not recorded because of the record systems insufficiencies or difficulties and the ones that are not because of the simple reason that they are not considered by the record system. The first type is the one that draw our attention even though the other one has a great importance given that it is necessary to take it into consideration for the point of view of the international comparisons.

4.2.2. Under-recording and Missing data

It happens when a given type of information (e.g. alcohol consumption or vehicle brand and model) is not collected for a ratio of accidents, being for the impossibility to get the data, because they raise special difficulties for its fulfilment or for any other circumstances that are difficult to specify.

This is essentially a problem of missing data, so we are talking about the *non-response error*.

If the missing data systematically and only appear in particular types of accidents, we would talk about potential *under-recording* bias or non-random missing data.

The problem of the missing data is that it makes difficult the data analysis:

- At the processing level: the calculation of given statistical techniques (like for example, the linear regression or the main components analysis) is based on matrix operations that require the disposal of all the complete information.
- At the results level: the existence of given standards of lost values may cause more important problems than the incomplete data on its whole (as these are random along the data matrix, any method that controls data will generate similar results, while if they depend on fixed variables, included in the analysis, or even unknown, the results are not acceptable).

4.2.3. Bias and errors

They represent the consequence of the measurement or response error, as well as the data encoding or recording error.

Regardless of the cause, we talk about *bias* when a type of data or information is systematically wrongly collected, showing a trend towards a certain value or value ranks.

We talk about *errors* when the incorrect data do not show any trend, and are distributed in an approximately random way.

The measurement or response error may show a random nature as well as a bias nature, while in the encoding or recording error (e.g. wrongly pressing a key) is much more difficult to be random.

In the following section some studies regarding the traffic accident data quality problems are reviewed.

4.3. Methods to study the Traffic Accidents data quality

To assess the quality of the traffic accident records, it is possible to make use of several methodologies. The type of strategy that we use depends on the objective and the type of limitation present in the data. This way, we may differentiate three categories:

- Methods to study Underreporting: Linkage and capture-recapture
- Methods to study Under-recording
- Methods to study bias and errors

4.3.1. Methods to study Underreporting

First of all it is necessary to define the measurement units used to analyse the accident rate. Voas (1993) emphasizes four elements traditionally used as accident rate measures: 1) the number of traffic accidents, 2) the number of fatalities and/or injured persons caused by the accident, 3) the number of involved drivers and 4) the number of dead and/or injured drivers. While the first one refers to the traffic accidents strictly speaking, the three others refer to the casualties of the mentioned accidents. Event though the four traffic accident rate measures have a strong correlation degree, it is important to specify which one we are using in each case.

Most of the revised studies use the number of traffic accident casualties as criterion. This is mainly due to the fact that the revised studies use two basic strategies when comparing the representativity of the police records: compare with the health care system records, or use the survey data. In both cases the comparison data, by a majority, makes reference to the accident casualties.

When the contrast data comes from the hospitals, the comparison may be done at the aggregate level as well as at the individual level. At the aggregate level, what is usual is simply to compare the number of traffic accident casualties, collected by the hospitals in a fixed geographical area, with the ones collected by the police in the same area.

This approach raises many difficulties, particularly being able to properly determine the geographical areas that belong to each hospital centre. This problem is highly minimized when the area covers the whole country.

One of the few studies that exist and that use an aggregated approach to compare accident casualty police data with the hospital records, is the one carried out in **Sweden** by Hagen (1993). The author used as sample, data from 4 major hospitals that cover 10 % of the population. From the hospital data, he performed an estimate of the data for the whole country and subsequently compared them with the police records at the national level.

In the case of survey data, the comparison is always done at the aggregate level. These studies are using ad hoc research surveys or they make the most of the data of other surveys in which questions on traffic accident involvement and consequences were asked.

From another far more common perspective, there are the studies that use an analysis methodology at the individual level (a case is identified in a reference database and it is watched whether it has been recorded or not in another database from matching methods). Basically, it is possible to differentiate two types of studies. The ones that are based on the linkage data methods and the ones that use the capture-recapture methods.

The linkage methods are based on the use of data matching or linkage techniques to identify the casualties of the hospital records that do not appear in the police records (and vice-versa). With that, we manage to identify the traffic accident casualties that needed to be hospitalized or to be attended in the E.R. and that do not appear in the police records, which give us an estimate of the representativity of these records in connection with the casualties attended in the hospitals.

The variables that are usually used to match the records are the date and time of the accident, the place, the type of user, the date of birth (age) and the gender of the victim.

The method assumes that the total number of the real victims is the addition of the records that have been matched with the specific ones of each database (Figure 16).

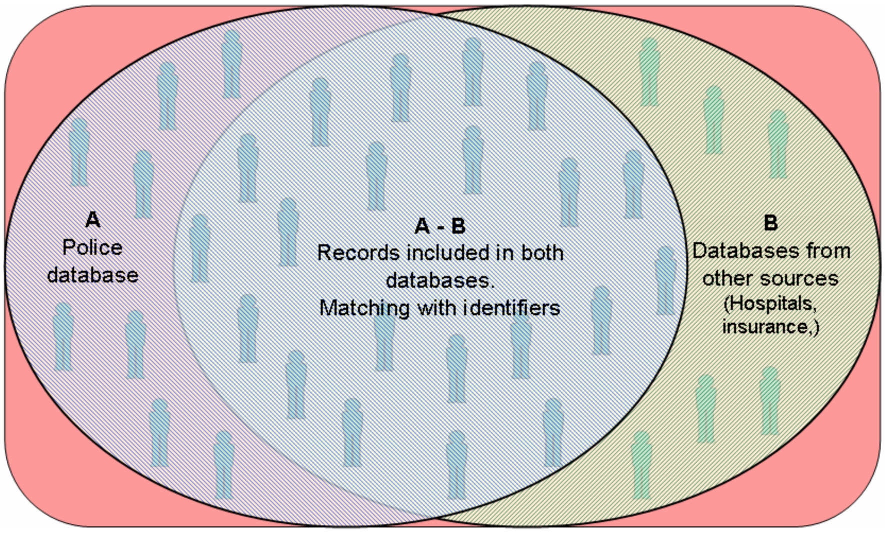


Figure 16: Illustration of the linkage method to study accident collection systems underreporting

The capture-recapture methods allow assessing the number of cases that belong to a fixed group, using for that two or more data sources extracted from a population to be studied. These lists are compared and the overlap degree is set. With this method, the sources are maintained as separate to assess the number of common records and estimate the number of cases that were not collected in none of them.

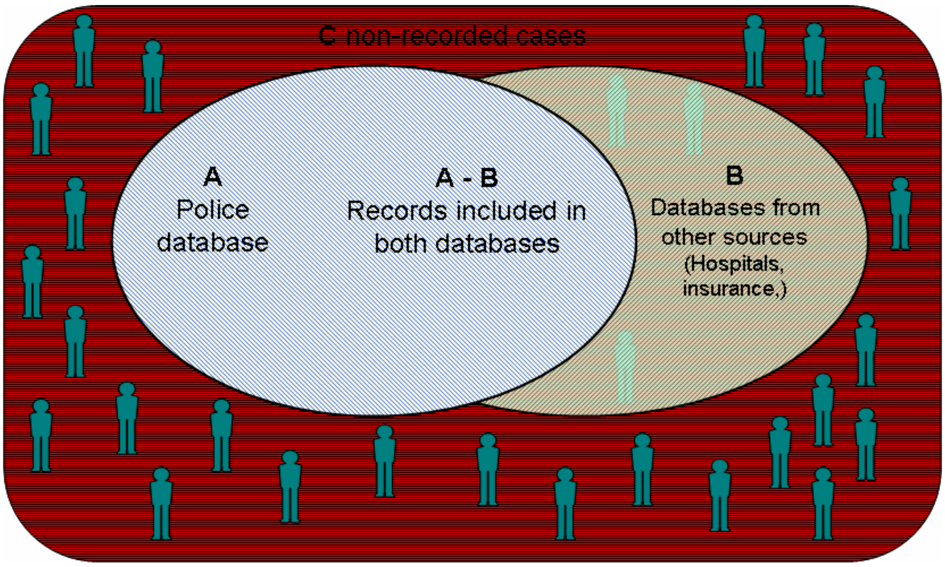


Figure 17: Illustration of the capture-recapture method to study accident collection systems underreporting

It had been used for the first time by Laplace in 1783 to estimate the French population. The technique was developed more particularly in the fields of ecology, biology and zoology, specifically to carry out estimates of animal populations. At the end of the eighties, the use of this method was notably extended to the field of epidemiology, in medicine more concretely, stressing its application to the study of cancer, infectious diseases, diabetes and drug addictions.

During the last decade, different experiences have proved its usefulness to estimate traffic accident fatalities and injuries from police and health information sources (Razzak & Luby, 1998; Morrison & Stone, 2000; Tercero & Anderson, 2004). This method allows identifying the coincidence of persons in the different information sources and, with that, to estimate the real number of cases or the total size of the study population and its confidence interval. In order to carry out the coincidences, the identifier fields are taken into account (name, age, gender, accident date…).

The sources to be used might be police records, hospital records, insurance records, death certificates…

The studies brought to light the deficiency – when separate - of the hospital records, the police records and the aggregate databases to carry out an appropriate magnitude representation of the traffic accident injuries, and recognized the usefulness of the capture-recapture methods based on the combined information available in the above-mentioned sources.

The Figure 18 shows in broad outline the information that is extracted from the capture-recapture method, where:

- "A" stands for the police database: it is composed of the victims not collected in the hospital database ("a") and the shared records ("AB")
- "B" stands for the hospital database: it is composed of the victims not collected in the police database ("b") and the shared records ("AB")
- "AB" stands for the records shared in both databases
- "C" stands for the victims that are not collected in any database: this is the only unknown data, which can be estimated through different mathematical methods. Among others we may stress on the maximum-

likelihood estimate derived from the Log-linear models, the almost unbiased estimate or the estimate from the application of the logit models (Freixa, Guàrdia, Honrubia y Peró, 2000)

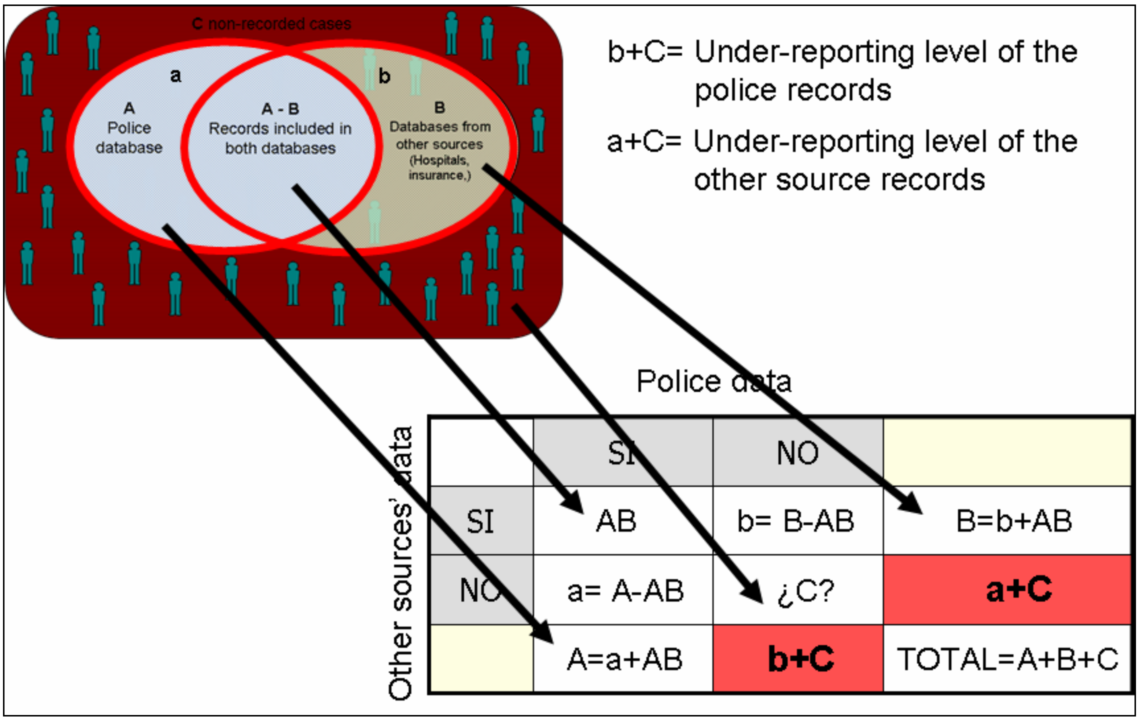


Figure 18: Illustration of the capture-recapture method to study accident collection systems underreporting

From this information, it is possible to estimate the level of underreporting of each database ("b+C" and "a+C") and the total number of "real" victims (we have to bear in mind that "C" is an estimation and therefore there exists an error margin in the calculation of the total number of victims).

In order to carry out a suitable application of this method, the information sources have to fulfil four basic conditions: 1) The population from which the list is set out has to be closed; 2) The data sources have to be mutually independent; 3) The capture probability has to be homogeneous for all the persons of the population; and 4) There must not have any loss of the identifiers assigned to each person.

When some of the criterions do not fit or when it has been decided to work with more than two sources, then other models have to be used (log-linear regression models, logit models).

Independently of the methodology that has been used, the representativity uses to be studied both as a whole and in a disaggregated way by sub-groups (by severity, user and/or accident type, age, gender, occurrence moment, etc.). The objective is to try to set which factors are related with the probability that an accident casualty will be collected by the police or, from another perspective, what is the representativity degree according to the sub-groups of traffic accidents or casualties defined in each case.

4.3.2. Methods to study Under-recording

In the case of under-recording, providing its presence and magnitude is something that is, in principle, quite easy. The mere inspection of the univariate frequency distributions of the different variables or fields of the accident questionnaire gives us information about the percentage of missing data for each one. Nevertheless the problem is more complex given that it is necessary to set to what extent the absence of data in a particular variable is related or not with the possible presence of biases that explain under-recording or non-aleatory missing data.

It is important to know the mechanisms that cause the missing data, given that these mechanisms, in conjunction with the statistical treatment given by the analyst to the data, will specify the type of impact that they might have on the analysis results.

Little and Rubin (1987) set different classification criterions of the mechanisms that cause missing data. On of them sets a missing data typology depending on the degree of randomness of these ones. This way, by taking into account the response probability of a variable Y (e.g. presence of alcohol intoxication), considering another variable X (e.g. accident occurrence moment) we may have three possible situations:

1 – That such probability would be independent from Y, and from X. In this case, it is about Missing Data at random (MAR). In our example, it would be supposed that the probability that the data will not be collected does depend neither on the alcoholic intoxication degree nor on the accident occurrence moment.

2 – That the probability depends on X but not on Y, which we would call Conditional Missing Data at random (CMAR). Following our example, the probability that the alcohol data will be collected depends on the accident occurrence moment (e.g. it is more probable in the weekend night accidents) but, considering this particular moment, the process is aleatory (e.g. the fact that, in the set of night accidents, the alcohol test will be done or not is an aleatory question).

3 – The probability depends on Y, and possibly also on X, case that involves that the data is not missing at random, and is neither MAR nor CMAR. Coming back to the example, in this case, the probability that the data will be collected depends on the intoxication level (to greater intoxication, greater external symptomatology and, consequently, greater probability that the police will carry out the test), and maybe also on the accident occurrence moment. This last situation is the one that may be the more probable for our example.

In this line, the authors point out that the missing data completely at random can be "ignored" when analysing the data. In this situation, as the under-recording is due to a completely random process, the records that have the full information represent a representative sample of the total of the records.

However, in the cases 2 and 3, in which determined under-recording biases appear, the situation gets complicated (particularly in the last case). Nevertheless, lots of techniques and strategies to deal with this type of situations have been developed (e.g. data allocation methods) though its revision remains out of the means of this work. Some revisions about this subject may be found in Valero (1999) and Valero and Young (2000).

Strictly speaking, we can only try to set the degree of randomness of the missing data in the case where we are reasonably sure that we have information on all the variables X that might be related with the Y value. On the contrary, it may happen that Y depends on another variable Z that we have not collected and which values, therefore, are unknown. In this case, we would also be talking about non-aleatory missing data, even though we could not be able to specify them.

If we assume that all the variables related with Y are included (or at least the more important ones), there are several methods to try to estimate the degree of randomness of the missing data in the traffic accident records in a quite easy way, basically through a bivariate or multivariate data study with an exploratory and descriptive approach.

Finally, with regard to the studies of the traffic accident record under-recording or missing data, we want to point out that these studies are generally carried out by using the data of the own accident record, without turning to other records or external data sources, things that do happen for the bias and error studies, and in those that analyse representativity.

4.3.3. Methods to study biases and errors

In the possible strategies to contrast the data quality, as for the biases and the errors, Valero and Young (2000) differentiate between the methods directed to check the data reliability and the ones which objective is to examine its consistency or coherence.

The first ones (data reliability) try to check whether the information finally entered in the accident database matches with the one collected in the accident questionnaire. Basically, it as about trying to detect possible errors generated during the different transcription processes, so we are referring to coding or entry errors.

In order to avoid this type of errors different methods have been raised, like the double entry or the revision by several operators. These methods are thought for survey or questionnaire simple data for which, generally, the only thing that has to be revised is the possible errors during the entry process. However, in the case of the accident reports, on the one hand, the volume is huge and, on the other hand, the information goes through several transcription processes. This way, by taking as example the usual procedure used in Spain to collect and process accident data, firstly the main data is usually collected on the spot, normally for the writing of the report, which lies in a descriptive narration of the facts (first transcription). Then, the report information is used to fill in the accident report (second transcription), of which data is next entered in the database (third transcription). In the cases where there is also a differentiated local accident report, we have up to four transcription processes. The time and resources cost that would suppose revising and correcting the possible errors derived from these transcription processes, makes it extremely difficult to control this type of errors once the process is over.

When the objective is to consider the consistency or coherence from a logical point of view, the situation turns up to be more favourable. We are talking about impossible data like "Sunday and Working day" or "single accident and five involved vehicles", or very unlikely or very suspicious like "heavy rain and dry road" or "driver and 10 years old". We are talking about measurement or response errors that might be detected through a control of the consistency between different fields or variables of the accident report. Naus (1982) has set two types of tests:

a) The exact ones or also called deterministic or logical. These ones would be tests to determine impossible data.

b) The approximate ones (empirical or probabilistic). Their objective is to identify very unlikely data.

Again, we observe a huge number of tests and methods of both types. Among them we stress on the out-of-range value detection ones (e.g. 30th of February), and the checking of the consistency of some data combinations in several fields of the accident report.

In both cases, the controls may be done by computer, during the data entry process, or later. This way, when there are some deterministic relations between different fields or variables, it is possible to define filters in the accident database that detect when an impossible combination is entered. For example, not allowing indicating working day when the day of the week is Sunday. Some of the data can also be automated, like for example, setting the day of the week from the date (assuming that this one is correct just as it has been entered).

This type of quality control procedures is more usual in the case of the deterministic inconsistencies, but it can also be applied to some extent to the probabilistic situation, setting up some type of "warning" in the database. In the case of detecting not very plausible combinations, the operator may check whether there has been some kind of error in the data entry and correct it on the instant, either by directly knowing the right data, or by carrying out some searches with the persons involved in the accident (casualties, witnesses, police officers…).

When the controls are carried out after the data entry process, it is usually done through several statistical exploratory techniques. We may stress on some easy ones like the univariate search to detect out-of-range values or outliers (which do not have to always be errors), or the bivariate search, by crossing pairs or groups of variables to detect inconsistent values, either impossible or improbable, that could be due to errors made in some of the contrasted variables.

In spite of the importance of the process of debugging and quality control of the information, the traffic accident data is defined with some particular characteristics that incite the added use of specific work strategies to contrast its quality. A great part of the measurement or response errors come from the great difficulty that means reliably fixing which one is the "true" answer for many fields or variables in the accident report, even for data with a high level of objectivity. To put an example, for the police agents is really difficult to reliably fix the type of injuries or the severity of these or also the road type and features of it in places like the intersections or the cross-town links. The topic gets more complicated in the case of "less objective" variables like the accident cause.

In view of these situations, other data validation strategies are needed. They consist in contrasting the police record data with other external records of which data is assumed to be more reliable or specialised. A typical example is to compare the data on casualties, injuries and severity with hospital data.

The record linkage methods make reference to the process for which data from different records belonging to the same person or event is brought together. As early as 1946, Dunn described the usefulness of this approach to study public health, given that it allows bringing together information on the same persons from different sources: health, accident, death records, etc. Even though this strategy has a long tradition inside the public health and epidemiology, it is only during the second half of the eighties when the first experiences were carried out in this line (Agran and Dunkle, 1985; Barancik and Fife, 1985; Fife, 1989; Agran et al., 1990).

The linkage methods may follow two strategies: the deterministic or the probabilistic ones.

The deterministic methods are the ones that pair up the records according to a perfect match in a variable or, preferably, in a set of variables. These methods may be applied to the records in which highly discriminating data like the name, the ID card or the insurance policy number are collected. In these cases the situation is relatively simple and the pairing between the records is reliably done (whenever there are not any error in the identifier data entry or missing data). An example could be the linkage of the accident records with the vehicles records by the registration number (unique identifier), and the brand and the model (validation criterions to prevent the effects of possible errors during the data collection or entry).

When we link records from two files or databases without unique identifiers, the combined use of a set of variables is needed so that it allows identifying each person or case in a unique way. In this context the concept of discriminatory *ability* or power makes reference to the probability of false positives or, in other words, the probability of the complete random coincidence for all the variables of the identifier set, among records that do not belong to the same subject or event (Newcombe, 1998). The discriminatory power is directly related with the number of variables that compose the set of identifiers, the range or different values that may take each one of these variables, and the frequency distribution between the mentioned values.

However, even though we may fix a set of variables of which values be unique for each accident or casualty, the data, as we already saw, shows a certain percentage of errors and/or missing data in the set of identifiers, or with "double or multiple matching" when a case matches with two or more cases of the other database.

For that reason, it is more appropriate to use probabilistic methods. These more complex methods start from the verification that not all the variables have the same discriminatory power (e.g. the age is more discriminatory than the gender). In this case, not only the number of matches between the identifier variables is taken into consideration, but also the discriminatory power of the ones that meet. For that, weightings are assigned to the variables according to their discriminatory ability. For example, the date of birth or the age are far more discriminatory than the gender or the civil status: take two subjects at random, it is far more probable than they will have the same gender than the same age, or, even less, the same date of birth.

In short, the probabilistic procedures provide the link between databases without unique identifiers or with a certain level of error in the data. They allow the matching of records without the need of a perfect match between identifiers. The geographical and temporal localization, the vehicle type and other variables may be used to identify a concrete accident. The age/date of birth, the gender, the injury description, the name or initials may be used to find a given person/victim.

