Abstract

This paper presents a comparative analysis of two large scale databases with disaggregate data, the CARE system of the European Union and the FARS system of the United States. The FARS system is already operational since 1975, whereas the CARE system is operational in pilot phase since 1995. The comparative analysis of technical and operational characteristics of information input, output and processing methods used, identifies the possibilities for inter-state road accident analysis in the two continents. This analysis reveals not only the potential of the two systems but also the way road safety is treated in the two continents.

Key-words: road safety, road accident analysis, data bases
1. Introduction

State-of-the-art

A considerable work effort has been dedicated over the last decades in road accident analysis in the United States of America (USA) and in the European Union (EU). State level road accident analysis has led to numerous interesting results better explaining the road accident phenomenon and which were used for the development and implementation of road safety policies at state and local level. Efforts for road accident analysis at inter-state level are lacking behind equivalent efforts at state level due not only to the particularities of the inter-state variance of the road traffic and accident characteristics, but also to the difficulties arising from the incompatibilities among the various state data and the related insufficiencies of the central-level data bases.

This road accident analysis is supported by huge data collection and processing systems concerning road accidents at state level. Since the late fifties, computerised data bases have been developed in the various USA and EU states containing disaggregate road accident data. These data files evolve constantly in order to meet the constantly changing users' needs and technological possibilities. An important work effort has been dedicated to the improvement of these systems which in most of the cases were developed independently of each other.

Today, in the USA and in the EU, two data bases with disaggregate data on road accidents at central level exist, allowing for detailed road accident analysis on a inter-state basis. The USA FARS system was developed in the early seventies and is already operational for more than 20 years, whereas the EU CARE system was developed in the late eighties and is currently in the final phase of its pilot operation. Both systems present a number of similarities and differences arising mainly from the way they deal with coordination of the data collection and processing systems of the various states.

The objective of this research is to carry out a comparative analysis of the current and future potential of FARS and CARE systems allowing for the extraction of useful information on their utility and their further development. The results of this research can be especially useful for the CARE database which can benefit a lot from experience gained during the twenty-year operation of the FARS system.
Methodology

This work is extending part of a research on the potential of a number of international data files on road accident statistics and concerns only the two multi-state data files with disaggregate data on road accidents, FARS and CARE. In the framework of this research, preliminary information has been collected through extensive bibliography research allowing for the formation of a questionnaire for the support of the information collection during interviews with the persons responsible for the files. During these interviews very interesting information was collected for the technical and operational characteristics of each system as well as for the way towards their further improvement.

All information collected was put together and analysed in a uniform way allowing extraction of comparative results. Quantitative and qualitative comparative Tables were prepared allowing the formulation of general and specific conclusions. During this procedure, additional information was provided by the persons responsible for the two systems.

For text ergonomy reasons in the context of this paper, member states of USA and the EU are called states, even though they belong to two different types of State formations, the United States of America (USA) which is a federal state and the European Union (EU) which is a union of independent states. Special attention has been given to the description of the two types of State formations because behind their differences rely also some of the differences of the two systems.

2. The CARE system

General characteristics

The European Union (EU) is the actual form of an inter-governmental organisation, the European Community which was founded in 1957 and achieved in 1993 the single market in which goods, services, people and capital move about as freely as within one country and is in the process of preparing the way for further integration (economic and monetary union, common foreign and security policy, etc.)\textsuperscript{4}. The European Commission is the key institution of the EU whose role and responsibilities place it firmly at the heart of the European Union's policy-making process. The Directorate General for Transport of the European Commission is in charge for the execution of the EU transport policy in which the road safety chapter gained a growing importance over the last years.

Recent developments in the European society (integration of the European Community internal market, opening up of the eastern European markets) gave also a more international character to road traffic and safety problems and initiatives for international cooperation in the field of road traffic and safety are intensified\textsuperscript{5}. Approaches for dealing with the road accident phenomenon at European level are gaining more ground and the European.
Union has already set the basis for a European road safety policy\(^6\). In order to better support road safety policy at EU level, the European Commission has developed over the last years three road accident data files: a) the CARE project (disaggregate data), b) the Eurostat statistics (aggregate data since the seventies) and c) the Road Safety Quick Indicator (recent trends).

The CARE data base is an initiative of the EU member states to create a data base with disaggregate road accident data\(^7\). This initiative started in 1988 and is coordinated by the European Commission competent services, i.e. the General Directorate for Transport as far as road safety issues are concerned, Eurostat for statistics' issues and the Informatics Directorate for informatics issues. The first phase of the project (1988 - 1993) consisted of a feasibility study for the creation of CARE data base which led to a positive result and thus, to the European Council decision of December 1993 for the creation of a disaggregate road accident data base\(^8\). During the first phase of its pilot operation, which was decided for the period 1993 - 1996, the CARE data base had to deal with all operational problems and be ready for an overall evaluation. The results of this evaluation were positive and the European Commission with its communication on CARE data base opened the way for the further development of CARE into an integrated information system\(^9\).