The probabilistic procedures are based on iterative algorithms that involve successive steps and comparisons between the records so that, at each step, matchings can be ruled out. The development of probabilistic methods has gone on a par with computer developments that allow its application with a great number of records with identifiers composed of several variables. Nowadays there is already a certain amount of software - developed for that purpose - that have been applied to the case of accident records. An example could be GIRLS (Generalised Iterative Record Linkage System) developed in the Canadian Institute of Statistics by Hill and Mill (1981), and implemented to link police records of casualties with hospital records by Ferrante, Rosman and Knuiman (1993). Moreover, in the environment of the Statistical Analysis System (SAS), a set of macros and applications (SAS/Linkpro) has been developed to carry out record matching procedures that have been applied to accident records (Ross, 1995). Another example is Austin's research, who used the Data Query Language (DQL) of the DataEase software (version 4.2), to set algorithms in order to match records (Sapphire DataEase, 1989). Nevertheless, possible on a global scale, the more extended application for the accident record procedures has been the one carried out by the NHTS in seven states of the USA (CODES project), using the AUTOMACH software of Match Ware technologies, Inc. (Jaro, 1995).

4.4. Quality of the accident data: main results of the international research

4.4.1. Underreporting

The problem of the lack of representativity is probably of the most serious one in all the countries where an accident data statistical record is carried out. Generally, the police data only represents a percentage of the real accidents and casualties, percentage that is surely hard to specify.

Obviously, it is impossible to know the exact number of traffic accidents. In the practice, the police only do act of presence in a part of them, so the other ones are not reflected in the official statistics (normally because the own users do not request the police presence). Furthermore, the police presence in a particular traffic accident does not necessarily imply that an accident report will always be filled in. The level of representativity of the police statistics is hard to determine and varies along with time, from one country to another, and even in the same country, particularly when different police forces are in charge of filling in the accident reports.

In **The Netherlands**, for example, it is estimated that 99% of the fatal accidents are collected, 60% of the hospitalized casualties, 20% of the slight casualties, and only 5% of the damage-only accidents (SafetyNet, 2006). In **Germany**, the Federal Motorway Research Institute (BASt) estimates that only 5% of the accidents with casualties are not recorded in the national database, and it is thought that they mainly refer to accidents with pedestrians or cyclists (Crow, 2004).

The International Traffic Safety Data and Analysis Group (IRTAD) has recently carried out a survey study on accident victim data underreporting (Derriks and Mak, 2007). The Table 7 shows the estimates of several IRTAD countries regarding the representativeness of the data according to the severity.

Country	Fatalities	Hospitalized	Serial injured	Slight injured	Damage-only accidents
Germany	95		68	64	
Iceland	100				
Netherlands	94	60	14	5	30
New Zealand	100	100	67		3
Norway	100				
Slovenia	100		94	82	49
Spain	97		67		
Sweden	100	90	50	20	
Switzerland	98		77	25	
USA	100		95	75	50

Table 7: Percentage of the data representativeness in several Member States of the IRTAD group (the remaining Member States do not appear in the table given that the requested data is unknown)

Generally, it has been observed that a minority of participants can define and describe the quality of its accident data, especially when differentiated by the type of accident or its severity. And on the other hand, we may emphasize that many countries are quite trusting on the representativeness of their records, mainly when referring to the fatalities or the serious victims. This representativeness decreases when talking about slight injuries or material damages.

As an example, we want to stress on some of the reasons that might explain the shortages that happen in the exhaustive collection of accidents:

- The place of the accident is complex and stressful: the agents are facing a task conflict where they prioritize the most necessary tasks to protect the users (attention to the victims, traffic control, etc.)
- Some persons involved in accidents do not request police attention, because the consequences are not serious or because the intervention might harm them legally (e.g.: alcohol consumption, no driving licence, etc.).
- The existence of legislative criterions that limit the characteristics of what has to be recorded as a traffic accident for the purposes of the official statistics. This way, for example, some countries do not collect slight

accidents or damage-only accidents (criterion of the accident severity), pedestrian or cyclist accidents (motor-vehicle criterion), "special" accidents (e.g. suicides or murders) or the accidents where there are not any vehicles in motion (e.g. falls in buses).

- Lack of police means to attend to all the accidents.
- Lack of motivation, for example due to work overload: to the in situ action, a great amount of legal and administrative papers that the agent has to fill in subsequently is added.
- Police competence in several geographical areas: the police officers sometimes attend the accident but do not fill in the administrative papers given that they are not in the area of their competence.
- Local or autonomic politic decisions about what, how to collect the information, and about what is sent to the state record.

In the Table 8, the results of several studies, carried out in different countries, in which the level of representativity of the official traffic accident casualty records has been estimated, are summarized. The reviewed studies use two fundamental strategies when contrasting police records: compare them with hospital or survey records. What is presented in the table is the percentage of traffic accident casualties that appears in the police records in relation with the reference or comparison data in each case.

Study	Type of casualty	Contrast	%Reporting[4]
SPAIN	Hospitalized casualties (> 24 h.) (Aggregate data)	*Hospitals*	91% (1995) 66% (2001)
SPAIN	Casualties attended in hospital (emergency or hospitalization)	*Hospitals*	65%
UNITED KINGDOM	Casualties attended in hospital (emergency or hospitalization)	*Hospitals*	62%
NORWAY	Casualties attended in hospital (emergency or hospitalization)	*Hospitals*	53%
SWEDEN	Casualties attended in hospital (emergency or hospitalization)	*Hospitals*	40%
GERMANY	Casualties attended in hospital (emergency or hospitalization)	*Hospitals*	39%
DENMARK	Casualties attended in hospital (emergency or hospitalization)	*Hospitals*	32%

[4] The concept of reporting makes reference to the percentage of traffic accident victims that appear in the police records in relation with the reference or comparison data (hospital data or data collected through a survey)

Study	Type of casualty	Contrast	%Reporting[4]
DENMARK	*All the non-fatal casualties*	*Survey*	20%
THE NETHERLANDS	*All the non-fatal casualties*	*Survey*	24%
SPAIN	*All the non-fatal casualties*	*Survey*	20%

Table 8. Summary of the works that study the level of representativity of the traffic accident statistical police records in several countries.

It has been observed, in the results, that when the contrast data comes from the hospitals, the percentage of the resulting representativity is greater than when it is compared with the survey data. This is due to the fact that, during these last years, the methodology used allows obtaining a huge quantity of casualty information, generally the slight ones, that even if they did not need hospital help, they did suffer injuries.

The studies focussed on the data about casualties attended in hospitals meet a representativity percentage of between 91% and 32%. In general, the studies carried out in hospitals are only indicative and not representative of the situation of the country, given that they are normally delimited or local studies, being much influenced by the specific situation of the place or region where the study has been carried out and that can vary from the national average. The Dutch and Spanish studies have been based on national surveys and have as referent the whole population.

The degree of representativity of the traffic accident varies according to the accident severity, the transport mode, the user type, the age of the victim, and the moment or place of the accident. This implies that the accident databases are biased depending on the mentioned variables.

It is not a matter of chance that a particular accident or casualty is taken into account or not in the statistics. This way, the serious casualties and, particularly, the fatal ones are much more represented than the slightest ones. Moreover, underreporting is more frequent in particular types of accidents: for example in light collisions, knocking into or over pedestrians, falling off two-wheel vehicles (accident falls – due to trying to avoid collisions with other vehicles -, collisions with parked vehicles…), or passengers falling off inside a bus due to sharp manoeuvres or collision with another vehicle (Frantzeskakis et al, 2000).

In the Table 9, the results of a study carried out in **The Netherlands** are shown, study in which the police record representativity is analysed in comparison with other national data records (CBS-Causes of Death, NPR-National Patient Register e ISS-Injury Surveillance System). When in accidents, mote vehicles are involved, there is a higher probability of a police record, effect that is also observed as the casualty severity is greater.

Registration rate			
Transport type	Fatalities (CBS 2000)	In-patients (NPR 2000)	Injured persons treated in A&E (ISS 2000)
Car/lorry/motorcycles	95%	84%	25%
Moped	96%	67%	14%
Bicycle	88%	33%	4%
Pedestrian	94%	55%	21%

Table 9: Registration rate of traffic accident casualties by transport type and severity (comparison with the CBS, NPR and ISS records). A&E= Accident and Emergency. Source: SWOV, 2006

In the case of Great Britain, the study of the underreporting issue has a significant relevance within the improvement plans and the accidents data quality control. Thus, recently a complete report has been published that deals with exhaustive form this issue (DfT, 2006b). The revision which they do of different previous studies made in the country, confirms the existence of significant percentage of underreporting in the data on casualties of accidents, as well as problems in the classification of the level of gravity of the same ones (mainly in agreement the gravity is smaller and in certain types of users). In Table 10 some of the results of these reviewed studies are summarized:

Author, year	Type of study	% of all reported	Other % reported
Bull and Roberts, 1973	Police vs. hospital		Fatal 100 Serious 81 Slight 65
Nicholl, 1980	Police vs. hospital	50	
Tunbridge et al., 1988	Police vs. hospital	61	Fatal 100 Serious 66 Slight 55
Austin, 1992	Police vs. hospital		Cyclist 67 Pedestrian 75 Driver 61

			Passenger 60
Hopkin et al., 1993	Police vs. hospital (no fatality)	64	Slight 69
Simpson, 1996	Police vs. hospital	46	Bicycle 22 Pedestrian 60 Car driver 70 Car occupant 53 Motorbike 57 Serious 55 Slight 45
Cryer et al, 2001	Police vs. hospital (no fatality)	61	Bicycle 31 Pedestrian 72 Vehicle 67 Motorbike 69
Broughton et al, 2005	Police vs. hospital	61	Cyclist 43 Pedestrian 66 Driver 67 Passenger 57 Motorbike 60
Ward et al, 1994	Police vs. hospital		Pedestrian 74
Ward et al., 2005	Police vs. hospital	52-60	

Table 10: Summary of the previous studies on underreporting en the United Kingdom. Source: DfT, 2006b

From the comparison of the STATS19 police data, with a sample of data registered in the emergencies services ("A&E" data), they have estimated a percentage of representativity of police data between 54% and 55%. Based on the user type, pedestrians and cyclists were correctly registered in 70% of the cases, drivers of vehicles with two wheels in 60%, and vehicle passengers only in 50% of the cases. Depending on the age, the percentage of casualties register aged between 20 and 24 years was only 45%.

In the Simpson study (1996), some types of victims and accidents of which level of representativeness is usually high are emphasized:

- Victims whose vehicles were seriously damaged.
- Victims whose injuries were apparently noticed at the scene of the accident.
- Victims that were brought to the hospital by the emergency services.
- Victims attended in the hospital immediately after the accident.

- Accidents with more than one victim.

As we have already mentioned before, in several countries, like **Finland**, they have chosen to complement the police record information by the one supplied by the cause of death records, and consequently the level of lost cases with regard to traffic accident fatalities has considerably reduced. However, the same thing does not happen with the rest of accidents with casualties. This way, they estimate that approximately 20% of the total of accidents with casualties is collected in this country. So, the level of collection is worse in the case of injured cyclists in single accidents (in most of the cases they are light accidents that do not need police intervention). Given that, nowadays, only the collection of accident data with a fatality or serious casualty is compulsory, the level of representativity has considerably increased in that country.

In the PENDANT project (Pan-European Co-ordinated Accident and Injury Databases), the group in charge of the WP3 on data analysis, have carried out a study which deals with "linkage" techniques of police information with hospital data, used in several European countries (Kampen, Pérez y Martin, 2005). Concerning the Netherlands, estimated values related to hospitalized traffic casualties, which are included in the police data base are only about 59%. Moreover, they conclude that the inclusion of serious injury cyclists is rather low.

Regarding the victim age and the user type, Barancik and Fife (1985) fix a representativity percentage of 28% for under 16 years old and of 60% for the rest of the population, and also 74% for the drivers, in front of 45% for passengers, cyclists and pedestrians. Harris (1990) points out that 91% of the under 14 casualties do not appear in the police statistics. The same study concludes that by user type, the representativity is of 15% for cyclists, 25% for pedestrians, 34 for moped or motorcycle passengers, and of 41% for car passengers. With regard to the injury severity, the author estimates a representativity of 79% for in-patients; 26 for the persons attended in the emergency service and 11% for the persons treated out of the hospital. Moreover, a greater representativity for the accidents with more than one vehicle involved shows up. Ajo (1996) also shows that the representativity is lower for single-vehicle accidents (55%) and especially low for the accidents with bicycles (30%). Rosman and Knuiman (1994) get a representativity percentage lower than the average for the single-vehicle accidents (56%). The different studies agree that the night accidents are more represented than the daytime ones.

Finally, it seems that there are important differences between motorway and urban area. For example, in Spain, a survey was carried out among the municipalities of the Community of Valencia in which it was stated that only 28% of the municipalities always sent the accident reports to the Province Authorities, 10% only in case of fatal accident, 22% only sometimes and the remaining 40% never.

Generally, the main limitations of the revised studies are the following ones:

- They are studies based on accidents/victims samples usually restricted to particular geographical areas. Therefore the results are based on population inferences (estimates) that, depending on the case, are linked with greater or smaller error margins.
- The compared collection systems are sometimes very heterogeneous, which makes difficult to match cases. This heterogeneity is found in different accident/victim definitions, in different recording and classification criterions, in different available information fields, in socio-economical parameters that have influence on the record (e.g. quality and efficiency of the police and/or health system), etc.
- The estimate methodology used is little defined (in many cases, the used methods are not described and it is not specified to which compared records the estimated levels of representativeness refer).
- It has been observed a great variability in the obtained results, which makes difficult to set few clear patterns that allow explaining underreporting. The results vary according to the used sample, the compared data sources, etc.

The SafetyNet project includes the implementation of updated studies on underreporting in the participating countries, obtaining updated data and suggesting standardized methodologies for this type of studies. (SafetyNet, WP 1.5). When the national studies will have been carried out and when the corresponding report will be available, we will have a completely updated vision of the situation at the European level.

4.4.2. Under-recording

Given the complex and difficult circumstances in which the data collection is carried out in the scene of the accident, it is logical that problems related with the lack of information of some data appear. Moreover, the accident report normally considers a great amount of information (see Table 5, pág. 29), sometimes difficult to specify or even to obtain. To this, we should add other aspects linked with the process, like the complexity of the accident report, the evaluation of the importance of its completion, the training of the persons in charge of the task, the amount of transcriptions of the information, and the need to carry out other more urgent administrative formalities, like the proceedings, the accident report and, when needed, the technical report.

Not all the fields or information types of the accident reports are equally sensitive to this type of problems, given that particular information are more complex or costly to obtain than other, so it is more likely that they will remain "empty" or that the data is incorrect.

Regarding the **accident location**, the right information about the place of the accident is highly important from the point of view of the road safety management. Nevertheless, there are many cases in which precisely establishing this information is not an easy task for the police.

In urban area, the accident location is usually based on the name of the street where it has happened. For that, some municipalities have specific codes for each street that provide the data processing. The literal record of the names of the streets sometimes generate non-unique accident locations, that is to say, several accidents are generated in the same street but they are recorded with different names (e.g. "Avenue" and "Av." already generate two different locations in the databases).

The junctions are usually defined by the codes or names of the streets that make it. The locations inside the sections of a street are usually defined by the number of the buildings or, also, by the monuments, squares or other well-defined places.

In spite of this, in many occasions, this information raises problems. There are <u>areas of the road network where specifying the location may be very difficult due to a lack of reference criterions</u>. This could be the case of the accesses, roundabouts, crossings, tunnels, bridges…

Moreover, even in the cases where there are enough parameters to set the accident location, the data is not sufficiently precise for the problems like the complexity of many places where accidents happen (roundabouts, junctions…), the lack of application of homogeneous criterions (the location is where the accident started, or where the vehicles eventually ended? and what if each vehicle ended in different streets? which one is the more important for the accident?,…), the lack of accuracy of the collected data, or possible errors that happen during the data entry or transcription process.

The accuracy in the accident location considerably varies between countries (Frantzeskakis, Yannis and Handanos, 2000). This way, it has been estimated that in **France** the location is not collected in 17% of the accidents (4% in large cities), and is imprecise in 20% of the accidents where it has been collected (the location is exceeded in more than 100 meters). In **the United Kingdom**, the use of the Ordnance Survey Grid Reference method (information is mapped onto the national road network) makes the accident location more accurate (Figure 19).

Figure 19: Geographical reference grid of the accident location system used in the United Kingdom

The use of GIS systems has allowed the improvement of this information collection, given that they are based on objective parameters of geographical location.

The data on ***alcohol consumption*** of the drivers involved in traffic accidents are considered in most of the accident reports. However, a complex set of technical and administrative difficulties uses to make difficult the systematic achievement of the tests in every case, so it is finally limited to particular types of accidents or casualties. For example, in the case of accidents with fatality or serious casualty, it is very hard for the police to determine the presence of alcohol, given that the casualties are rapidly moved away from the scene of the accident. In these cases, the information has to be obtained in the hospitals or in the death certificates, provided that these corresponding tests have been carried out, and usually after an administrative procedure generally complex and laborious.

Furthermore, there are important differences between the different countries. For example, in **France** the information of the drivers' alcohol consumption is available for approximately 50% of the collected accidents, while in **Greece** this percentage does not exceed 2% (Frantzeskakis, Yannis and Handanos, 2000).

On the other hand, it has been observed a phenomenon of selection of specific types of casualties to which alcohol tests are usually submitted in particular cases of an accident (bias).

Back in 1972, Waller already detected that it was less probable that the police could submit alcohol tests to specific groups like the elderly, the pedestrians or drivers that were not responsible for the accidents.

In a study carried out in the **United Kingdom** (Ostrom, Huelke, Waller, Eriksson and Blow, 1992), the information about alcohol consumption recorded in a wide sample of fatal accidents was contrasted. Among the general results, they emphasized that in 57% of the accident reports the information about alcohol consumption was not collected. Moreover, the presence of alcohol was not contrasted for 59% of the men, neither for 82% of the women, particularly stressing on the case of the pedestrians (the alcohol tests are not carried out in

90% of these cases). On the other hand, these tests were carried out with greater probability to the drivers considered as being responsible for the accidents (in 93% of the cases).

The probability that the alcohol consumption be contrasted is also related with the accident severity and the type of casualty. Voas (1993) concluded that 75% of the drivers involved in fatal accidents were submitted to alcohol test, while this only was the case for 25% of the drivers hurt in less serious accidents.

In **Spain**, the situation seems to be particularly deficient in this aspect. In a study carried out in the province of Castellón, the accident reports, the forensic reports and, in that case, the autopsies of a sample of motorcycle accident casualties were reviewed, concluding that only 7,5% of the cases were submitted to a alcohol/drug tests (De Francisco, 1996).

The scarce emphasis that seems to be in the collection of this information clearly contrasts with the huge importance of the problem. In this sense, the **Spanish** National Institute of Toxicology (Rams, Ortega y Sancho, 2003) carried out blood tests to a wide sample of drivers involved in traffic accidents. 42% of the cases had a blood alcohol content higher than 0.8, also detecting drug abuse in 8% of the cases. This information, together with the one collected about medicines that may affect driving, allowed estimating that approximately 63% of the drivers involved in an accident were under the effects of alcohol, drugs or medicines.

According to experts from different police forces, to fix the ***type and severity of the injuries*** is one of the aspects that poses most difficulties when filling in the accident report. On the one hand, because they are not health professionals, being very difficult to carry out injury appraisals (particularly with the scarce time available at the scene of the accident). On the other hand, because there are difficulties to carry out a follow up of the victims, either if they are in the hospital or at home. For these reasons, sometimes the information is not specified in the accident report or estimates are made.

On the other hand, the presence of ***speeding offences*** is often hard to specify. In a research carried out in **Spain** by INTRAS, it was shown that the presence or absence of speeding offences was not proved in 20% of the drivers involved in motorway accidents, neither in 50% of the ones involved in urban area (Chisvert, Monteagudo y Pastor, 1998). By analyzing the most recent accident rate data, is has been noticed that this proportion of lost values has not improved.

As for the ***use of the seatbelt***, in **Spain**, it had been verified that this information did not appear for 6% of the car passengers in motorways and for 30% in urban area (in 2004, these percentages were located at 7% and 39% respectively).

With regard to ***demographic variables***, Austin (1995a) found out that the age was not specified for 4% of the casualties (study carried out in the **United Kingdom**). In **Spain** it was concluded that the percentage of under-recording of the victims' age was located between 0,5% and 13% according to the severity of the injury, the casualty type, the road type and the gender.

As far as the ***vehicle characteristics*** are concerned, Lindeijer (1987) concluded that in 33% of the accident reports, the brand and model of the involved vehicles were not specified, even more in the case of lorries and heavy vehicles. In the same line, Hughes, et al. (1993) pointed out the huge difficulties for the police to identify the brand and model of the lorries, particularly for the articulated ones. In **Spain**, the age of the vehicle was not specified in 8,1% of the accidents collected in motorway and in 24% of the ones collected in urban area (Chisvert, Monteagudo and Pastor, 1998).

An exampled of exploratory study about missing data or under-recording was the one carried out in **Spain** from the accident data collected in the A-7 toll-motorway (Ledesma, Sanmartín and Chisvert, 2000). The analysis focused on detecting fields sensitive to featuring missing data from the inspection of their univariate distributions, and in exploring patterns that might explain this under-recording. This way, the Figure 20 shows proportion of accidents in which such information had not been collected. The fields more sensitive to presenting missing data were the km covered from the origin and the trip duration.

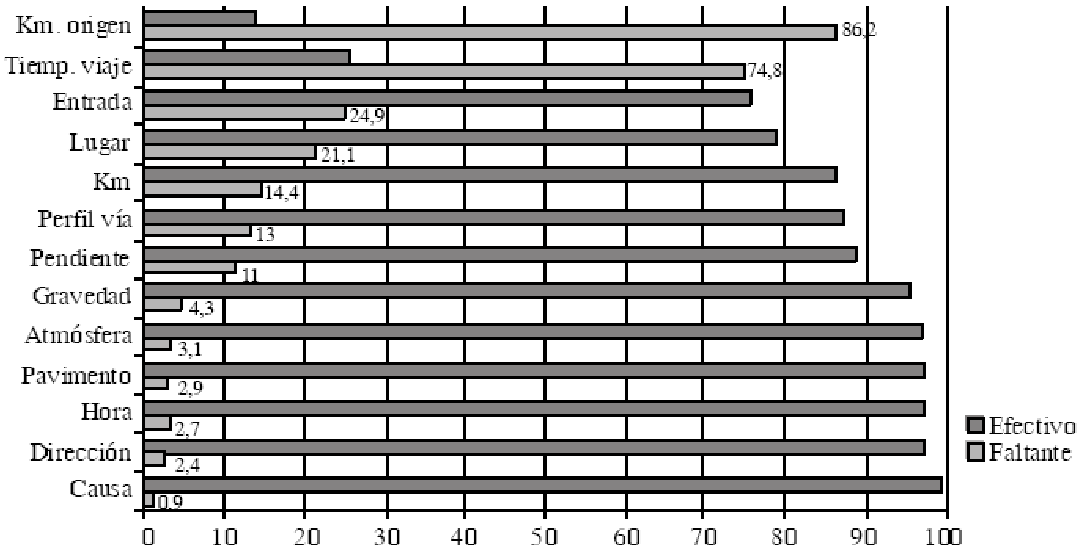


Figure 20: Proportion of the lost data in several variables collected in the AP-7 Spanish motorway.

A more detailed analysis allowed detecting patterns related to, for example, given stretches of the motorway (Figure 21), with given technicians in charge of the collection (Figure 22), or with changes in the data collection routines throughout the time.

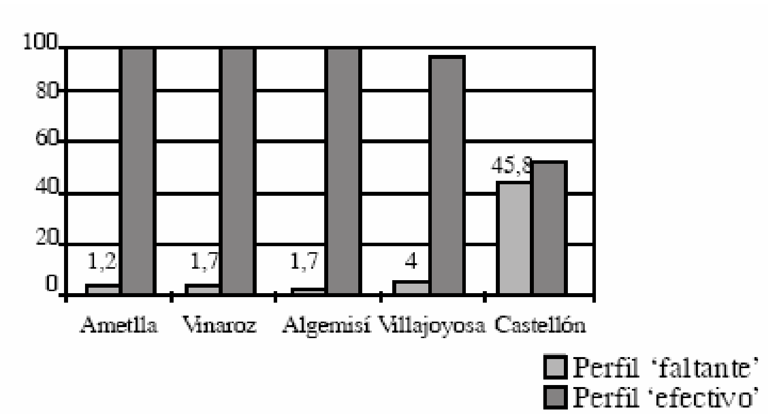


Figure 21: Proportion of missing data in the "Road profile" variable, depending on motorway stretches, in the AP-7 Spanish motorway.

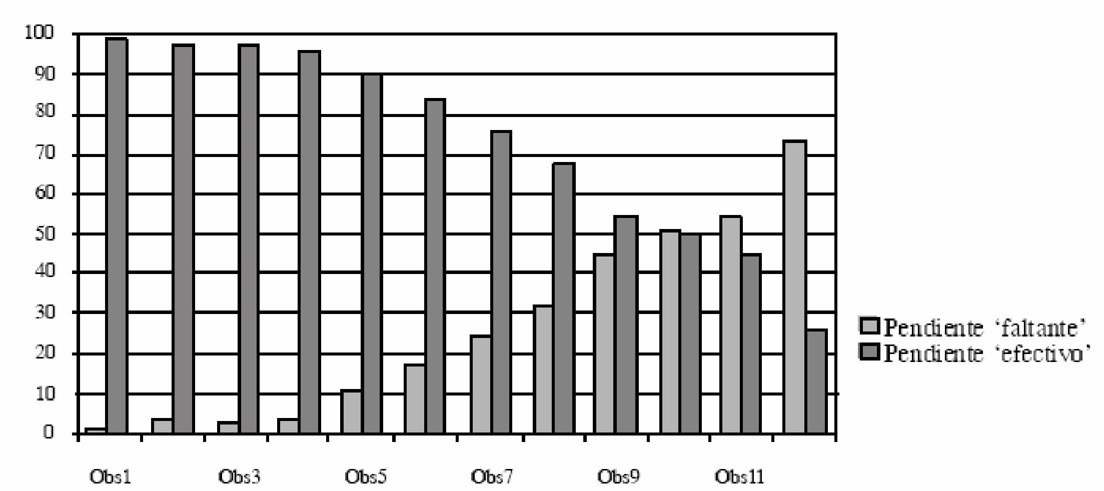


Figure 22: Proportion of missing data in the "Road gradient" variable, for different observers, in the AP-7 Spanish motorway**.**

4.4.3. Errors and biases

The inaccuracy of the recorded values may be explained, among other reasons, by the lack of appropriate specific training of the police forces in charge of the data collection. Hereafter we describe some of the studies focussed on this problem.

The difficulties already mentioned to set the ***severity of the injuries*** imply that this information features a low level of reliability. In a study carried out in the **United Kingdom** (Austin, 1995b), traffic accident casualty police data are compared with the hospital data for the same casualties, reaching the conclusion that the real number of serious casualties was 35% greater than the one pointed out by the police. Similarly and using the same methodology, another study in **Austria** (Rosman and Knuiman, 1994) concluded that 44% of the casualties that, according to the police, needed hospitalization (serious) were not hospitalized, while 31% of the ones that were hospitalized appear in the police reports as being slight (no hospitalization).

This data leads us to think about errors in front of the possibility of a bias, given that there are so many undervaluations as overvaluations of the severity.

In **France**, there are some studies that have analysed the exhaustivity and reliability of the data referred to the severity of the casualties from the surveys carried out to the health staff that attended them. The comparison of this information with the one of the police records has shown that, generally, the agents use to overstate the severity of the casualties. Moreover, they point out that the serious casualty record is much more correct for accidents where there are several involved vehicles than when there are none. Furthermore, in this last case, the record is worse when it is about two-wheel motor vehicles or cyclists, and when the casualty is less serious (Laimon, 2001).

The Transport Department in Great Britain has carried out a study after comparing the STATS19 data, with a sample from the "A&E" (DfT, 2006). They have concluded that the serious injury group could be twice greater than what the STATS19 data shows. This is due to the underreporting level which was explained before. However, they have also detected some mistakes related to the casualty severity classification. They realized that a significant proportion of slight casualties should be classified as "serious casualties" (in order to be roughly 25% more representative).

On the other hand, several studies reviewed by Voas (1993) have shown the presence of reliability problems for the data referred to the ***alcohol consumption***, though without showing any concrete numbers: Delay in the sample collection (the level of alcohol has reduced), intravenous treatments (the level of alcohol has reduced), loss of blood (the level of alcohol has increased) and states of shock (the rate of alcohol elimination varies).

As for the ***age***, Austin (1995b) compares the data about casualties collected by the police with hospital data, identifying errors in connection with the age in 12,1% of the cases. Ferrante, Rosman and Knuiman (1993) use a similar procedure, including death certificate data in addition. This way, they identified age errors in 18% of the cases, a third of which with differences of more than five years.

In addition to the errors, it seems that there is a rounding bias. In the same study, it has been detected a trend to round up to 25 or 30 years old the persons around these ages.

Hugues, et al. (1993) point out that the information in connection with the ***accident location*** is different from the ones that seem to raise problems to most of the agents. Some of them are the great variety of road types, the absence of kilometric reference points at the place of the accident and other information referred to the road, errors in the data transcription (e.g. problems of readability in the names of the streets), and difficulties with a particular type of roads like the junctions or the crossings. Austin (1995a) compared the data of the accident reports related with the road, in a specific area of the United Kingdom, with data collected in a geographical information system (GIS). He found the following percentage of errors: 4,4% regarding the road number, 6,8% in the speed limit, 15,3% on the presence of zebra crossing, 21% in the road type, 12,5% in the junction type and 13,2% in the junction signs. In addition to this type of errors, other studies like the one of Hughes, et al. (1993) identified a strong trend to round up in connection with the kilometric point of the accident, phenomenon also stated by INTRAS in the motorway accidents (INTRAS, 2005).

4.5. Data exploitation

This is a key topic. The effort and cost of the traffic accident statistical data collection and management are only justified when they are useful to detect and solve road safety problems, which undoubtedly pass through an appropriate exploitation of them.

At the national level, the usual model is that the different administrations, national as well as regional and local, collect the data that is centralised by a central administration. This administration produces a series of outputs of the collected data. Normally, they are statistical yearbooks (in most of the EU countries) and/or periodic publications (monthly, quarterly…), which are generally descriptive and that show a picture of the general situation at the macroscopic level.

Likewise, it is usual to carry out more specific studies, in a systematic or punctual way, according to the problems that appear, or to several issues that are raised. Likewise, they can be carried out from a macro perspective (general data on the population), or with a more micro or detailed approach, centred in more specific or local aspects and/or localisations.

Many public and private, national, regional or local organizations or institutions may benefit from these data, through the outputs performed by the administration in charge of their management (in the section 3.7, p. 46 there is a description of the links to several official WebPages where these outputs may be consulted), or by directly accessing the data[5] and carrying out ad hoc exploitations.

To European level, the CARE database has an application to the consultation online of data accidents and casualties. In the following figures, an example of interactive data exploitation in CARE appears.

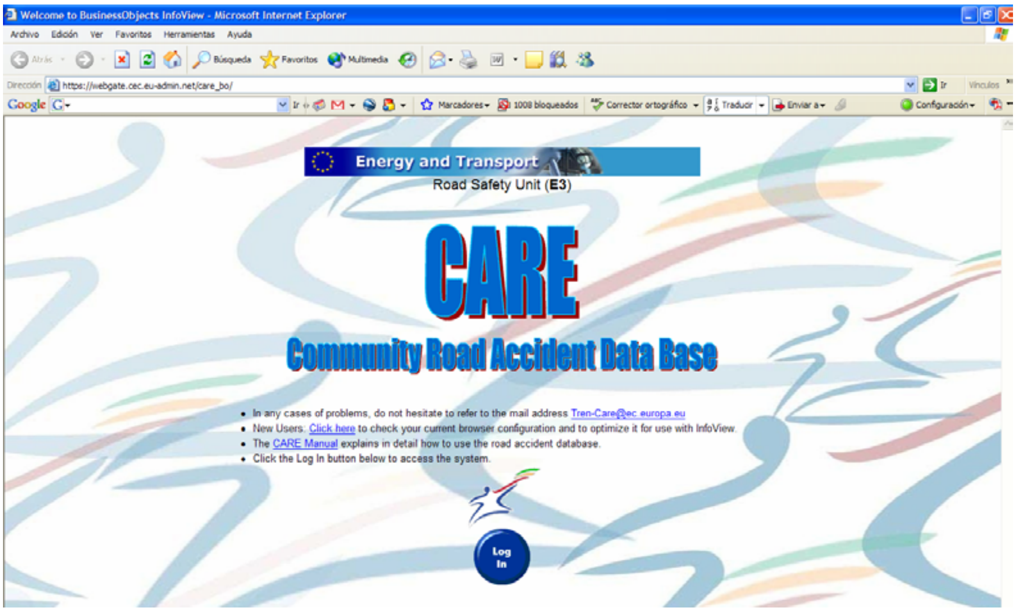


Figure 23: Initial screen of CARE database

[5] Always respecting the confidentiality and anonymous non-identifiable character of the data.

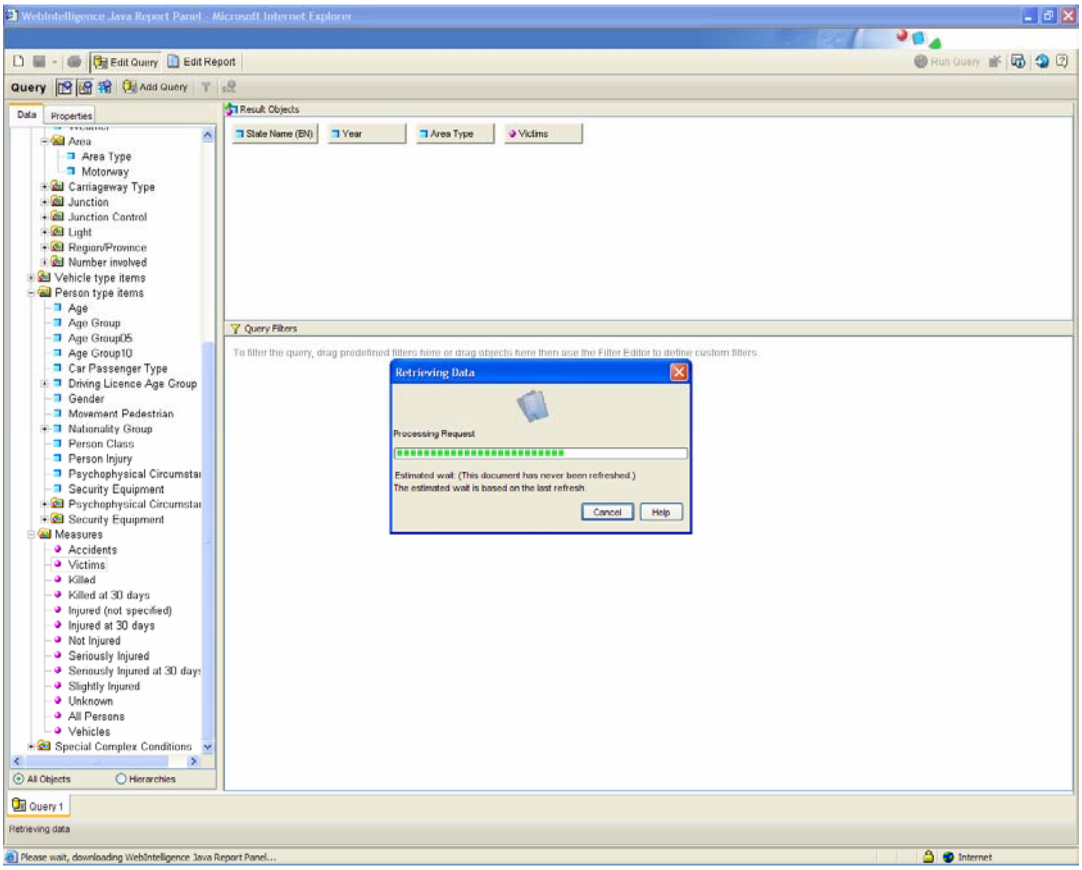


Figure 24: Consultation to CARE database (casualties by country, year and area)

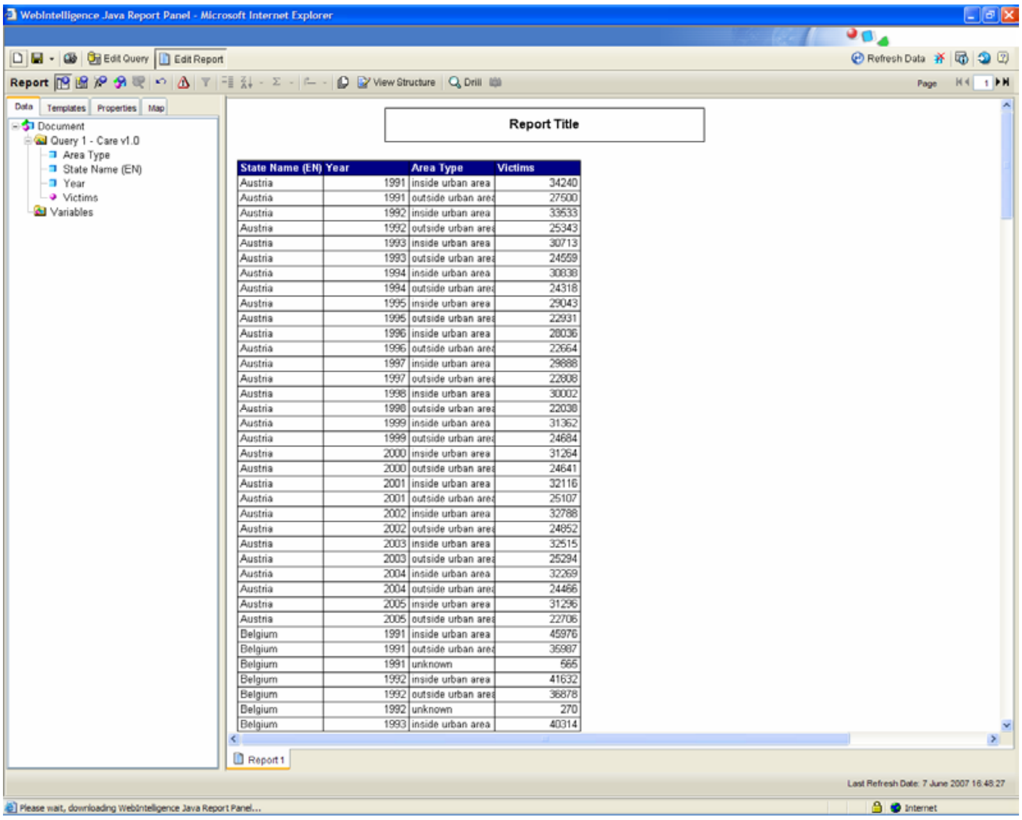


Figure 25: Output of results in CARE

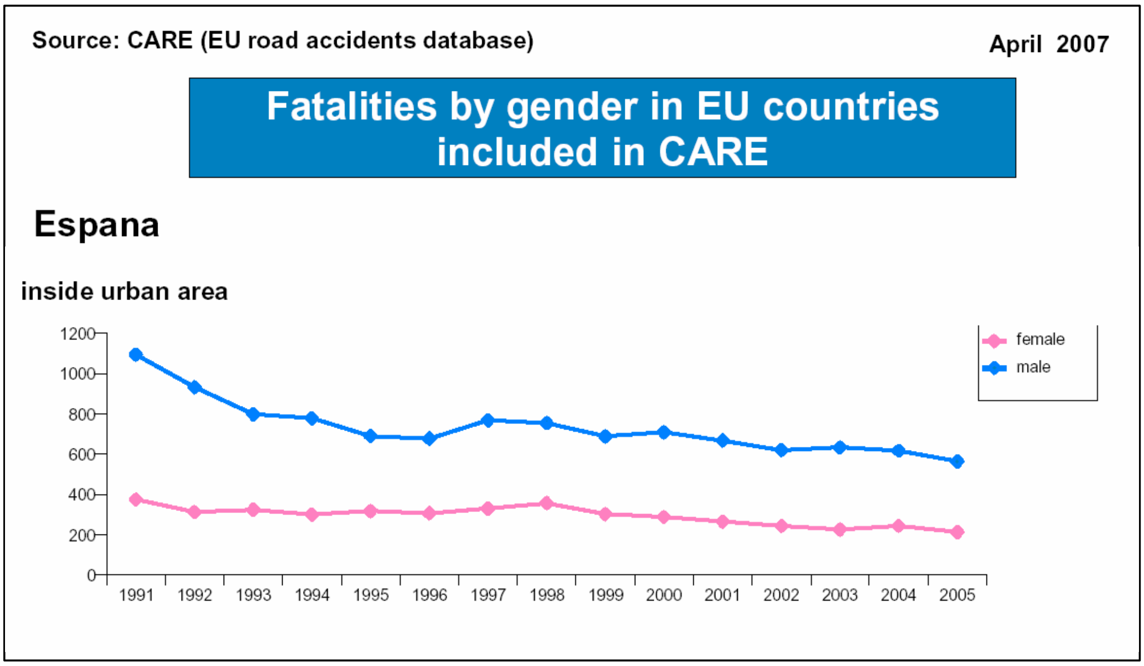


Figure 26: Graphical data representation in CARE

At the national level, an example of interactive data exploitation can be found in **The Netherlands**. In the web of the Institute for Road Safety Research (SWOV), there is the Powerplay tool (Cognos Powerplay Web), that allows working on-line with information selections, creating and modifying tables, creating graphic representations, and carrying out basic operations with the accident data recorded in the national database (Figure 27).

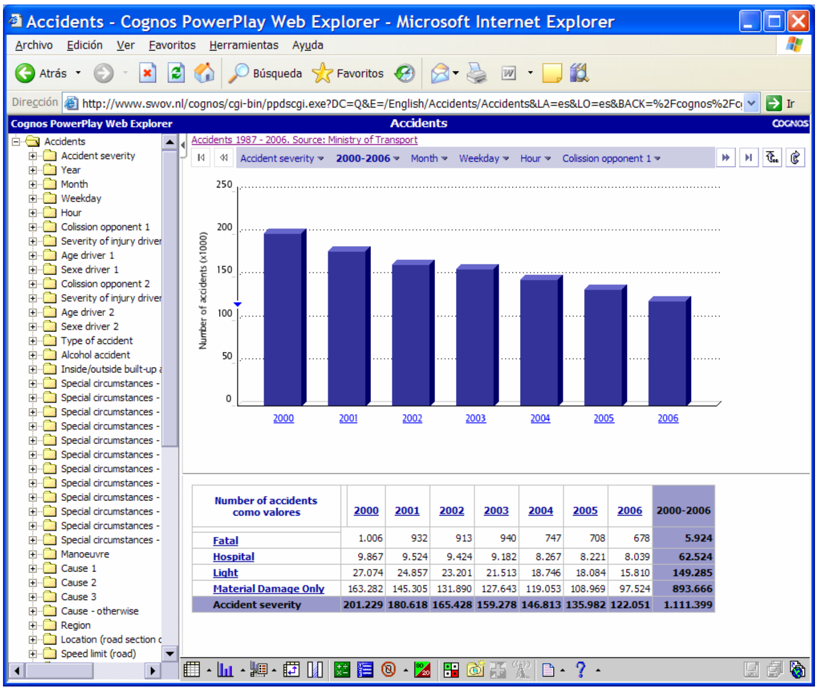


Figure 27: Example of table and graph in the interactive web of Powerplay Cognos

Another example can be found in **Hungary**. In the web of the Hungarian Central Statistical Office (HCSO), there is the KSH Web Statinfo tool, to analyse lots of statistical indicators. In order to visualize the data on traffic accident rate, the path is: “3. Society” – “3.1 Public Health” – “3.1.3. Accidents”.

The application allows creating statistical and graph tables from a selection of variables related with the accidents (road type, accident type or cause…), or with the victims (age, gender, severity of the injuries…). Moreover, it provides specific data of the accidents related with alcohol consumption. In the Figure 28 and Figure 29, it is possible to see some results obtained with this data exploitation tool.

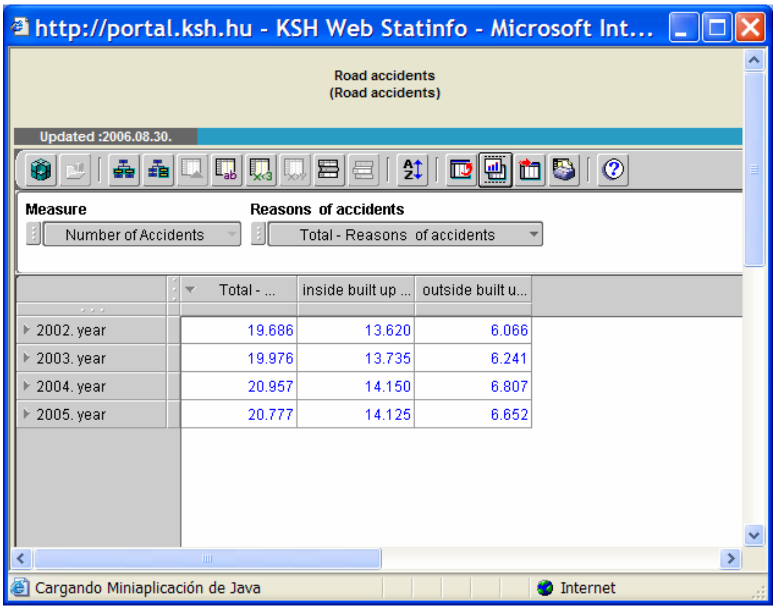


Figure 28: Example of a table in KSH Web Statinfo

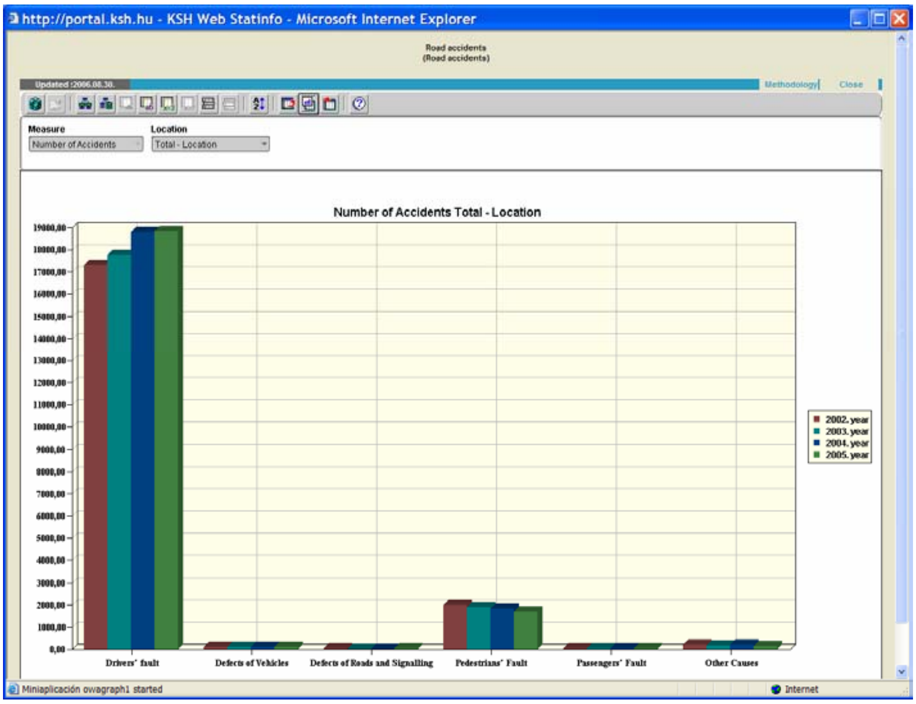


Figure 29: Example of a graph in KSH Web Statinfo

All these organisations have as common denominator their involvement in several activities related to road safety, each one from their own competences (competencies) and objectives. These activities include the more general, systematic or continuous ones (general planning, road infrastructures building and maintenance, urban planning, traffic policy and vehicle safety standards follow up, etc.) as well as other more punctual or specific ones (particular road design, awareness and road education campaigns – alcohol, speed… -, driver training programmes, black spot detection and action, road maintenance, signposting, etc.).