Data collection and processing

CARE is a huge data base containing all national files with disaggregate road accident data in their original form (without any aggregation). These original data files are sent to CARE data base in electronic form (magnetic tapes, disks, etc.) once per year. Information for every single accident that has been recorded in the European roads is contained inside CARE data base and can be used for the production of aggregate reports (reports containing aggregate statistics)\(^10\). To get an impression of the magnitude of this data base it is mentioned that the computer memory required for the storage of the information available for all road accident data in ten member states for one year is more than 1 Gigabyte.

However, a number of incomparabilities make the coherent operation and use of CARE data base a procedure which is far than straight forward\(^11\). The different national accident data collection systems, the different national informatic file structures, the different information collected and the different corresponding definitions used\(^12,13\) are the basic problems that CARE project has to overcome in order to be fully operational and useful for the European road safety community.

CARE is developed on a main frame system using the Acumen application. The actual state of CARE programming is the outcome of an important amount of work effort for more than eight years. A considerable number of structural incomparability problems have been solved and some more are in the process to be solved. The system time response and memory management are already sufficiently developed and future improvements are expected to concern mainly the optimisation of user query procedures. The
on-line access of the actual users (national administrations) to CARE is a technical issue which when established does not present difficulties; the time response of the system depends on the telecommunication lines and is already considered today satisfactory.

The actual user interface allows the production of multi-dimension aggregated reports in a standardised way, combining a wide range of information. The user can choose at the moment among data concerning at least 5 years (1991-1995), 14 EU states (all except Germany), more than 400 variables and more than 2,500 values (not necessarily comparable between states).

Seventeen accident variables containing directly comparable values (basic information on the date of the accident and the characteristics of the persons involved) are available for processing since 1995. More variables and values can become directly comparable only by the application of special correction factors transforming national values (with different definitions) to common-definition values. In this way no intervention in the national data collection and processing system is necessary. Thus, an additional number of variables has been recently examined (CAREPLUS project\textsuperscript{14}) and today 28 accident variables containing some 600 comparable values (400 values refer to person’s age) are directly comparable. In the second phase of CAREPLUS project which will start within 1997, some more variables (about 15) and respective values will be examined.

Output reports and exploitation policy

There is no official publication issued by CARE as it is a project in pilot operation. Since 1996, several versions of test reports\textsuperscript{15} were issued which were destined only to the members of the governmental experts’ group on road accident statistics. These test reports concerned mainly accident information which is directly comparable (number of killed persons, age and sex of the road user, weekly and monthly distribution of the accidents, etc.) presented only in percentages. The further development of these test reports will constitute the future official CARE publication.

As far as access to data is concerned, on July 1997, seven EU member states have validated the accuracy of their national data contained inside CARE and gave thus their approval for access to their data to other EU member states under the reciprocity principle. Five other countries are in the process of validating their national data and three other states will enter the validation process phase when their particular technical problems will be solved.

CARE program structure does not give anybody access to primary disaggregate data as output reports are derived through specific partial aggregation procedures applied on disaggregate data. The outcome of a query is always a report produced by aggregating on the basis of a number of
parameters of the fully disaggregate data. This CARE program structure protects the individual data confidentiality.

All CARE reports can be produced today in English language. Additionally, CARE reports referring to a specific state can be produced in the corresponding state language. In the near future, the possibility to produce CARE reports in French and later in German language is foreseen, whereas in the long term all eleven EU languages may be supported.

Today, CARE is a closed system with access restricted to specially appointed national administrations. As the first phase of CARE pilot operation is reaching its end, exploitation plans are under development. More improvements in CARE technical and operational characteristics will take place progressively in the years to come towards facilitating full exploitation of the accident data base.\(^\text{16}\)

3. The FARS system

General characteristics

The United States of America is a federal state. The executive power of the USA is vested in the President who is supported by his administration. According to the US federal system, matters that lie entirely within state borders are the exclusive concern of state governments. There are, of course, many areas of overlap between state and federal jurisdictions but where the federal government exercises responsibility in the states, programs are usually adopted on the basis of cooperation between the two levels of government, rather than as an imposition from above\(^\text{17}\). The Department of Transportation (DOT) establishes the nation’s overall transportation policy which comprises among others the responsibility of the 68,000-kilometer interstate highway system as well as the road safety policy.