The local administrations may use this data for several purposes. Nevertheless, due to a series of difficulties that are described hereafter, in practice we find out that, either the data is not systematically used, or its use is only limited to following up the main aggregate data evolution – in term of accident frequency – and to using it with the objective of identifying accident black spots.

Among other constraints, this situation is due to:

- **The need of technical and human resources**. The statistical analysis of the traffic accident rate is a multidisciplinary task that may reach an important level of specialisation and complexity. In the local field, it is not always possible to have the sufficient technical and human resources to carry it out with an appropriate level of depth.
- **The need of appropriate tools**. The previous problem may be partly solved with the introduction of "expert" computer systems that provide the information analysis, through automatic procedures of analysis and study. Nowadays in Europe, the central administrations have developed scarce systems of this kind, and the experiences in the local administrations are scarce. Nevertheless, there are relevant positive experiences, like, for example, the case of France, which we will see in the section 5.1.1, p. 97.
- **The data delay**. This is an important aspect that has a great influence in the use of the data from the point of view of the urban road safety management. The inadequacy of many current procedures implies that there is an important temporal interval between the moment the accidents take place and the moment in which the data are available to be studied. For that, several countries have established specific deadlines for the data reception. So, for example, in the **United Kingdom**, the data should be received no later than three months following the end of the month to which they relate, whereas in **Denmark**, the data should be received within 5

weeks after the accident has been collected. In **The Netherlands**, the data should be received no later than the month of March of the following year to which they relate, even though there is an estimation that stated that almost 99 % of the accidents are already recorded within 60 days.

The electronic transmission of the data has also improved this deficiency.

4.6. Measures to improve the data quality

The improvement of the accident statistics quality and of the exploitation methods is a constant concern in many countries, developing specific plans aiming at obtaining faster and more reliable information.

4.6.1. Francia

Measures:

a) Launching of a data collection and management plan, based on quality indicators (decrease of the delay, complete information on the data and consistency) and creation of a quality file (documentation and procedure).

Since 1998, a programme of accident data quality control called SAXO (Serveur Accident sous uniX et Oracle – Accident Server under Unix and Oracle) has been started. Its main functionalities are the centralization of road accident with victims information, control the consistency of such information, allow the correction of incoherencies, create data files, create and spread national results and finally file and manage the history of the treated data.

The system has four types of controls:

- Detection of duplicate record of accidents.
- Control of the record order and structure.
- Control of the out-of-range values
- Control of the consistency between the recorded data in the same accident.

Another type of more informal quality control is based on setting and finding the accidents that have not been transmitted for the centralized record in the accident database. The experience shows that under-notification uses to happen for local problems (among the police forces or during the data transmission) that induce that the reports are not filled in or sent out.

In 2000 a monitoring group of the statistical production was set up, it is formed with officers from the "direction générale de la gendarmerie nationale" and from the "direction générale de la police nationale", responsible for collection the data and managing the accident rate records.

In the modernisation plan of the accident collection system, the attention has been focussed specifically in the improvement of the quality and, for that, a series of control indicators have been set up: the delay in the data spreading, the exhaustive record of all the information fields, and the detection of specific inconsistent or erroneous data in order to correct them.

Other objectives of the plan are the creation of reference manuals to fill in the accident reports and the improvement of the data transmission tracks on the short-term.

b) Simplification of the record system.

The simplification of the record system is focussed on the suppression of the information fields that are not used or that are irrelevant, the modification of others, the creation of new fields that summarize the information of several fields that can be grouped (for example, for the "type of accident" variable they try to develop a library of descriptive schemes of the contexts in which the accidents take place), the use of annexe files to replace some variables, the use of GPS systems to track down the accidents (initially in the motorway and subsequently in urban areas also), and the creation of a simplified accident report for the slight accidents.

The main modifications of the new French accident report (BAAC 2002) that came into operation in January 2004, are the following: suppression or modification of 11 variables, creation of the "GPS data" variable to track down the interurban accidents accurately, extension of the categories on "type of vehicle", improvement of the definition of some variables, introduction of the "drugs" variable.

c) Decentralisation of the accident record by modifying the computer architecture of the data recording system.

It lies in the creation of local databases, available at central level as well, that allow a direct and faster access for the regional-local entities to the available information on the accidents (correction, query and exploitation). It is foreseen that for 2008, it will be possible to have full access to the accident rate record through the new computer system (for a query as well as for the data modification).

4.6.2. Belgium

In June 2004, the federal police launched an action plan to control the quality aiming to correct the detected errors without losing the system efficiency. For that, the quality concept is focussed in the information flow and in the optimal data entry. The current situation presents a series of limitations and problems on which corrective measures have to be established. There are 4 main axes (CFSR, 2007):

1. Data exhaustivity (collection of the accidents and of their attributes). Among the measures to avoid underreporting, it has been raised that there is a need to link the accident databases with the hospital databases. However, the estimate of this problem from the hospital data is impossible for the moment given the different collection methodologies they are using. For that, another option might be the use of insurance companies' databases.

   On the other hand, using the results obtained from the accident in-depth analyses allows a better appraisal of the real causes and the circumstances in which they take place. These qualitative data are a complement to the quantitative data. However, even though there is an institutional will to bet on this type of studies, there exist technical and legal difficulties that impede the data query and handling.

2. Delay in the data availability (how many time is needed for the definitive data entry and for the data to be ready for its statistical treatment).

Nowadays, the complete annual accident data is not available before 4 months, so the statistics are usually delayed until the following month of June. There are administrative problems that make difficult the fast access to the database for a statistical purpose.

3. Data uniqueness (to avoid information duplication).

4. Collected data accuracy.

Concerning the problem of the fatality within 30 days data, up until now the police is doing the appropriate checks and sends the data through a manual form. In the framework of the "Optima" study and of the study on "Road safety data exploitation", it has been proposed to use the whole data flow of the police services. This requires clear agreements between the hospitals, the Public Prosecutor's Office and the police.

On the other hand, we have to pay attention to the specific training of the police officers to fill in the accident form. At present there is not any refresher training regarding this subject which is directly affecting the collected data reliability.

Finally, there is not any clear definition regarding the casualty severity. This lack of specific criterions to differentiate between an unhurt person from a slight injured or seriously injured one, causes different interpretations among the persons in charge of codifying the information.

In order to improve these aspects, the federal commission for the Belgian road safety has implemented several recommendations. The main one, in terms of road safety information quality, is focussed on the urgent development of an expert accident collection system, based on a single data entry, aiming to be able to have reliable and complete accident data, rapidly available to be analysed. The data exploitation has to be the most optimum as possible. This is expressed in the following objectives:

- To carry out a road safety barometer from the monthly accident count.

- To give priority to the development of the "Pol Office" data integration computer system (section 5.1.5, page 113), both for its generalized application in all the police services and for the creation of official statistics in the best terms.

- To make available the Statistical Directorate of the SPF Economy data to the registered users automatically and freely. This procedure implies the amendment of the privacy contracts in terms of individual data transmission.

- To develop a controlled training on the Pol Office system securing the required attention to the aspects in relation with the proper collection of accident data.

- To publish, on the Intranet, a guide to the data entry. The objective is to unify the information entry and coding procedures.

- To implement necessary measures to guarantee a proper record of the accident exact location.

- To assess the level of *underreporting* for the accidents with casualties and put forward corrective measures.

- To look for and use existing information sources that allow complementing the data of the accident databases (considering the features of the hospital collection systems, in principle its use is excluded in the short term). For that, the central data bank structure has been widened.

- To define the information flows related to the fatalities within 30 days data, from the hospital to the official database.

- To approve the changes and contributions coming from the AGORA study, on the information of the new accident collection system.

- To create an inventory of risk exposure data available in Belgium. This type of data is highly useful to analyse accident data.

- To set up clearer definitions to identify the severity of the casualties, fitting with the criterions defined at the international level.

4.6.3. Great Britain

The quality control plans on the traffic accident collection system in Great Britain is focussed on three main work areas. On the one hand, there is the continuous review of the accident data collection tool (STATS19) in order to adjust it to the statistical needs, on the other hand, there is the evaluation and correction of the wrong collected data, and finally there is the evaluation of the sub-notification problem and the implementation of measures to control it (in the section 4.4.1, p. 66 some results of this evaluation are described).

Review of the STATS19 collection system

Each five years, a review of the STATS19 collection system is carried out. The objective is to see that it provides enough information to the government, minimizing the workload of the local authorities and the local polices in charge of collecting data. The sixth quality revision was carried out in 2002 by several working groups led by the Standing Committee on Road Accident Statistics (SCRAS), an independent adviser, and by the director of the quality programme of the National Office of Statistics (DfT, 2006). The consulted organizations were divided in different groups: Police Forces, Road Safety Organisations, Policy Divisions, Scottish Executive, Local Authorities and associated bodies.

The report sets a series of advised changes in the STATS19 accident data collection and processing system, and improvements in the spreading and access to the accident statistics. The aspects that are taken into account in this review are the following ones:

- Evaluation of 1997 review changes. Were they effective?
- Concerns about current data reporting and coding practices.
- Casualty reduction targets and proposals for changes in the severity definition.
- Proposals for the addition of new variables and values.
- Proposals for the deletion of under-used variables and values.
- Proposal to formally adopt the collection of contributory factors.
- Data linkage with health, crime and socio-economic statistics.
- International perspectives and commitments.
- Availability and presentation of the national road accident statistics.
- Data protection.
- Standard formats for the data transmission from STATS19 to the organizations in charge of it statistical treatment (DTRL/SE/NAW[6]).

[6] DTRL=Department of Transport, Local Government, and the Regions; SE= The Scottish Executive; NAW= The National Assembly for Wales

Collected data quality

The police (and sometimes the local road authorities) process the STATS19 data in order to electronically transmit it to the DfT, in a way that the information could be added to the national database of accidents with casualties. The "STATS21" document sets a series of minimum validation checks that have to be applied to the collected data (and the actions related to its correction), before they are sent to the DfT.

Once this organization has received the data, the DfT applies its own validation rules, and sends back any record with errors or suspicious values.

Finally, before carrying out the annual statistics, the DfT makes a last quality evaluation, ensuring that the number of files fits in with the total number registered by the data suppliers.

The validation system establishes two types of errors: the serious ones (structural errors or errors detected in key variables) and the slight errors (ranking errors or inconsistencies with other variables)

According to the nature of the errors or to the amount of information to correct, the complete corrected file will have to be sent back again, or it is asked to make the appropriate modifications through an on-line system. If the errors persist, an analysis is carried out as well as an adjustment of the local validation procedures. An example of the computer validation codes may be looked up in the Scottish version of STATS21[7] (Government Statistical Service, 2005).

4.6.4. Spain

In the Spanish road safety strategic plan (DGT, 2005), a series of action areas are set. Specifically in the road safety research and analysis area, a series of actions aimed at improve the data collection and processing systems is planned:

- Development of the ARENA system for the traffic accidents computer record and storage, including the training programme for the users.

[7] The validation code used by the DfT is not open to the public, in order to protect the information electronically transmitted (given that it includes the system format code).

- Evaluation, improvement and development of the coordination/interrelation between the different databases (hospital, police, forensic experts, insurance, road, etc.) that allow obtaining better quality information.

As for the urban area, the same plan sets the development of specific municipal plans. These have to include, among other actions, the accident rate analysis and the setting-up of indicators to assess its evolution and comparison with other municipalities. This way, in the action and objective decalogue, of the urban safety type plan (DGT, 2007), it has been included the study of the mobility and the urban road accident rate from the implementation of monitoring systems in order to improve data collection and analysis. Likewise, among the priority actions there is the implementation of a municipal traffic accident statistical record, and centralized it in a unique database.

However, for this to be feasible, the local police authorities have to collect accident data and fill in the municipal reports, transferring them to the official statistics (through autonomic or state institutions), premise that is not fulfilled in many cases because there does not exist any systematic and scientific data collection municipal method (it is estimated that between a third and half of the real accidents are not detected by the Traffic General Directorate, sometimes because they have not received enough data from the local police).

In this line, some autonomic communities and municipalities have taken into account this issue when developing their road safety plans.

So for example, in the Autonomic Plan of Galicia (2006-2010), the implementation of a computer application to homogenize the accident data collection done by the police has been raised (Xunta de Galicia, 2006). In the Autonomic Plan of Navarra (2005-2010), they have proposed the maintenance and improvement of the local accident database (Government of Navarra, 2006). Among the objectives of the Autonomic Plan of the Basque Country (2003-2006), it is stressed on the improvement of the quality of the available information, through the establishment of collaboration and complementary information exchange agreements with all the involved agents, the integration of accident rate information in interurban and urban roads, and the development of an information system that guarantees the knowledge of the final consequences of an accident on the health of the casualty (Basque Government, 2003). In the Road Safety Catalan Plan (2005-2007), it is insisted on the improvement of the quality of the basic accident rate data, which use to come from very scattered sources (Servei Català de Trànsit, 2005). The efforts of this Community as for the improvement, modernization and computerization of the accident collection systems will be explained in detail in the Deliverable II: Cases study.

On the other hand, in the Municipal Plan (2007-2010) of Oviedo (Asturias) a plan to operatively modernize the local police has been set, giving them the necessary technical means for: a) a better collection and subsequent information exploitation on road death rate, b) a better internal management that would free the staff dedicated to administrative tasks for the moment, c) a better accident investigation and creation of Police Technical Reports, and d) a more agile and effective communication with the Traffic General Directorate (Municipality of Oviedo, 2007). The Municipal Plan (2006-2009) of Donostia-San Sebastián (Basque Country) insists on the theoretical-practical training of the agents that act in the accident zone, to improve the data collection and analysis procedure (Municipality of San Sebastián, 2006).

5. Urban Accident Analysis Systems

There is an unquestionable interdependency between the national accident data system and the different local collection systems. The traffic accident data collection procedures, in the local areas, are seldom used for their own and exclusive purpose, being usually limited to the larger municipalities. In general, the urban accident rate analysis at the local level uses the procedures and infrastructures already implemented for the national data collection. This may be effective, roughly speaking, in two ways:

1) The data collected locally may be locally used before being sent to the national record. This may be done, for example, by creating a local accident database, with the national accident report format, and entering the data in it.

2) Using the national database data corresponding to each particular city, once they have already been entered in that database.

The type and objectives of the urban accident rate analysis, as well as the countermeasures that may be derived from and the implementation mode of these ones, feature significant differences with the national procedures.

Generally, the different actors involved both in the data collection procedure and in the analysis, design and implementation of countermeasures show a greater cooperation level in the urban or local scope, given that they all operate in the limits or in the scope of the same city. Consequently, the relationship between the accident rate analysis and the implementation of measures to reduce the accidents is far more efficient at the local level than at the national one.

In this line, some local administrations have been designing and implementing their own systems to collect, enter and store the data coming from the traffic accident police investigation. As it has already been pointed out previously, several local administrations haven been designing and implementing their own systems to collect, enter and store the data coming from the police investigation on traffic accidents. The quality and complexity of theses systems and, therefore, its usefulness from the point of view of the accident rate statistical analysis, depends on aspects like the available economic and technical resources, the amount of accidents to record and the technological developments in connection with the data management systems.

These local systems, even though they are useful given that they allow a monitoring of the accident rate and its descriptive study, frequently feature some problems and limitations.

So, in many cases, the developed systems are focussed on the data entry and storage, but they have very limited means from the point of view of its statistical exploitation. Generally, the systems only consider the creation of some summary tables and data queries, being necessary to export the data to specialized statistical software for a deeper analysis. However, in the local administrations they do not use to have an appropriate technical staff to carry out this task.

On the other hand, another usual problem is the performance of duplicated tasks. In most of the cases, to record the information in their own databases, as well as fill in the local accident form, they have to fill in the national accident form (which has to be sent to the central administration). This task and format duplication implies frequent disadjustments among the data handled by the different sources (local and national ones).

For these last years, several countries have been developing computer applications or systems aimed at accident data analysis that try to solve the mentioned problems.

Sometimes the applications have been developed by the central public administrations for a local use (e.g. AURORE, CONCERTO, COPRA and PROCEA systems in **France**). With this practice it is possible to optimize the profitability of the application development, as well as to apply standardized homogeneous criterions from the different local administrations, with the resulting easiness This way, for example, the accident analysis system developed by the Transport Department of **United Kingdom**, as part of the national accident database (following the structure of the STATS19 questionnaire), allows to fix the level and severity of the accident rate recorded locally (cities, villages,…). Moreover it allows that the local authorities might investigate given problems that cause such accident rate, as well as to assess specific intervention measures.

The **Danish** Road Administration has developed several accident analysis systems for urban and interurban areas. In urban area, specific computer tools are used depending on the size of the municipalities (*Black Spot on PC, ROADMAN…*). The GIS tools are usually applied to the larger municipalities, given that they use to collect and store a great amount of information on housings, streets, accidents, traffic flow, lighting, etc.

Several research centres have also developed and distributed this type of systems o tools (Frantzeskakis, 2000).

In the **United Kingdom** we find the case of the application MAAP carried out by the TRL (section 5.1.2). In **Germany**, several states are using computer information systems like NIVADIS (in the Lower Saxony) or the EUSKa system, developed by the German Institute of Traffic Engineering (section 5.1.3). In **Italy**, the Civil Engineering Department of the University of Brescia, in collaboration with the local authorities, has developed a system for the accident data analysis and its location based on geographic information systems. The **Greek** system has been developed by the Department of Transportation Planning and Engineering of the National Technical University of Athens, in collaboration with the Ministry of Public Works. In **Austria**, the KfV (Kuratorium fur Verkehrssicherheit) started in 1995 the accident analysis system called UNDAT (Unfalldatenbank – Accident Database). In the **Czech Republic**, the PVT Litomerice, in collaboration with technical experts from the National Road Administration, has developed a software to identify black spots. In **The Netherlands**, the AVV Transport Research Centre has developed applications to depict black spots, to present data in Internet and to develop standardized technical reports (Figure 30).

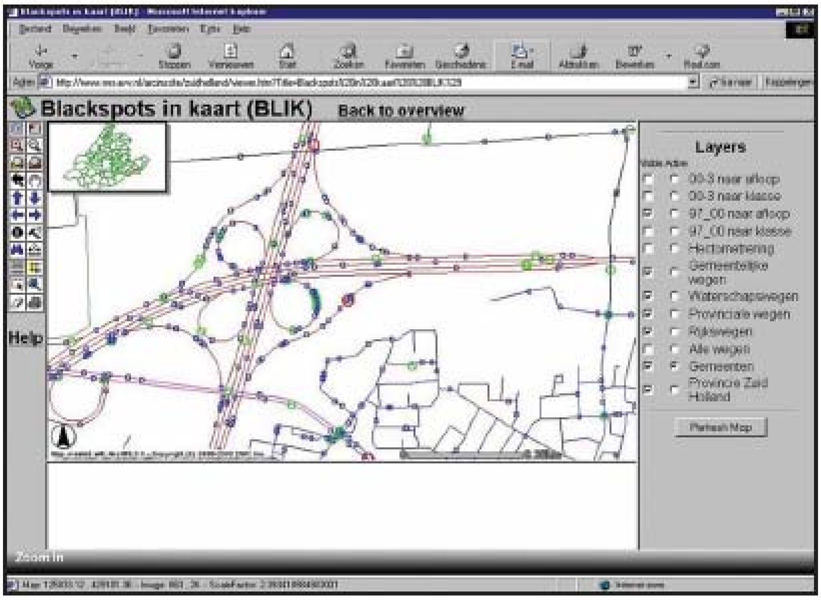


Figure 30: Application to depict black spots in The Netherlands

Some software companies have also developed and commercialised applications of this type. An example is the TIES software developed in the **USA** by the CISCO's *Safety Software* company. In **Sweden**, Aerotech Telub (SAAB group) has developed the *STRADA* system (Swedish Traffic Accident Data Acquisition). In **Belgium**, the Computer Science Corporation company (CSC) took part in the development of a computer platform that integrates the traffic accident data: the Pol Office system.

Finally, sometimes, especially in the larger municipalities, it is the local administrations themselves that raise the development of new accident data management and analysis systems, turning to being assessed by experts (universities, research centres and software companies) many times. A good example would be the PACTOL system developed by the Urban Community of Lille in **France**.

5.1. Some examples

Among these examples we want to emphasize – for their generalization and potentials in the field of urban accident rate – the set of applications developed in **France** (AURORE – PACTOL – LISPACTOL – CONCERTO), the MAAP system (*Microcomputer Accident Analysis Package*) developed by the TRL in the **United**

Kingdom, the EUSKa system developed in **Germany**, the *STRADA* system used in **Sweden,** the *Pol Office* system lately developed in **Belgium**, and the SEWIK system in **Poland**.

5.1.1. FRANCE: AURORE–CONCERTO, PACTOL (COPRA) Y LISPACTOL (PROCEA) [8]

In **France**, the creation from the central administration of standardized tools for the urban accident data management and analysis at the local level, and its spreading among the municipalities, has been a usual practice since mid-eighties.

The AURORE system (*Accidents Urbains sur Ordinateurs*) has been specifically designed to manage and analyse accident rate data in urban areas. The system had been developed in France by the DSCR (*Direction de la Sécurité et de la Circulation Routière)* in 1988 and is nowadays managed and updated by CERTU (*Centre d'études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques*). The latest version (AURORE 3.5) is from 1997.

The system has been developed for the entry and analysis of data coming from the BAAC (*Bulletin d'Analyse des Accidents Corporels*), that represents the protocol or *accident report* used to nationally compile accident data. All the data compiled in the BAAC and locally entered in the AURORE system are alphanumeric. Added to data entry and management, the AURORE system allows standardized local exploitations of the data. The analyses that are incorporated into the system have a simple descriptive feature: case selection, counting, crossed tables… Eventually, specific softwares developed locally allow carrying out some simple statistical tests.

The AURORE system has been the database on which the greatest part of the local traffic accident records have been managed in France.

The CONCERTO application (interurban network) is the latest developed tool to analyse accident rate in France. It has been developed by CERTU and SETRA (*Service d'Etudes Techniques des Routes et Autoroutes*) on the initiative of the

[8] In CERTU (2004) the characteristics of these French applications are described with more details.

ONISR (*Observatoire National Interministériel de Sécurité Routière*). The objective of this application is to progressively substitute the AURORE software as management and analysis tools for the data coming from the BAAC.

CONCERTO (the version 1.7 is from 2006) is a tool that incorporate a GIS (Geographical Information System) application, added to a statistical analysis module that is more complete and sophisticated than AURORE. It has been conceived to manage a local accident database of which data entry is looked after by the police forces. CONCERTO is fed on data coming from the BAAC, being directly entered or imported from AURORE or from PACTOL (see next page).

Regarding the data exploitation, it enables the creation of different tables (univariate and/or bivariate), trend studies, as well as to calculate different safety indicators. This information may be completed with the accident data display and spatial viewing. To show this spatial dimension, CONCERTO uses the road network as a geographical referent. It is possible to directly choose accidents by selecting them in the map, or indirectly, through available geographical objects (municipality, junctions, public road, school, bicycle paths, etc.).

If there already exists a SIG application in the municipality where CONCERTO is going to be implemented, the system has foreseen data exchange procedures with the mentioned SIG application.

Moreover, the system may be completed with data in connection with the features of the infrastructures and the urban environment, and make selection in the analyses according to this information: e.g. study of the accidents in school zones, in roads having bicycle path, in roundabouts, etc. The selection may be directly done on the map manually; depending on alphanumeric search criterions or by combining these two procedures. Alphanumeric data exchange is also considered in the EXCEL environment, so greater analysis possibilities are added.

The PACTOL application (*Procédures d'Accidents Corporels Traités par Ordinateur à Lille*) represents a computer tool to develop and manage information on urban accident **reports**. The system was developed and put in practice in the Urban Community of Lille, even though the objective was a standard use across the country.

The problem of the report privacy is solved by means of the creation, for each record, of a version in which there is no identification data.

The type of information that is entered is both verbal (comments, descriptions) and graphic (accident sketches and graphs). Therefore, the main advantage of the system comparing with AURORE is that it gives more detailed information, available in the reports, and very useful, with the aim of having local diagnostic research.

Nowadays, the French Ministry of the Interior has proposed a progressive replacement of PACTOL by the PROCEA application. The objective is to have a standardized tool to be used all over the country, avoiding the excessive development of software having the same purpose at the urban level. This application was internally developed by the Compagnies Républicaines de Sécurité (CRS) of the National Police, and adapted to the features of the urban area. In 2005, different French police stations started using it.

The analysis and exploitation of the data collected through the PACTOL or PROCEA system is carried out by the LISPACTOL (COPRA) [9] application. This application also allows the automatic creation of the BAAC from the data entered in PACTOL (or in PROCEA), as well as the direct transfer of the summarized alphanumeric data (the ones that are considered in the BAACs) to the AURORE and CONCERTO software.

The LISPACTOL (COPRA) application has been complemented by different utilities that enable the accident data spatial viewing, as well as the interconnection with other data files like the infrastructure inventories, traffic data, speeds, etc.

---

[9] Nowadays there is a new application for the data exploitation more advanced that PACTOL (and also for the PROCEA data) called COPRA (the latest version of 2006 is 1.10, however, the version 1.20 is already being developed), which objective is to replace LISPACTOL.

5.1.2. *UNITED KINGDOM: MAAP (Microcomputer Accident Analysis Package)*

MAAP represents an integrated analysis system, made up of several work tools that allow both accident data management and the information statistical analysis and geographical viewing.

It has been designed by the Overseas centre of the TRL to be used by the police forces and by the administrations and organizations in charge of traffic and road safety. Nowadays it is being used by a great number of central and local administrations in the United Kingdom and in other countries.

The software is customized for each administration that requests it (language, table fields, information encryption…) so it depends on the format of the accident report used in each place. The data is stored in Access or in another database programmed in SQL language and structured in three tables: general data, vehicles and occupants.

The system shows a data entry screen in which lots of validation criterions are applied (Figure 31).

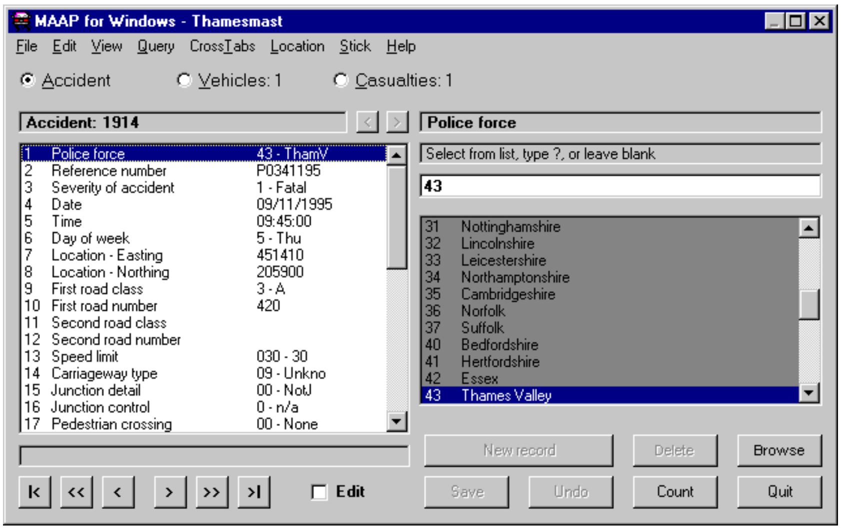


Figure 31: Data entry screen of the MAAP application.

The system enables to make a selection of accident cases or groups, vehicles or users, by using the software commands or, for the more advanced users, with SQL complex queries (Figure 32).

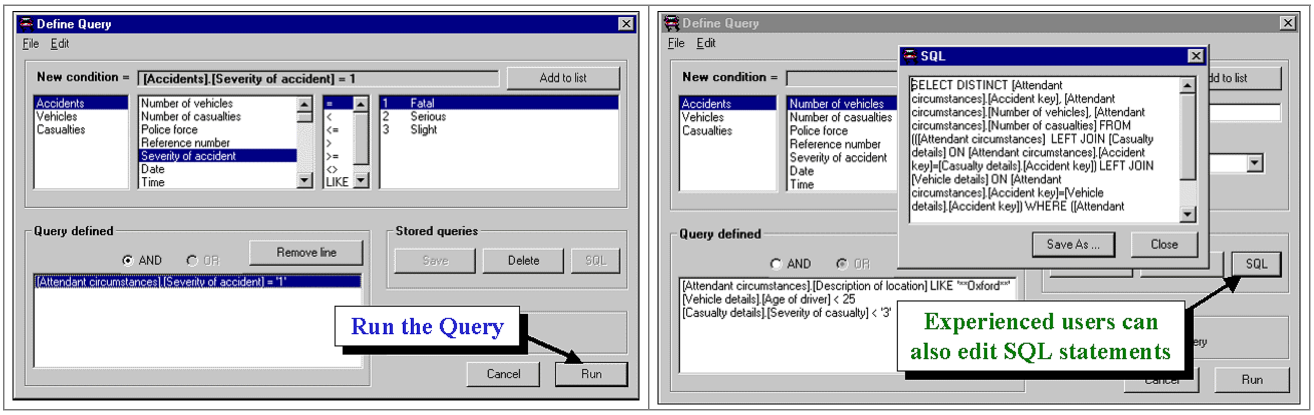


Figure 32: Case selection in MAAP.

The analysis module allows the flexible development of crossed tables of two or more variables according to the user's needs, as well as several types of format for the presentation of results (Figure 33).

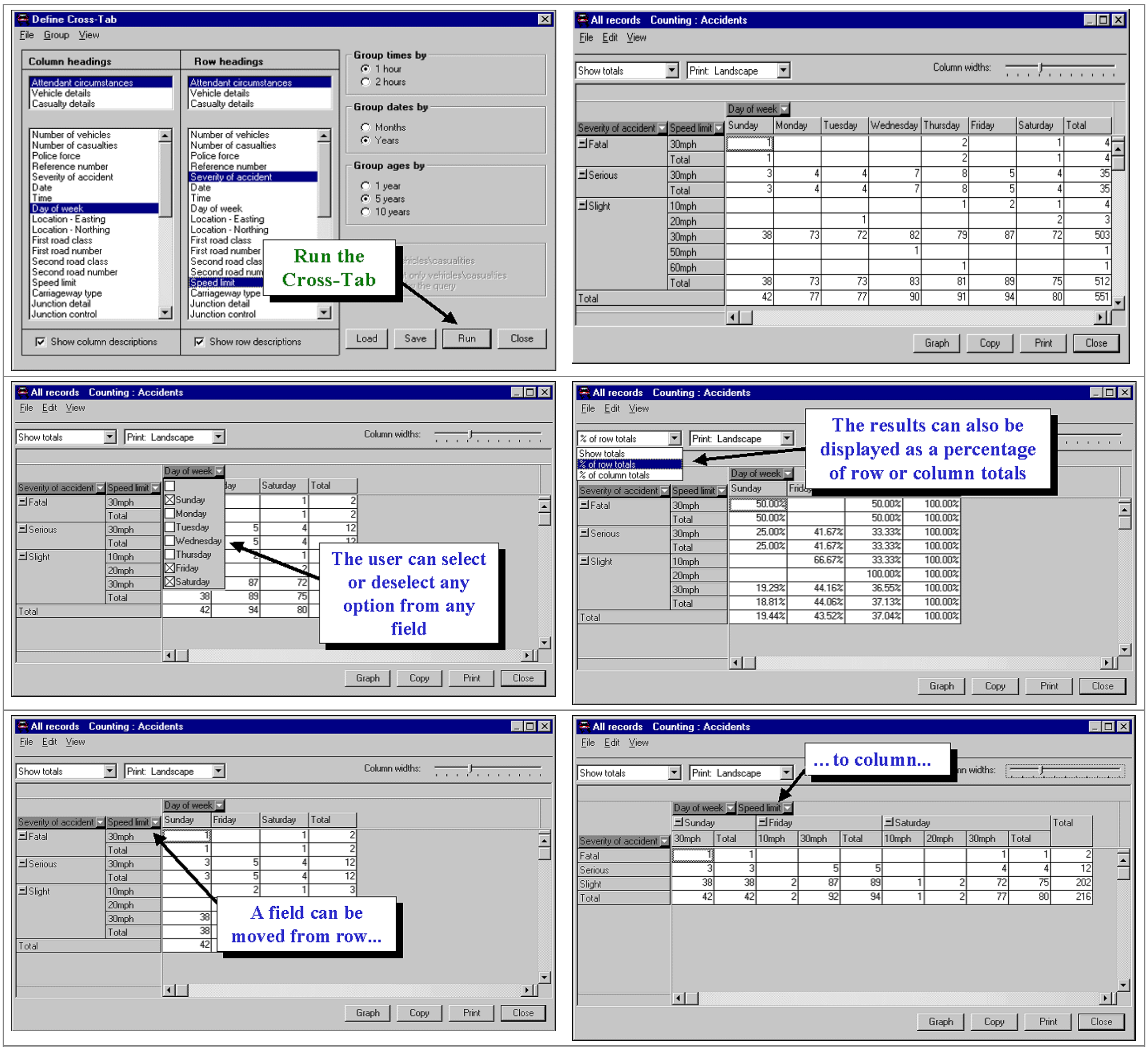


Figure 33: Examples of crossed tables in MAAP and presentation of results

It also integrates a graphic module to display the results of the analyses requested by the user (Figure 34)

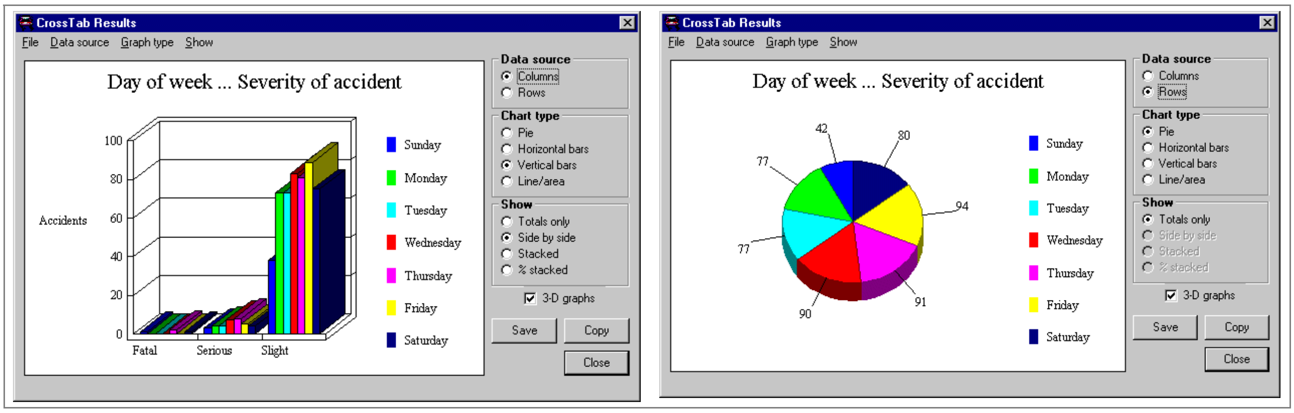
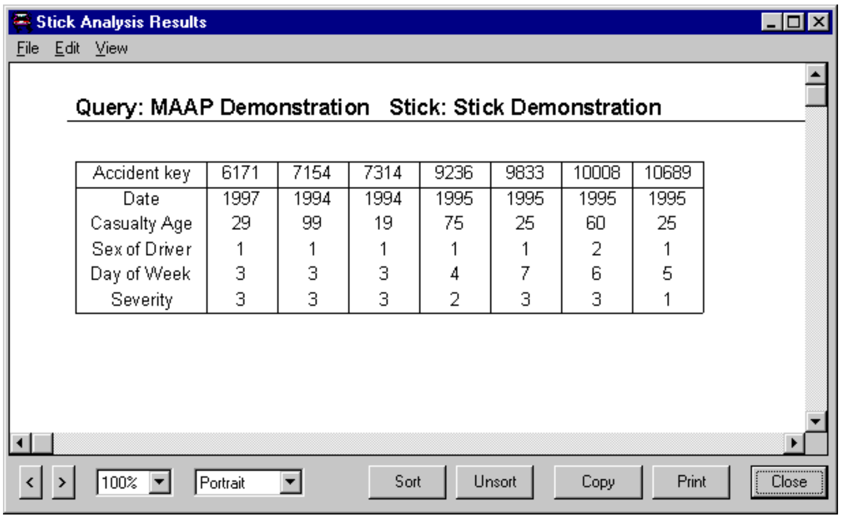


Figure 34: Graphic display of the results of the statistical analysis in MAAP

Moreover, there is the possibility to isolate specific cases, to study them in detail aiming to identify the patterns, as shown in Figure 35.



Accident key	6171	7154	7314	9236	9833	10008	10689
Date	1997	1994	1994	1995	1995	1995	1995
Casualty Age	29	99	19	75	25	60	25
Sex of Driver	1	1	1	1	1	2	1
Day of Week	3	3	3	4	7	6	5
Severity	3	3	3	2	3	3	1

Figure 35: Selection and viewing of specific cases in en MAAP.

Finally, it also enables the spatial analysis of the information, given that it has a geographical information system with several possibilities to select and analyse data (Figure 36).

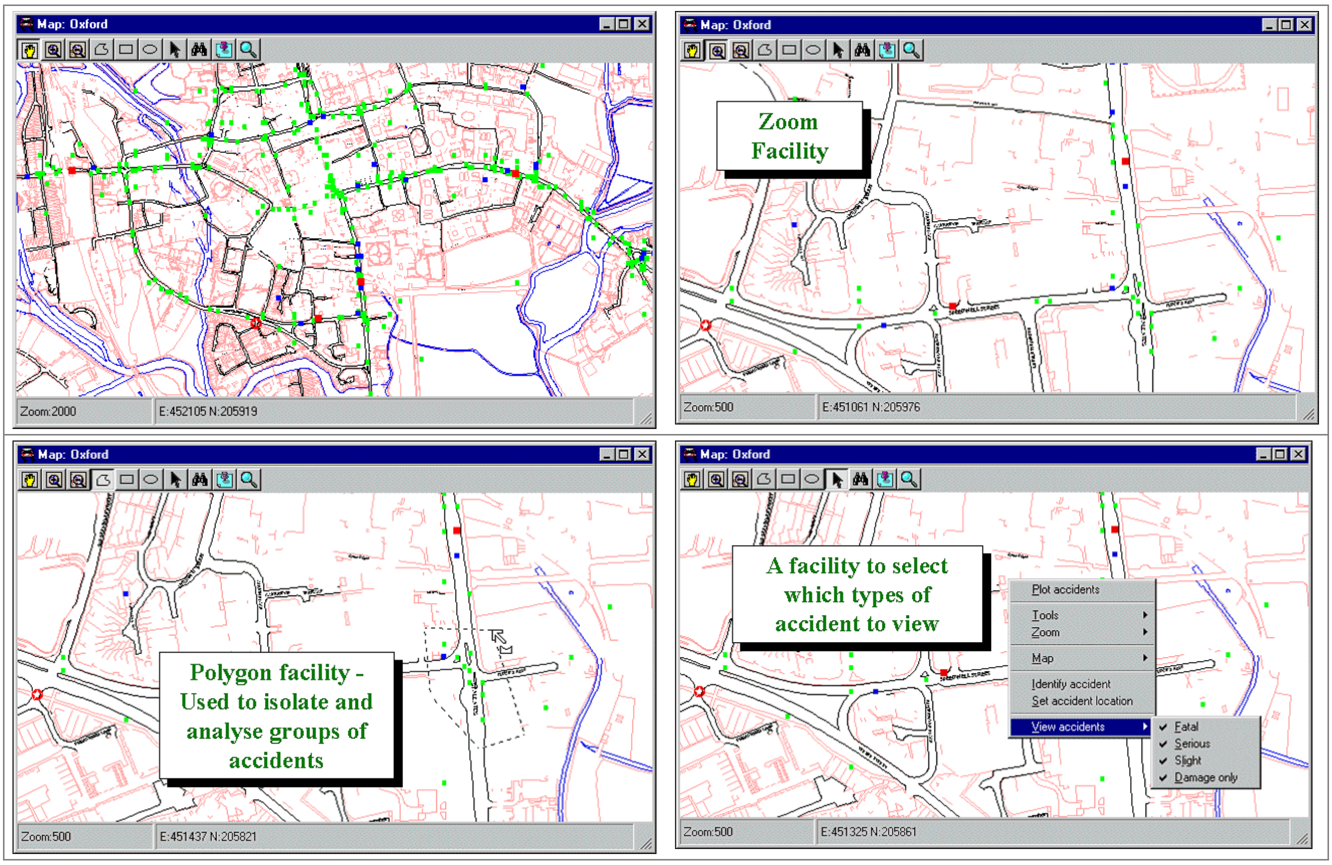


Figure 36: Geographical information system in MAAP.

5.1.3. GERMANY: EUSka[10]

EUSka is accident data collection tool developed by the German Institute of Traffic Engineering. This Windows-based software allows making easier for the police officers the accident data collection. The control and correction mechanisms ensure that all the information has been entered before the record can be made official. The system includes the necessary variables for the national accident database. Moreover, it uses geographical information systems to display accidents in specific maps of the municipalities.

All the police stations of five states are now using this software (Thuringia, Baden-Wuerttemberg, Saxony-Anhalt, Saxony and Hesse). This year, North Rhine-Westphalia will start using it.

The user may select any accident in a digital map (Figure 37) and get a complete report on that accident.

[10] More information at http://www.ptv.de/cgi-bin/traffic/traf_euska.pl

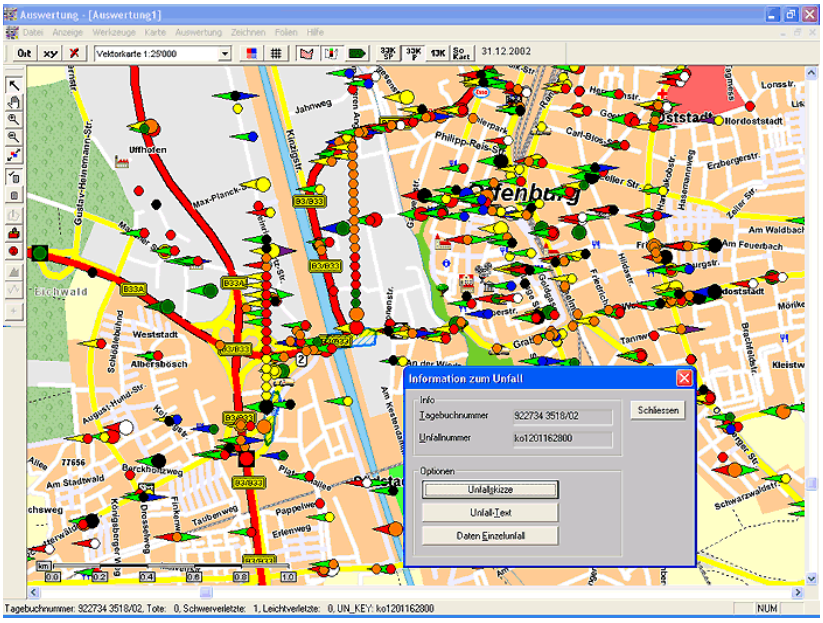


Figure 37: Geographical viewing of accidents in EUSka

Moreover, in the same place, it is possible to analyse both the frequency of the collected accidents and to compare different accidents (Figure 38).

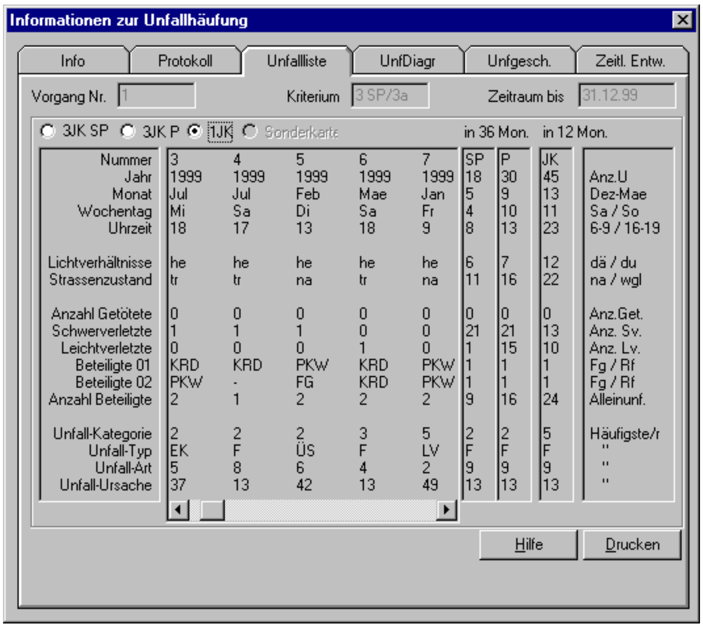


Figure 38: Report on the accident characteristics in EUSka

5.1.4. SWEDEN: STRADA (Swedish Traffic Accident Data Acquisition)

Four institutions were involved in the work team that developed STRADA with the aim to provide a reliable national system for recording accidents: the Swedish National Road Administration, the cities, the police and the hospitals. The system started to be introduced in 2003 (De Mol and Boets, 2003).

Its main task is to combine police and hospital data before they are sent to the national database. Once the records are complete and corrected, they are sent – encrypted - through the Internet.

The police records use three types of information. To locate the accident, they use geographical positioning systems (GPS) that supply precise information on the place of the accident. Moreover they carry out a description of the accident (features of the accident, the vehicles and the involved persons) and they finally carry out a road description (Figure 39 y Figure 40).

The police report data entry is carried out in a computer placed in the police vehicle. This way the data can be directly processed and subsequently sent through the Internet. The GPS system of the vehicle shows the precise location of the accident. Other information fields are directly filled in to ease the collection (e.g. the day, the time or the person in charge).

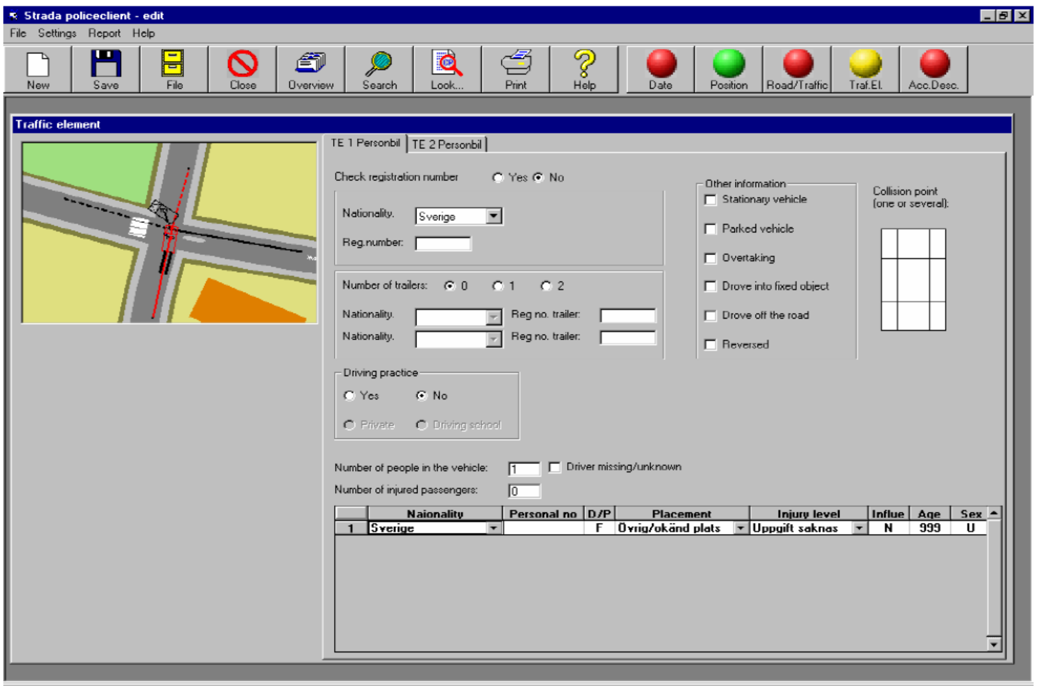


Figure 39: Description of the accident and of the involved persons in STRADA

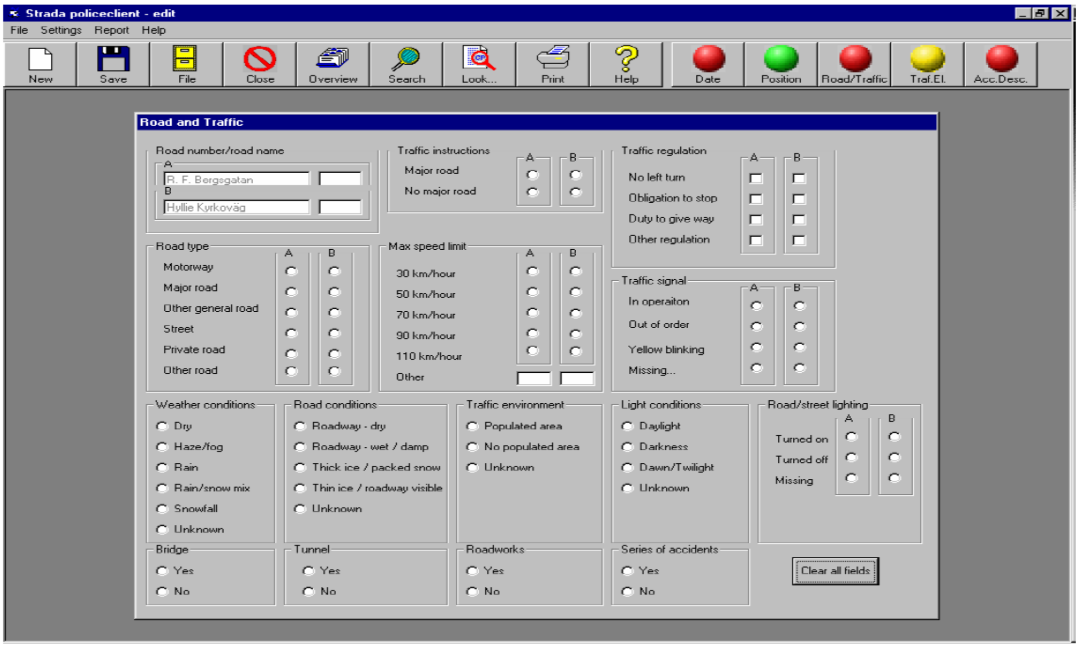


Figure 40: Road description in STRADA

The hospital data contain general data on the accident and the victims, a description of the injuries based on the ICD-10 codes (from which the injury severity index is calculated), and a digital map, where the accident is located, based on the GPS data of the ambulance that helped the victims (Figure 41 y Figure 42).

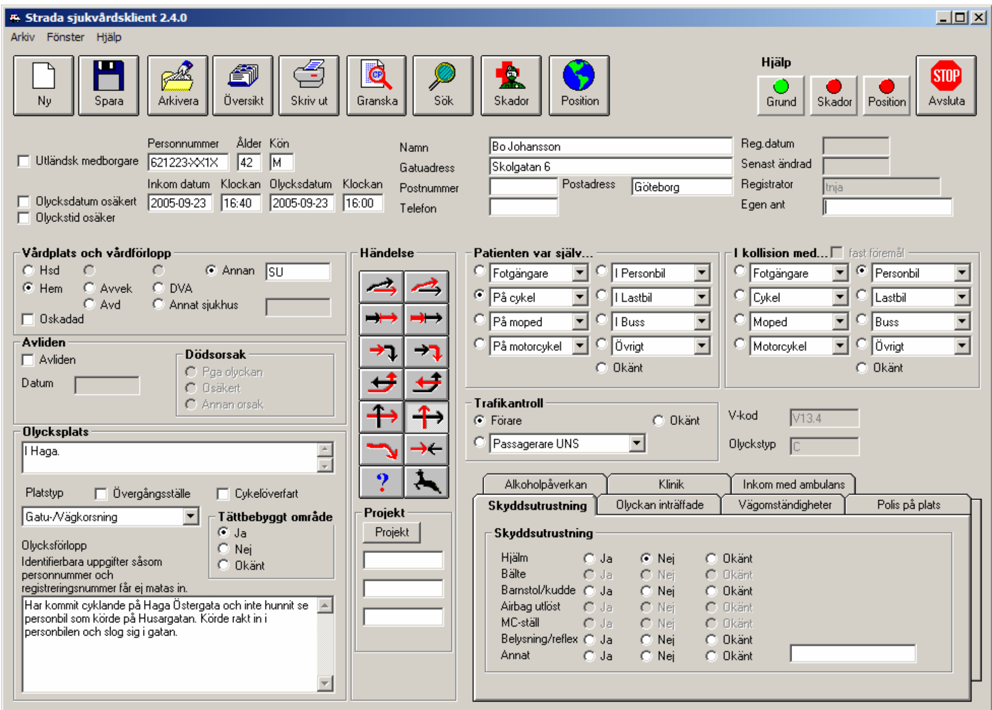


Figure 41: Hospital information about the injured and circumstances of the accident in STRADA

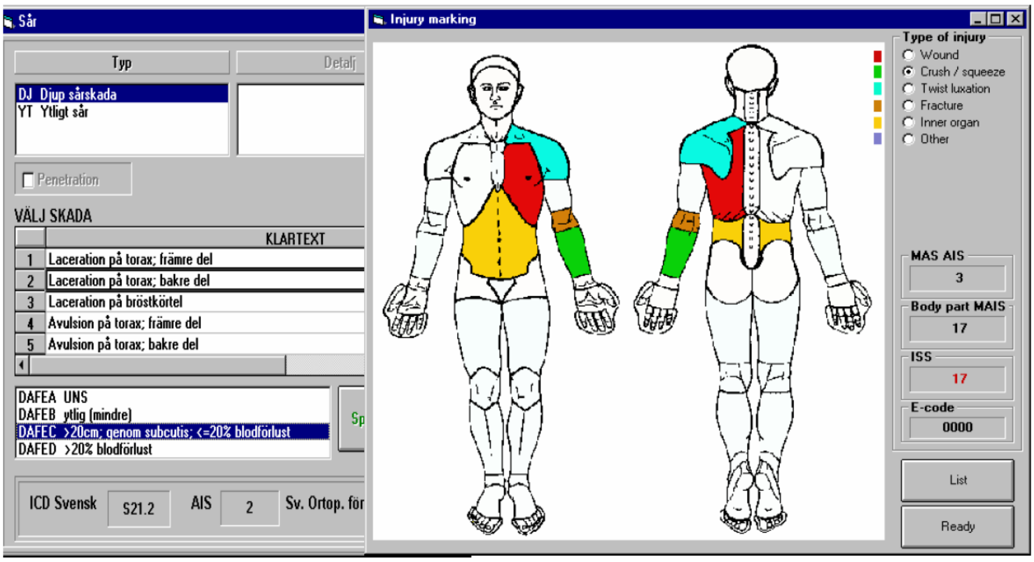


Figure 42: Hospital information about the injuries in STRADA

The system provides an information output in SQL language that makes easier the data handling and analysis. It also allows developing standardized result reports and data graphic displays (Figure 43).

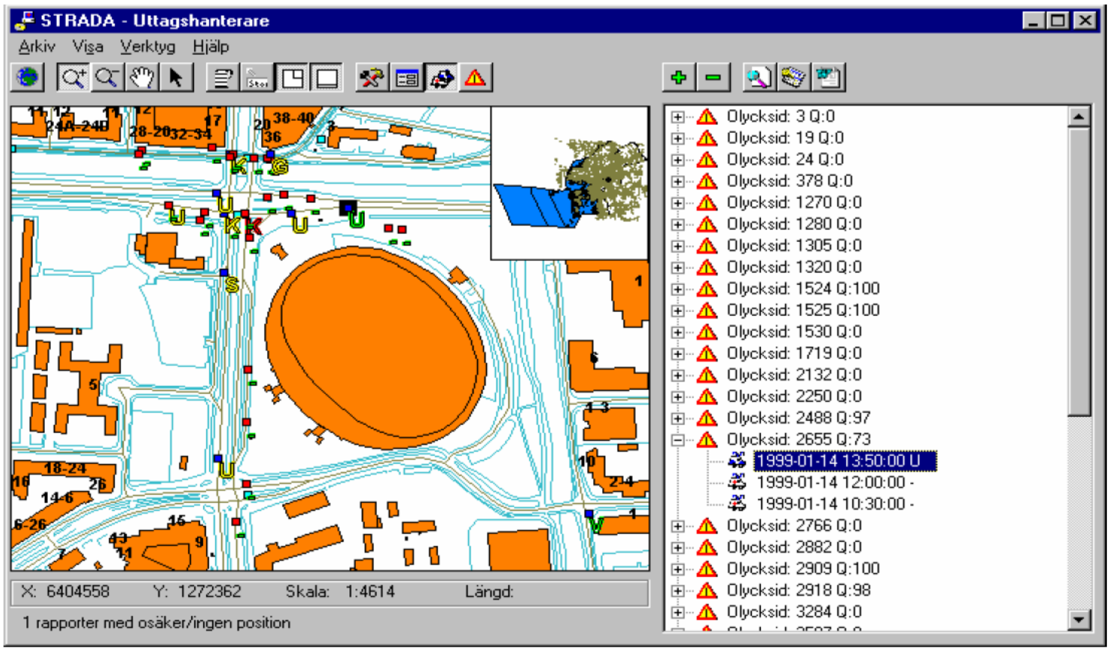


Figure 43: Accident graphic display in STRADA

5.1.5. BELGIUM: ISLP/FEEDIS-Pol Office

The computerisation of the services carried out by the Belgian police[11], is a project that started in the nineties with the PIP development (*Projet Informatique Police*), mainly developed for the municipal scope. The police reform, carried out in 2001, sped up this process of computerisation. The reform included the creation of a police structured in two levels: the federal police and the local police. This new structure favoured the development of two collection systems, the ISLP (Integrated system for the Local Police), at the local police's service and the FEEDIS system, used by the federal police (successor of the POLIS-brigade system of the ex-Gendarmerie and of the PV record of the ex-judicial police).

Both systems are linked and feed the general national database (Figure 44). From the data, the different bodies may use the "Datawarehouse" tool to make a statistical exploitation of the data.

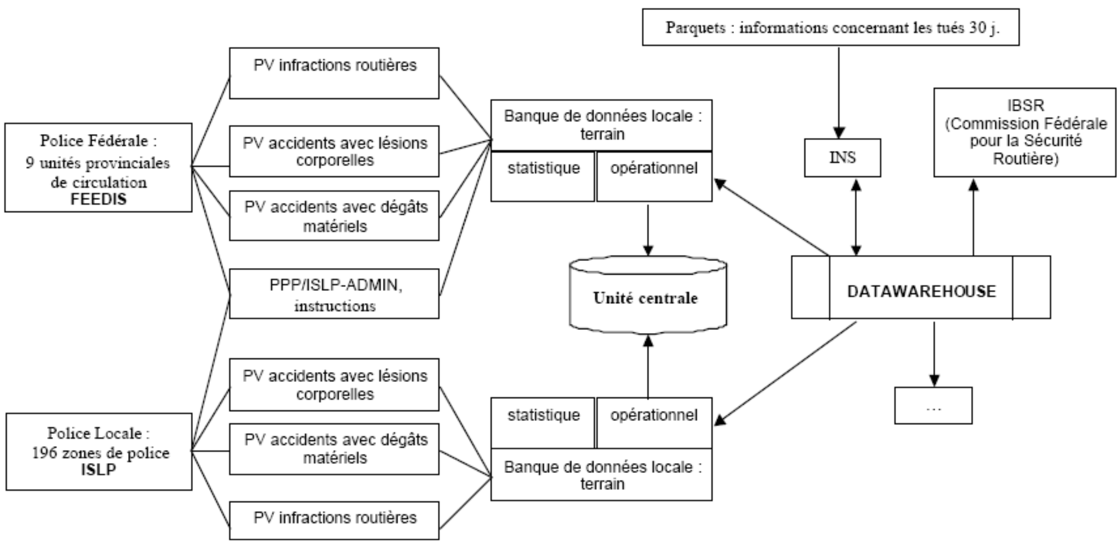


Figure 44: Schematic display of the data flow of the FEEDIS and ISLP systems in Belgium (Source: IBSR, PHL, LUC, 2004)

In the framework of the entry into force of the traffic law modifications in 2006, a new collection system called "Pol Office" is being developed. Nowadays the federal police is already using it and for 2008, it is hoped that it will be fully integrated in the Belgian local police.

[11] The Computer Sciences Corporation (CSC) took part in the whole process of computerisation and modernisation of the Belgian accident collection system.

The Pol Office system is a unique platform which objective is to integrate the ISLP local application and the FEEDIS federal application, as well as other internal services typical to the police (human resources, logistics, document-retrieval…).

5.1.6. POLAND: SEWIK

In Poland, the traffic accident database was created in 1975. After the latest modification of the format and the contents of the accident form (2004), the SEWIK system has been integrated to the Polish police database system (KSIP).

This database is the basic information source for the National Roads Directorate (GDDKiA), for the local authorities that have developed their own computer systems (Warsaw, Gdańsk, Bydgoszcz, Kraków) and for the research institutes and the Universities.

The new central database is being equipped with the following applications (Malasek, 2005):

- The list of all towns and villages with roads numbers and streets names for given country region.
- A tool to analyse road safety in the most important road routes.
- A tool to search for accidents according to their spatial location.
- A tool to list accidents according to determined individual variables (e.g. by year, accident type…) or combined variables (e.g. alcohol consumption per accident type).
- A tool to export data to the road administration databases and to import external data.
- An automatic identification of cases when the given threshold of accidents and collisions numbers in a certain place or within the time period was over passed.
- Specific levels of accessibility for different database users.

In the Figure 45 and the Figure 46, are presented several screenshots of the Polish collection system.

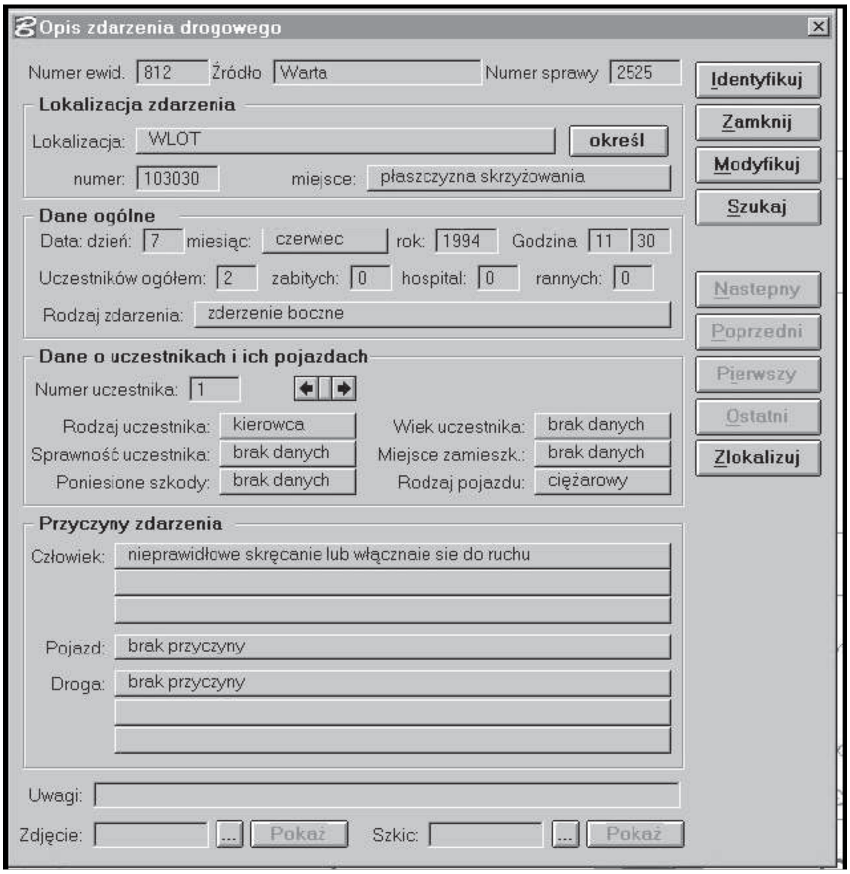


Figure 45: Accident collection screen in SEWIK

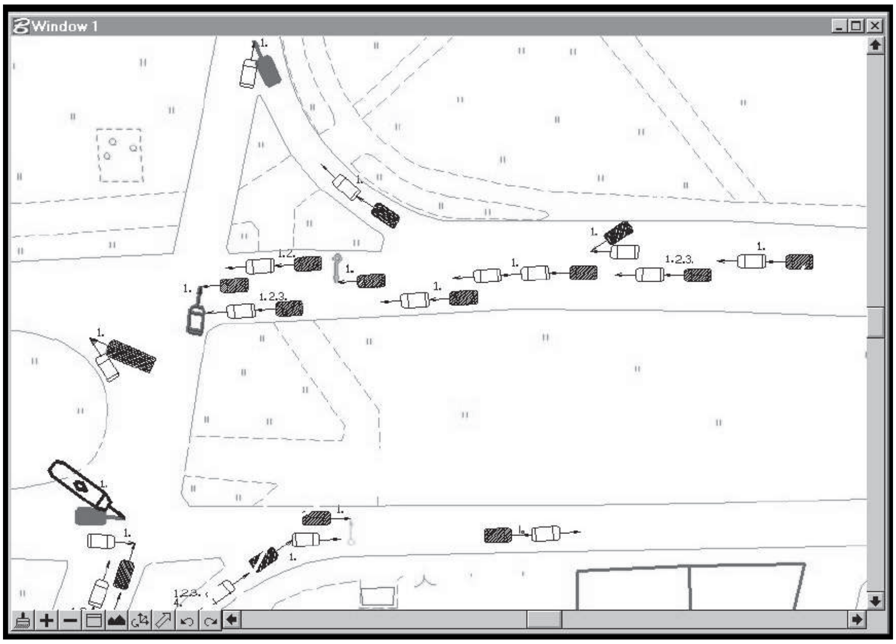


Figure 46: Screen of the accident sketch in SEWIK

5.1.7. SPAIN: CIAT y SIDAT (specific case in Catalonia)

In Spain, the current accident data collection and storage system integrates the information collected by the different police forces (Figure 47). From the centralised database, the reports that compose the yearbook and the statistical bulletins are produced, and the information is extracted for the applications developed for purposes of more specific exploitation.

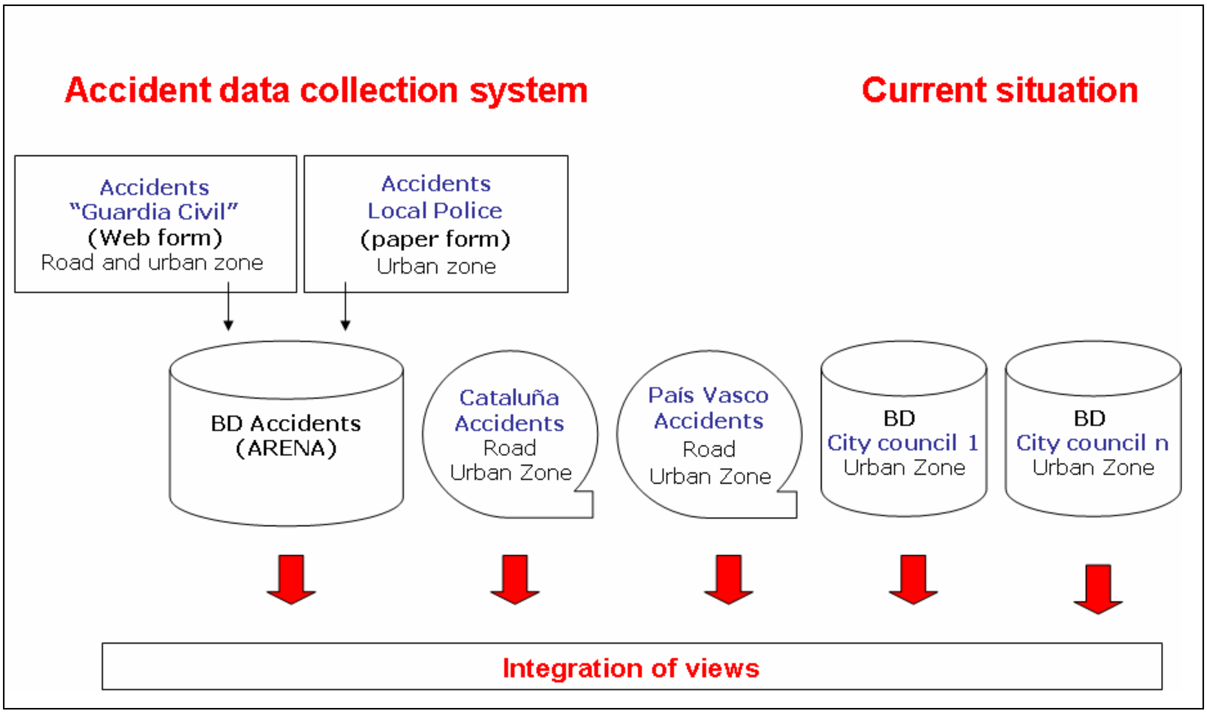


Figure 47: Current Spanish road accident data collection system

The main core of the data is obtained from the ARENA computer system, introduced in 2005 as a tool to collect the accidents attended by the traffic guardia civil or by the local polices (information centralised in the traffic provincial headquarters).

The system allows validating and contrasting the entered data during a processing prior to its integration in the central database, so that it will not allow incoherences, informing to the user the need to carry out modifications before storing the information. Moreover, it allows storing sketches and photos linked with the accident.

ARENA counts on a module that allows issuing basic reports, but it is not much extended given that the data exploitation is carried out through the creation of a data warehouse and its related analysis tools.

The remaining information that the system has to integrate comes from the autonomous polices of the Basque Country and Catalonia (they send the information by electronic device from their own independent systems), and from some municipalities having their own collection systems, like La Laguna (Canary Islands) and Madrid.

In the forthcoming months the launch of a new accident data information system is foreseen: the Traffic Accident Information Concentrator (CIAT).

The Figure 48 shows the architecture of the system. The main proposed idea is to unify the accident rate system at all the levels:

- Unification of the different ways of entering the information in order to dump the data in a unique concentrator.
- Unification of the different patterns of accident rate information.
- Unification of the different ways of terminal data exploitation. They can be on line, through a unique Web interface that enables the access to the exploitation tools that are intended to be implanted on CIAT (generators of lists or of small reports, data mining tools, corporative report tools, or GIS advanced tools), or they can be departure lines of the information to other remote systems, through exploitation.

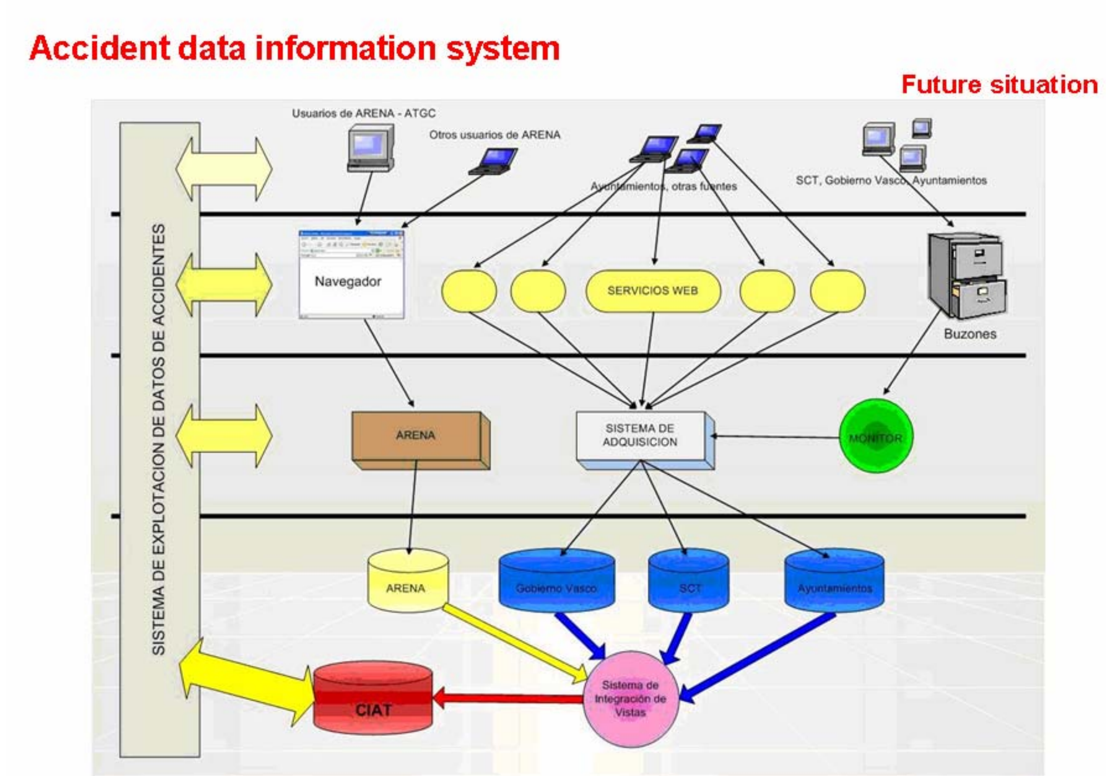


Figure 48: Architecture for the Traffic Accident Information Concentrator (CIAT)

<u>Autonomous Community of Catalonia (Servei Català de Trànsit): Traffic accident data collection integral system (SIDAT)</u>

In 2005, the Servei Català de Trànsit, organization responsible for traffic and road safety in the Autonomous Community of Catalonia launched a new traffic accident data collection integral system (Servei Català de Trànsit, 2006).

Its main objective is to achieve a common collection system for the different police forces in Catalonia that allows, as far as possible, a fundamental improvement of the quality of the accident data used for diagnosing and assessing road safety. This objective is reached form the following specific objectives:

- Creation of a common set of essential contents that all the police units must collect being homogeneous criterions.
- This data set must correspond with the needs and criterions that are posed by the European Community road accident database (CARE).
- Reduce and simplify the data collection task, increase the quality and avoid the numerous information duplications existing in the collection process.

- Development of a comprehensive, exhaustive and conceptually clear accident data collection manual.
- Computer definition of the selected fields of information: set up of filters to enter the information, rules to detect and reduce errors and inconsistencies, description and rules to enter missing data, and the set up of fields that are automatically filled in.
- Development of a system of data access licences from the users so that it enables its use and analysis locally. It is pretended to allow and promote the exchange of data between the central autonomous service (Servei Català de Trànsit) and the municipalities.
- Use the current information technologies (networks and client/server architectures) to be able to dispose of the information in a relatively updated way.

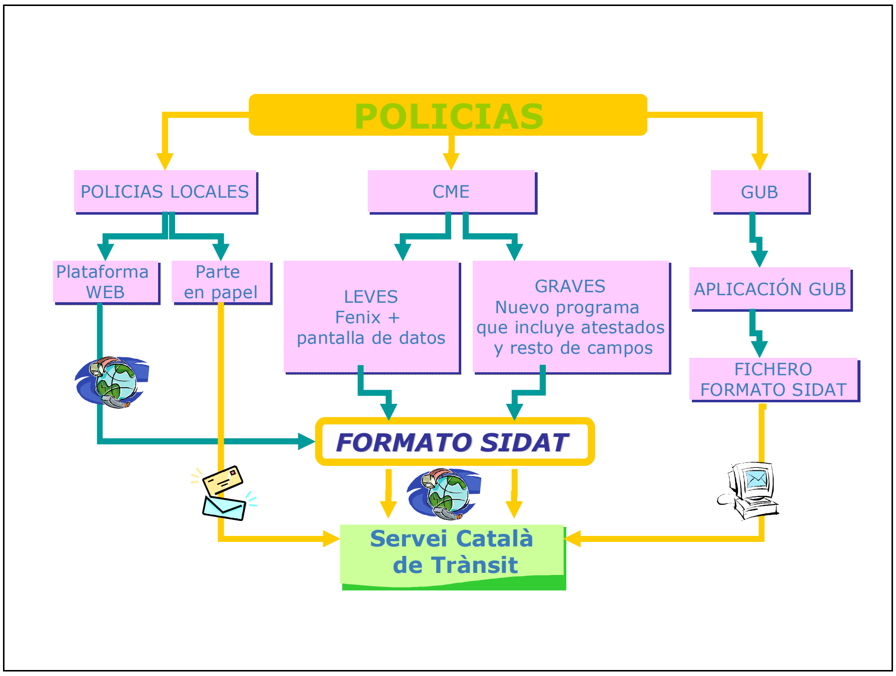


Figure 49: Information collection system of the different police forces in Catalonia. CME (mossos d'esquadra), GUB (urban guard of Barcelona)

The following figures show some screenshots of the SIDAT system.

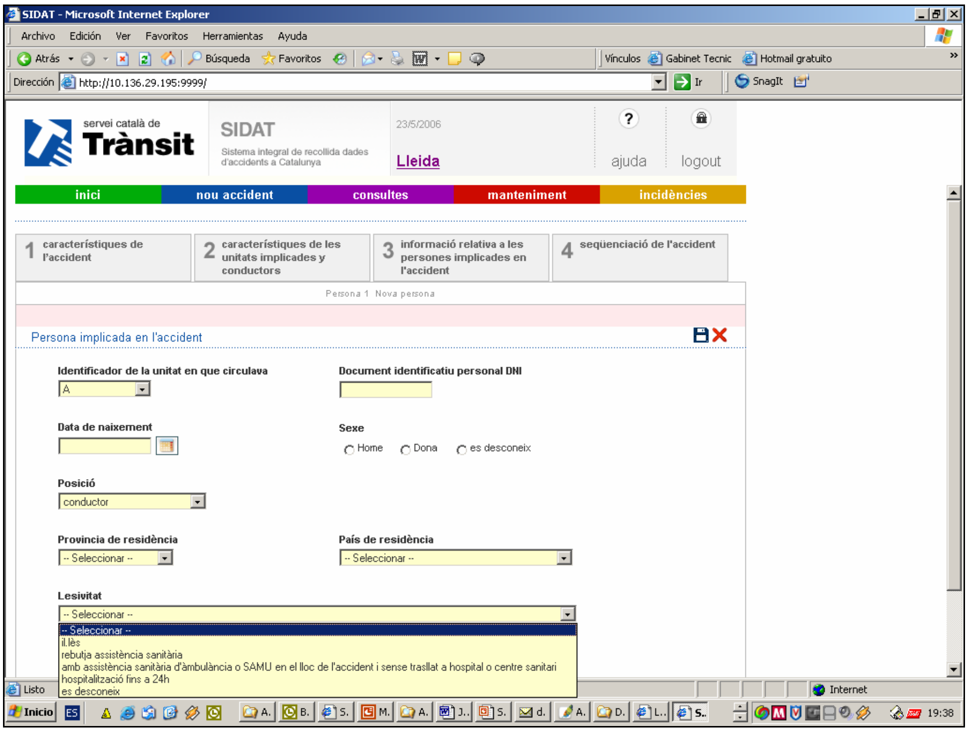


Figure 50: Screenshot of the SIDAT system. Data of the persons involved in the accident

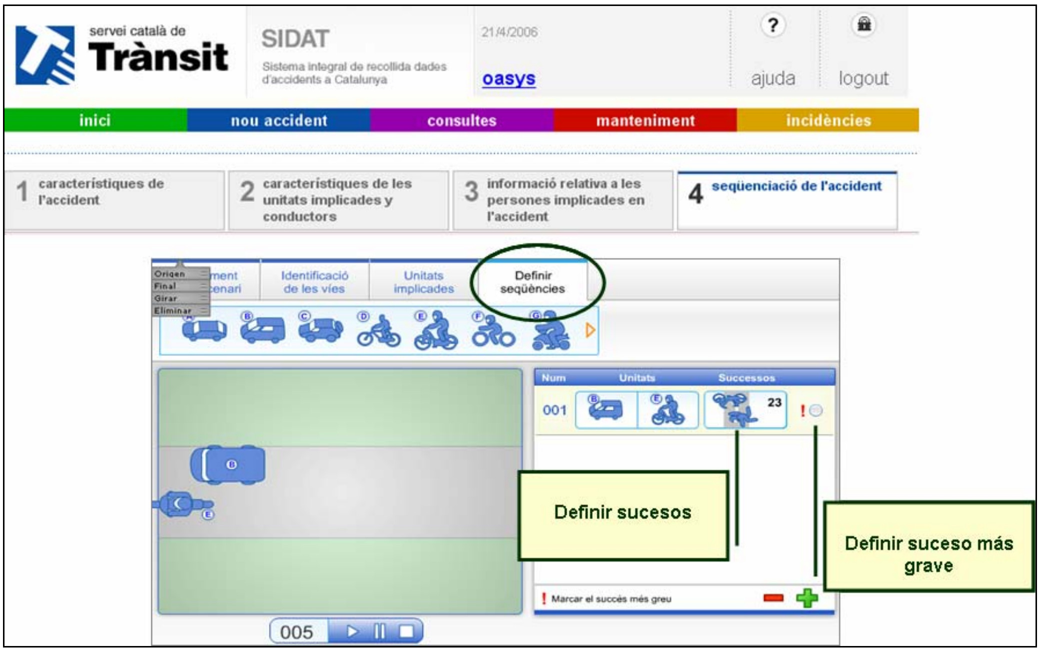


Figure 51: Screenshot of the SIDAT system. Data on the accident sequence

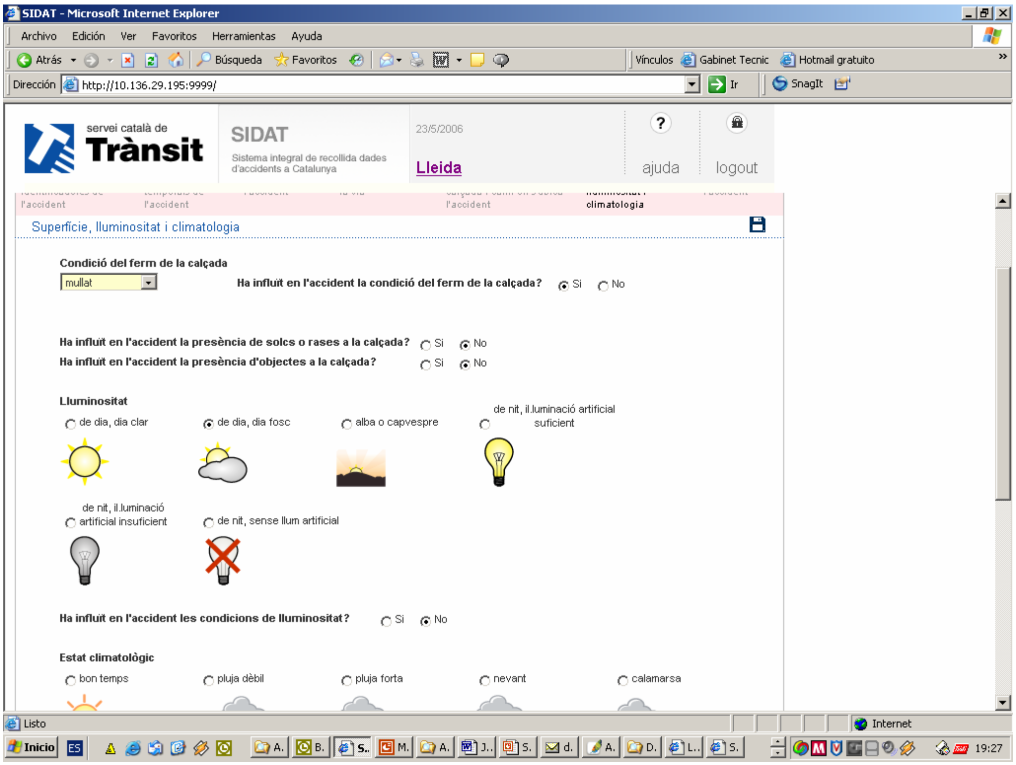


Figure 52: Screenshot of the SIDAT system. Data on the environmental conditions (road surface, light and weather conditions)

5.2. General reflections on the accident record collection systems

In this section, the objective is to extract the most important aspects of the reviewed systems, aiming to help specifying some relevant requirements that the accident collection systems have to satisfy.

According to Benavides and Serra (2003), it is possible to establish a series of parameters to evaluate the information systems quality (in this case, referred to the health systems, even though they can be adapted to any other scope of study) (Table 11).

Criterion	Definition
Simplicity	The structure (documents, circuits, deponents, etc.) and the procedures (classifications, indicators, etc.) must be easy to handle.
Flexibility	The capacity to adapt to new information needs.
Acceptability	The positive opinion of the users and of the persons that take part as deponents.
Predictability	The proportion of notified cases that are real cases.
Representativity	The description of the interest phenomenon characteristics (age, gender, etc.), throughout time, a for a defined population, is correct.
Punctuality	The rapidity or delay in getting the information.

Table 11: Definitions of the parameters proposed to evaluate the health information systems quality (adapted from Flauke, by Benavides and Serra, 2003)

INTRAS' experience in these procedures to improve traffic accident rate data systems shows us that its computerisation has to be raised intimately linked with the procedures of definition of the accident report contents, and with the working procedures for the collection of such information.

However, nowadays it is possible to reach solutions too sophisticated as for the accident rate records computerisation, being necessary to define the appropriate level according to the needs and means available in each case (a balance has to be reached between sophistication and efficiency of the system).

This way, independently of who develops the system, a series of common elements or points that the accident data management and analysis have to share may be defined:

1. **Relational structure**. Nowadays, the accident databases systems are constituted from a data relational structure, in which three data tables or entities are considered: 1) general data of the accident, 2) vehicle data and 3) persons' data.
2. **Use of Geographical Information Systems (GIS)**. The incorporation of these systems is very usual for the accident rate visualization and spatial analysis.

3. **Integration**. The general philosophy is to set up an integrated tool that considers both the data entry, and the queries and statistical analysis carrying out, as well as the data visualization and spatial analysis on a cartographic display.
4. **Importation-exportation**. These tools must have the possibility to import data (e.g. in case the entry is carried out in another database), as well as export data, either for the statistical analysis in a specialised software, or for its transmission to the different regional and national administrations (avoiding the need to entry the data twice).
5. **Easiness of use**. Normally the final users of the system are not experts in databases management and statistical analyses. This has to be taken into account while designing the interface and the procedures of use of the system. Nevertheless, the users' training is fundamental.
6. **Automation**. Related with the easiness of use, the systems usually have the possibility of queries and statistical analyses in an automated or semi-automated way, analyses defined according to their relevance in the accident rate local study.
7. **Flexibility**. However what has been previously said, the systems do not have to be strict, allowing the possibility of new queries and analyses according to the users' needs.
8. **Cohesion with the environment**. The traffic accident data management is articulated in a wider system in which many procedures and information take part. This would be the case of the documents generated by the police investigation and the judicial procedures of the accident, like could be the reports, proceedings, technical reports, etc. The implementation of a new system has to take into account all these procedures and integrate itself with these, in a way that it tries to reduce the work generated for each accident by avoiding duplication of tasks.
9. **Possibility of linkage with other databases**. There are data related with traffic that are systematically recorded and that are useful from the perspective of the accident rate analysis: Traffic (density, volume…) and infrastructure (characteristics, signposting…) data. With a statistical purpose only, the system could consider the automated linkage of these

external data with the ones recorded in the accident database. Likewise, the possibility of linkage with other data files (like the health assistance or also the offences and complaint ones) might be valued and studied. This would avoid the duplicity of efforts in the collection of the same information, it would simplify the number of variables to record, and would favour a maximum production of the system (greater information with the use of less means). In order to achieve this objective, it is very important to establish unique identifier variables that allow the linkage between the different databases.

10. **Maintenance and adaptability**. The system has to be little expensive (technically and economically) as for its maintenance. On the other hand and directly related, it has to allow and provide the necessary changes to be adapted to future changes in the accident questionnaire or database formats, as well as changes in other related elements, like GIS.

6. GENERAL CONCLUSIONS AND SOME INITIAL RECOMMENDATIONS

Nowadays, the traffic accident data collection, management and analysis systems represent a fundamental tool for the Road Safety management. There is a normative, at the national level, which determines such data collection practice, as well as the guide through more or less standardized procedures. However, such practice may present particularities and adaptations in its local application.

In general terms, the data collection aiming to the statistical analysis is integrated in a wider procedure of accident investigation, reconstruction and judicial-administrative management. This may lead that, for the police units in charge of filling in the accident questionnaire, this task often implies an additional work, summed to the accident investigation, and to the different technical and judicial documents that derive from such investigation (accident report, proceedings, technical report, etc.), generating a work overload.

On the other hand, huge deficiencies in the accident rate records have been detected particularly as for the underreporting and the information quality, especially in the case of the urban accident data.

Moreover, the results of the local data exploitation – when existing – are usually offered with an important delay and, sometimes, with such an aggregate level that it means that they loose a great part of their utility for the persons in charge of their collection at the local level.

Everything said before involves that the completion of the accident report for statistics ends being perceived as an added task with an administrative purpose, which has to be done because it is stated in the normative, but that has relatively few applied utility in the local field. Most probably this is a factor that, among others, is having an influence in the data quality problems to which we have been referring in this report.

So, apart from the need to increase and improve the available human and technical resources, it should be stressed on that, beyond an administrative procedure, the accident reports represent a research tool necessary to improve road safety.

In this sense, and as it has been pointed out throughout this report, the main international organisations with competence in traffic and road safety (European Commission, ETSC, OECD…) influence on the need to improve and optimize the accident data collection systems, in a way that their use is maximized at all the application levels of the road safety policies and actions: international, national, regional and local.

At the local level, this implies the need to carry out diagnoses, in the sense of "quality control", that value the performance of the current systems mainly in relation with two dimensions: (a) working procedure and quality of the resulting data and (b) use and exploitation of the data collected locally. From these diagnoses it is possible to start establishing what could be the general lines for the possible improvements.

This way, apart from the current procedures evaluations that are carried out for each case, a series of general guidelines or recommendations are posed and may set the future standards in the traffic accident data collection, management and analysis at the local level.

1. **Data local management and exploitation**. There is a need to favour the systems that provide the use of data at the local level, both for diagnostic and evaluation purposes. This passes by (1) the use of the systems and procedures introduced at the national level, avoiding duplications and transcriptions in the data entry tasks, as well as the delays for its availability, and (2) the development and implementation of the "expert" systems and tools that allow and provide to the local authorities the accident rate data management and analysis[12].
2. **Introduction of GIS in the accident rate analysis systems**. The incorporation of Geographical Information Systems (GIS), in the urban accident data management and analysis procedures, is going to be one of

[12] This type of systems allows obtaining a constant feedback of the work done, which implies data quality improvements and the proportion of collected accidents.

the most noticed trends in the next years. In this sense, and focussing on the analysis level, the most traditional approach to evaluate accident concentration stretches or points has to be complemented with the new developments coming from the accident data spatial analysis, in which the concept of "areas" of accident rate (in opposite to the point concept) with a more dynamic and wider scope acquires importance.

3. **Use of other data records related to traffic and transport**. An important part of the information that is needed in the accident reports may be obtained from other collection data sources or systems related to road safety (hospitals, vehicle record, infrastructure inventory, traffic data…). The linking data methods allow contrasting, while completing, the information of the police records. This way, for each accident or casualty we may have more complete and reliable records in relation to the accident, as well as to the injuries and consequences (hospital), vehicles (vehicle records), or road characteristics (road inventory/GIS).

4. **Use of accident and casualty data coming from insurance companies**. In some country like **Finland**, they are carrying out a systematic use of the accident rate data coming from the insurance companies to complete the data collected by the police. In other countries, the current situation is very different. In this sense, there is the need to start facilitating the necessary ways and requisites to establish working methods that allow using insurance companies' data in the road safety studies.

5. **Use of casualty data coming from the Health system**. The more reliable information on the casualty severity is the one provided by the health system. In **Spain**, for the moment there is not any normalized procedure through which the hospitals have to communicate the cases of traffic accident casualties. In **France**, for example, this procedure has been implemented. This way, the hospitals have to systematically notify the state of the victims that are still hospitalised six days after the traffic accident took place.

   On the other hand, some countries are establishing several systematic collection systems of general public health data and in which the traffic accident data has a relevant place. For example, in **The Netherlands**, some highly useful health records have been implemented in order to analyse accident rate (SWOV, 2000):

- LIS (Injury Information System) of the Institute of the Consumer Safety: centralized record of all type of accidents (or diseases) that need an emergency treatment.
- LMR (National Patient Register) of the Centre of Health Information: centralized file of hospital data.

6. **Application of new technologies in the data entry process and improvements in the working procedures**. With that, it is possible to increase the data quality at the same time of reducing the time and effort to collect it.

   An example of use of new technologies is the <u>completion of the accident report being assisted by expert or intelligent systems of help for the data entry</u>. It would be interactive systems, through which the agent in charge is answering questions posed by the computer according to the information he is entering. The system identifies the questions that are necessary to be answered according to the type of accident, place, etc. This system features several advantages:

   - Missing data are reduced (as it is a guided entry, it does not allow going to the next field until the previous one has not been filled in)
   - Errors and inconsistencies are reduced (data check)
   - The time needed to entry the data may be reduced. The guided entry only features for each step the needed items according to the answer in the previous items.
   - They may be used to collect data from several action fields of the agents, not only the accidents, and consequently make beneficial the implementation and maintenance costs.

7. In the case of a manual data entry, it is important to **set automatic filters and checking systems in the database**. The objective is to detect errors, incoherent and/or impossible data during the data entry process.

8. **Improvements related to training**: more training and better manuals. This is particularly important in the cases where there are no groups specialized in traffic. This would be the case of the greater part of the local polices. One of the key objectives in training has to be the homogenisation of the data completion criterions.
9. Carrying out **periodic systematic reviews of the accident report and of the collected data** aiming to delimiter problematic fields and types of information. From these reviews, it is possible to move on to possible modifications that, in case they are implemented, it will have to be taking into account the comparability between the data before and after the change. On the other hand, the persons in charge of collecting data have to actively take part in the decisions to be taken about such modifications.
10. **Incentive schemes for the agents in order to improve the quality of the data** on the collected accidents. This has to come with an appropriate feedback on the task they carry out usually.
11. The municipal authorities have to establish agile channels of **collaboration and exchange of information in relation with traffic and road safety**. This favours the exchange of field experiences, with the resulting use of the successful experiences and avoidance of errors.
12. Together with the central administration, develop **proposals to establish new standardized procedures that allow optimizing the current local practices**, as well as maximizing the use of the great quantity of accident data that are now produced by the local administrations.
13. Finally, it would be convenient to establish cooperation experiences and information exchange between several EU cities and, on the other hand, lay out appropriate ways to **increase the participation and introduction of proposals of the municipalities in the different European Commissions** in which the guidelines that the future traffic and road safety information management and analysis systems will follow are being established.

Bibliographical References

AGRAN, P.F. & DUNKLE, D.E. (1985) A comparison of reported and unreported noncrash events. *Accident Analysis and Prevention*, 17(1), 7-13.

AGRAN, P.F., CASTILLO, D.N., & WINN, D.G. (1990) Limitations of data compiled from police reports on paediatric pedestrian and bicycle motor vehicle events. *Accident Analysis and Prevention,* 22(4), 361-370.

AJO, P. (1996) Investigación en Seguridad Vial a través del análisis de los siniestros. Experiencia Finlandesa. Trabajo presentado en XII Jornadas Comunitarias del Seguro del Automóvil. Madrid, 1996.

AUSTIN, K. (1995a). The identification of mistakes in road accident records: part 1, locational variables. *Accident Analysis and Prevention* 27(2), pp. 261-276.

AUSTIN, K. (1995b). The identification of mistakes in road accident records: part 2, casualty variables. *Accident Analysis and Prevention* 27(2), pp. 277-282.

AYUNTAMIENTO DE OVIEDO (2007) Plan Municipal de Seguridad Vial (2007-2010). http://www.dgt.es/dgt_informa/observatorio_seguridad_vial/seguridad_vial_local/Plan_Municipal_de_Seguridad_Vial_de_Oviedo.pdf

AYUNTAMIENTO DE SAN SEBASTIAN (2006) Plan de Seguridad Vial Donostia-San Sebastián (2006-2009): Objetivos y líneas de actuación. Dirección de Movilidad y Jefatura de la Guardia Municipal. http://www.dgt.es/dgt_informa/observatorio_seguridad_vial/seguridad_vial_local/ss_objetivos.pdf

BARANCIK, J.I., FIFE, D. (1985) Discrepancies in vehicular crash injury reporting: Northeastern Ohio Trauma Study. IV. *Accident Analysisand Prevention*, 17(2), 147-154.

BENAVIDES, F.G. & SERRA, C. (2003) Evaluación de la calidad del sistema de información sobre lesiones por accidentes de trabajo en España. *Arch Prev Riesgos Labo*r, 6(1), 26-30. http://www.upf.edu/cexs/recerca/ursl/fitxers/ornatart/fgb3aprl.pdf

CARE (2006) CARE-Glossary. European Commission/Directorate General Energy and Transport. http://ec.europa.eu/transport/roadsafety_library/care/doc/care_glossary.pdf

CERTU (2004) Les logiciels d'accidentologie. http://www.certu.fr/ROOT/customer/documents/pdf00102347.pdf

CFSR (2007) Etats Generaux de la securite routiere 2007. Dossier: Statistiques. Groupe de Travail Statistiques. Commission Fédérale pour la Sécurité Routière http://www.cfsr.be/Docs/Groups/CFSR%20GT%20Statistiques%20DEF.pdf

CHAPELON, J. & LOONES, F. (2006) Le fichier nacional des accidents corporels de la circulation routière en France. Observatoire national interministériel de sécurité routière. http://www.securiteroutiere.equipement.gouv.fr/IMG/pdf/Fichier_national_des_accidents.pdf

CHISVERT, M. (2000). *Calidad y representatividad de los datos de accidentes de tráfico: Revisión, estudio del caso español y desarrollo de propuestas para la mejora de los sistemas de recogida y tratamiento de la información sobre accidentalidad.* Tesis Doctoral. Universitat de València., Estudi General. Valencia.

CHISVERT, M., MONTEAGUDO, M.J. & PASTOR, G. (1998). *Aspectos y problemas metodológicos en la investigación estadística de los accidentes de tráfico: recogida y tratamiento de la información y análisis estadístico. II Congreso Iberoamericano de Psicología: Congreso Sectorial de Psicología del Tráfico y Seguridad Vial*, Madrid, 13-17 de Julio de 1998.

COM(1997) 131 final (1997). Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions. *Promoting road safety in the EU*. The programme for 1997-2001.

COM(1997)238 final (1997). Report from the Commission on progress with the project and its future prospects CARE. *Community Database on Road Traffic Accidents*. Council Decision of 30 November 1993 (93/704/EC).

COM(2000) 125 final. (2000) Communication from the Comission to the Council, the European committe of the regions - *Priorities in EU - road safety progress report and ranking of actions*.

COM(2003) 311 final (2003) Salvar 20 000 vidas en nuestras carreteras. Una responsabilidad compartida. Programa de acción europeo de seguridad vial

COSTA, J. & ARNAU, J. (1989). Els sistemes d'informació en els accidents de trànsit. Bancs de dades a Catalunya. En: Gerència de Seguretat Vial (Eds.) *Els Accidents de Trànsit a Catalunya*. Generalitat de Catalunya. Departament de Governació.

CROW, M., HALLADAY, M.L., MARTINOVICH, S., HILGER, B., HARKEY,D.L., MCNAMARA, D.J., ELLISON, J.W., LACY, J.K., SERIAN, B.L., GRIFFITH, M.S. & MACGREGOR, S. (2004) *Traffic Safety Information Systems in Europe and Australia.* Federal Highway Administration. U.S. Department of Transportation. Technical report FHWA-PL-04-010

DE FRANCISCO, E. (1996). *Accidentes de tráfico con motocicleta: Repercusiones Sociales, Económicas y Médico-Legales*. Tesis doctoral no publicada. Universitat de València.

DE MOL, J., BOETS, S. (2003) Optimisation of Traffic Accident Statistics (intermediary report). Interim Report. Belgian Public Planning Service Science Policy. Brussels. http://www.belspo.be/belspo/home/publ/pub_ostc/CP/CP39_en.pdf

DENCO, Universia di Brescia, INRETS, TRL & Aristotle University of Thessaloniki (2000). *The DUMAS Project. Developing Urban Management and Safety. WP-4. Accident Investigation*.

DERRIKS, H.M. & MAK, P.M. (2007) Underreporting of road traffic casualties. IRTAD special report

DFT (2005) Instructions for the Completion of Road Accident Reports (Stats20). http://www.dft.gov.uk/pgr/statistics/datatablespublications/accidents/casualtiesgbar/stats20instructionsforthecom5094

DFT (2006a) Review of Road Accident Statistics. National Statistics Quality Review Series. Report No.45. http://www.statistics.gov.uk/about/data/methodology/quality/reviews/downloads/NSQR45.pdf

DFT (2006b) Under-reporting of Road Casualties-Phase 1. Road Safety Research Report nº.69. Department for Transport. London.

http://www.dft.gov.uk/pgr/roadsafety/research/rsrr/theme5/underreportingofroadcasualti4788

DGT (2005) Plan Estratégico de Seguridad Vial: Plan de Acciones Estratégicas Claves 2005-2008. Observatorio Nacional de Seguridad Vial. DGT. Ministerio del Interior. http://www.dgt.es/dgt_informa/investigaciones/Plan_Acciones_Estrategicas_Claves_2005_2008.pdf

DGT (2007) Plan Tipo de Seguridad Vial Urbana. Guía de apoyo para la actuación local. Observatorio Nacional de Seguridad Vial. DGT. Ministerio del Interior. http://www.dgt.es/dgt_informa/investigaciones/Plan_Tipo_de_Seguridad_Vial_Urban_2007.pdf

DUNN, H.L. (1946) Record linkage. Technical Report. *American Journal of Public Health*, 36, 1412-1416.

EUROPEAN TRANSPORT SAFETY COUNCIL (1997). *A strategic road safety plan for the European Union*. European Transport Safety Council. Bruxelles.

EUROPEAN TRANSPORT SAFETY COUNCIL (1999). *Exposure data for travel risk assessment: Current practice and future needs in the UE*. European Transport Safety Council. Bruxelles

EUROPEAN TRANSPORT SAFETY COUNCIL (2000a). *EU transport casualty databases: Current status and future needs*. European Transport Safety Council. Bruxelles.

EUROPEAN TRANSPORT SAFETY COUNCIL (2000b). *Transport accident investigation in the European Union*. European Transport Safety Council. Bruselas.

EUROPEAN TRANSPORT SAFETY COUNCIL. (2001). *EU Transport accident, incident and casualty databases : current status ansd future needs*. Bruxelles.

FERRANTE, A.M., ROSMAN, D.L. & KUIMAN, M.W. (1993). The construction of a road injury database. *Accident Analysis and Prevention* 25(6), pp. 659-665.

FIFE, D: & CARDIGAN, R (1989) Regional Variation in Motor Vehicle Accident Reporting: Findings from Massachussetts. *Accident Analysis and Prevention*, 2, 193-196.

FRANTZESKAKIS, J., YANNIS, G. & HANDANOS, J. (2000) *The potential of accident analysis systems for the evaluation of road safety measures in Europe*. Proceedings of the XIII ICTCT Workshop "Evaluation, validation, implementation of measures to improve transport safety", Corfu, pp. 45-57. http://www.ictct.org/workshops/00-Corfu/Yannis.pdf

FONTAINE, H., GOURLET, Y., L'HOSTE, J. & MUHLRAD, N. (2003) *Inventaire critique des données nécessaires a la recherche en sécurité routière*. Rapport final sur convention 02/70/013 DSCR / INRETS http://www.securiteroutiere.equipement.gouv.fr/IMG/pdf/RR_Rapportfinal_inventaire.pdf

FREIXA, M., GUÀRDIA, J., HONRUBIA, M.L., & PERÓ, M. Estimación de la prevalencia a partir de los métodos de captura-recaptura. Psicothema, vol.2, supl.2, 231-235

GOBIERNO DE NAVARRA (2006) Estrategia Navarra de Seguridad Vial (2005-2012). http://www.dgt.es/dgt_informa/observatorio_seguridad_vial/seguridad_vial_ccaa/nv_estrategia_sv.pdf

GOBIERNO VASCO (2003) Plan Estratégico de Seguridad Vial de la CAPV (2003-2006). http://www.trafikoa.net/dinamikoak/planak/1_es_archivo.pdf

GOVERNMENT STATISTICAL SERVICE (2005) STATS21 (Scottish Edition): Procedures for Submitting Road Accident Data to the Scottish Executive. http://www.scotland.gov.uk/Resource/Doc/933/0031792.doc

HAGEN, K. E. (1993) "*Samfunnsokonomik regnskapssystem for trafikkulykker og trafikksikkerhetstiltak*" (Socio-economic accounting system for traffic accidents and countermeasures). TOI-Report 182/1993. Oslo, Norway. (Summary in English)

HAKKERT, S. & HAUER, E. (1988). The extent and implications of incomplete and inaccurate road accident reporting. En ROTHENGATTER y bruin (eds.). *Road User Behaviour: Theory and Research*. Van Gorcum. Maastricht, The Netherlands.

HARRIS, S. (1990) The real number of road traffic accident casualties in the Netherlands: A year-long survey. *Accident analysis and prevention*, 22(4), 371-378.

HILL, T. & PRING-MILL, F. (1981) *Generalised Iterative Record Linkage System: GIRLS (Glossary, Concepts, Strategy guide, User guide).* Technical Report. Statistics Canada. Ottawa.

HIRSTO, J. (1995). *ECMT accidents statistics*. En IRTAD, International Road Traffic an Accidents Databases. Seminar Proceedings. OECD, Paris

HUGHES, W.E, REINFURT, D., YOHANAN, D., ROUCHON, M, & McGEE, H.W. (1993). *New and emerging technologies for improved accident data collection*. Departament of Transportation, Wahington, DC.

IBSR, PHL & LUC (2004) Exploitation de données en matière de sécurité routière. *Projet AGORA*: *Mission de soutien scientifique aux banques de données administratives* (AG/EE/054). http://www.belspo.be/belspo/home/publ/pub_ostc/agora/rageer054_fr.pdf

INTRAS (2005) *Estudio de los datos de incidencias y vehículos implicados*. Proyecto Segura III: Factores de Seguridad en Autopista. Informe técnico.

IRTAD (1994). *Underreporting of road traffic accidents recorded by the police, at the international level*. Special report, OECD, Paris

IRTAD (1996). *Follow-up of traffic victims during the 30 days period after the accident*. Special report, OECD, Paris

IRTAD (1997). *Methods and necessity of exposure data in relation to accident and injury ststistics*. Special report, OECD, Paris

IWG (2003) *Glosaire des statistiques de transport*. Troisième edition. Conseil Economique et social. Nations Unies. TRANS/WP.6/2003/6. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-BI-03-002/FR/KS-BI-03-002-FR.PDF

JARO, M. (1995) Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Trampa, Florida. *American Statistical Association Journal*, 84(406), 414-420.

KAMPEN, L.T.B, PEREZ, C. & MARTIN J.L (2005) Análisis Protocol. Deliverable D7 WP3. PENDANT Project. SWOV

KOORNSTRA, M. J. (1995). Current statistical tools, systems and bodies concerned with safety and accident statistics. En OECD (Ed.) Seminar Proceedings. International Road Traffic and Accident Databases. 11-13 septiembre. Helsinki, Finlandia.

LAPLACE, P. S. (1783) Sur les naissances, les mariages, et les morts. In Histaire de I'Academie Royale des Sciences. Paris, Impr. Royale. p. 693.

LEDESMA, R., SANMARTÍN, J. & CHISVERT, M. (2000) Sesgos en la recogida de datos sobre accidentes de tráfico: evaluación de la actuación del personal técnico en autopistas de peaje. Psicothema, 12, Supl. 2, 320-326.

LITTLE, R.J.A., & RUBIN, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.

MALASEK, J., KORNALEWSKI, L. & KRZYSZTOFOWICZ, B. (2005) Road Accident Information Storage System and Cooperation between Police and Roads Administration. Presentation to the First FERSI Scientific Road Safety Research Conference. Bergisch Gladbach. Germany. http://www.fersi.org/downloads/September_7_8_2005/WS1/FERSI%20Conf.%20-%20Jacek%20Malasek.doc

NAUS, J.I. (1982) Editing Statistical Data. En Kotz & Johnson (eds.) *Encyclopaedia of Statistical Sciences*. New York: Wiley & Sons.

NEWCOMBE, H.B. (1988) *Handbook of Record Linkage*. Technical Report. Methods for health and statistical studies, administration and business. Oxford University Press. Oxford.

MORRISON, A. & STONE, D.H. (2000) Capture-recapture:a useful methodological tool for counting traffic related injuries? *Injury Prevention*, 6, 299-304.

OBSERVATOIRE NATIONAL INTERMINISTÉRIEL DE SÉCURITÉ ROUTIÈRE (2007) Le programme d'action de l'observatoire. http://www.securiteroutiere.equipement.gouv.fr/IMG/rtf/prog_action-2.rtf

OSTROM, M., HUELKE, D.F., WALLER, P.F., ERIKSSON, A. & BLOW, F. (2002) Some biases in the alcohol investigative process in traffic fatalities. *Accident analysis and prevention*, 24(5), 539-545.

PARLAMENTO EUROPEO (2001) *Textos aprobados en la sesión del jueves 18 de enero de 2001*.

PLASÈNCIA, A. (1995). *Epidemiologia de les lesions per accident de trànsit a Barcelona, 1990-91*. Tesis doctoral no publicada. Universitat Autònoma de Barcelona. Barcelona, 1995.

RAMS, M.A., ORTEGA, A. & SANCHO, M. (1993). Influencia del alcohol y las drogas en los accidentes de tráfico. *Revista Española de Drogodependencias*, 247-252.

RAZZAK, J.A. & LUBY, S.P.(1998) Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture-recapture method. *International Journal of Epidemiology*, 27, 866-870.

ROOS, LL. & WADJA, A. (1991) Record Linkage Strategies. Technical Report. Methodological Information in Medicine, 30, 117-123.

ROSMAN, D.& KNUIMAN, M.W. (1994) A comparison of hospital and police road injury data. *Accident Analysis and Prevention*, 26, 2, 215-222.

SABEY, B. (1990). Accident analysis methodology. *Journal of International Association of Traffic and Safety Sciences*, 14, (1), 35-42.

SAFETYNET (2006) *Database Transparency – Deliverable D4.2.* http://www.erso.eu/safetynet/fixed/WP4/sn_inrets_D4%202_final_03_02_06.pdf

SERVEI CATALÀ DE TRANSIT (2005) Plan de Seguridad Vial (2005-2007). Generalitat de Catalunya. http://www.dgt.es/dgt_informa/observatorio_seguridad_vial/seguridad_vial_ccaa/ct_plan_catalan_sv_2005_2007.pdf

SERVEI CATALÀ DE TRANSIT (2006) Sistema integral de recollida de dades d'accidents de trànsit SIDAT. Quaderns de trànsit 3. Generalitat de Catalunya.

SCOTTISH GOVERNMENT STATISTICAL SERVICE (2005) Procedures for Submitting Road Accident Data to the Scottish Executive. STATS21. Scottish Edition

STATISTICS FINLAND (2006) Tieliikenneonnettomuudet 2005: Vägtrafikolyckor (Road Traffic Accidents). Official Statistics of Finland. Helsinki. http://www.liikenneturva.fi/fi/tilastot/liitetiedostot/Tieliikenneonnettomuudet_2005.pdf

SWOV (2000). *Data Sources in Road Safety Research In the Netherlands*. http://www.swov.nl/en/kennisbank/gegevens/index.html

TERCERO, F. & ANDERSSON, R. (2004) Measuring transport injuries in a developing country: an application of the capture-recapture method. *Accident analysis and prevention*, 36, 13-20.

TRL, INRETS , SWOV , Danish Road directorate, DENCO, Universia di Brescia, BAST, Kuratorium fur Verkehrssicherheit, Transpor Research Centre & Aristotle University of Thessaloniki (2000). *The DUMAS Project: Developing Urban Management and Safety. Research Report.*

TRL, INRETS, SWOV, Danish Road directorate, DENCO, Universia di Brescia, BASt, Kuratorium fur Verkehrssicherheit, Transpor Research Centre & Aristotle University of Thessaloniki (2000). *The DUMAS Projec: Developing Urban Management and Safety.WP-1. Existing experience.*

VALERO, P. (1999) *Proyecto docente: Proceso de datos en Psicología*. Facultad de Psicología. Universidad de Valencia

VALERO, P. & YOUNG, F. (2000) *ViSta: The visual statistics system. Missing data analysis*. Technical report. LC. Thurstone Psychometric Laboratory. North Carolina University – INTRAS (Univ. de Valencia). http://www.uv.es/valerop

VOAS, R.B. (1993) Issues in cross-national comparisons of crash data. *Addiction* 88 (7), 959–967.

WALLER, J.A. (1972) Nonhigway injury fatalities. The roles of alcohol and problem drinking, drugs and medical impairment. Journal of Chronical Disease, 25, 33-45.

WHO Ad Hoc Technical Group (1979). Road Traffic Accident Statistics, EURO Reports and Studies 19, Regional Office for Europe, World Health Org., Copenhagen.

XUNTA DE GALICIA (2006) Plan de Seguridade Viaria de Galicia (2006-2010). http://www.dgt.es/dgt_informa/observatorio_seguridad_vial/Plan_de_Seguridad_Galicia2006-2010.pdf

YANNIS, G. & EVGENIKOS, P. (2006) *Work Package 1: CARE Accident Data*. Paper presented to Safetynet Conference. Prague.

More information on the SAU project at:

http://www.uv.es/sau/