The National Center for Statistics and Analysis (NCSA) is an office of the National Highway Traffic Safety Administration (NHTSA), an agency in the Department of Transportation. NCSA is responsible for providing a wide range of analytical and statistical support to NHTSA and the highway safety community at large.\(^\text{18}\) To address the many analytic requirements, data are obtained from a variety of sources. The major data sources are NCSA’s own Fatal Accident Reporting System (FARS) and the National Accident Sampling System (NASS). FARS was established in 1975, and contains data on the most severe of road accidents, those in which someone was killed. NASS is composed of two systems - the Crashworthiness Data System (CDS) and the General Estimates System (GES).\(^\text{19}\)

The Fatal Accident Reporting System (FARS) of the United States of America contains data on all fatal road accidents within the 50 states, the District of Columbia, and Puerto Rico. The data system was conceived, designed, and developed by the National Center for Statistics and Analysis (NCSA) to assist the traffic safety community in identifying traffic safety
problems, developing and implementing vehicle and driver countermeasures, and evaluating motor vehicle safety standards and highway safety initiatives.

Data collection and processing

NHTSA has a contract with an agency in each state to provide information on fatal road accidents. FARS analysts are state employees who extract the information and put it in a standard format. Each FARS analyst attends a formal training program, and also receives on-the-job training. Data on fatal road accidents are gathered from the state’s own source documents, and are coded on standard FARS forms and transferred electronically to the central FARS data base. The analysts obtain the documents needed to complete the FARS forms from various sources including the Police Accident Reports (PARS), State vehicle registration files, State driver licensing files, State Highway Department data, Vital Statistics, Death certificates, Coroner/Medical examiner reports, Hospital medical records, Emergency medical service reports.

To be included in FARS, a road accident must involve a motor vehicle travelling on a road customarily open to the public, and must result in the death of a person (either an occupant of a vehicle or a non-motorist) within 30 days of the accident. The FARS file contains descriptions of each fatal accident reported. Each case has more than 100 variables that characterise the accident, the vehicles, and the people involved, containing more than 3,200 values (1,200 concern car makes). The specific data elements may be modified slightly at times, in response to users’ needs and highway safety emphasis areas. All data elements are reported on four forms:

- The Accident Form asks for information such as the time and location of the accident, the first harmful event, whether it is a hit-and-run accident, whether a school bus was involved, and the number of vehicles and people involved.
- The Vehicle and Driver Forms call for data on each accident-involved vehicle and driver. Data include the vehicle type, initial and principle impact points, most harmful event, and drivers’ license status.
- The Person Form contains data on each person involved in the accident, including age, gender, role in the accident (driver, passenger, non-motorist), injury severity, and seat-belt use.

Data quality control is a vital system feature. One important part of the quality control program is a series of consistency checks, which ensure that no inconsistent data are entered. For example, if an analyst codes 11:00 am as the time of the accident and “dusk” as the light condition, these codes would be rejected as inconsistent. Other checks are for timeliness, completeness, and accuracy. Statistical control charts are also employed to monitor the coding of key data elements over time. According to FARS standards, 95% of the total number of fatal road accidents must be entered in the system 90 days after the accident.
FARS data present a **high degree of homogeneity** as a result of long efforts of data harmonisation between road accident data collected at State level. This homogeneity is also due to the availability of an important annual budget (exceeding 4 MECU) devoted to suitable data transformations in each State, as well as to a strong negotiation power of the federal administration towards the States' governments, which undergo financial penalties if they don't use common data definitions.

The FARS system uses **links with other data** in order to produce accident indicators as well as additional information on the vehicle and the road. For example, national traffic counts are used for the formation of accident indices and the Vehicle Identification Number (VIN) recorded for each vehicle involved in the accident is matched with the corresponding number of the national vehicle registration file in order to extract information related to the vehicle (make, age, etc.).

The FARS system uses a **distributed data base architecture** based on PC using MS-DOS Fox-Pro software together with Pascal code under Windows 95. There exist two copies of the data base which are updated simultaneously, one in the headquarters and one in each State containing the corresponding State data. The complete FARS data file for one year is approximately 30 Mb.

**Output reports and exploitation policy**

**FARS output services** comprise a road accident statistics' yearbook, a CD-ROM version containing reports and original data based on FARS and NASS-GES systems, and on-line access to the fact sheets and other information via the NHTSA www page on Internet. Furthermore, files on computer tape are available free of charge in several formats amenable to analysis enabling users to process the data using their own computer system. Recently, a new FARS prototype allowing on-line queries became available to everybody on Internet and in the near future it will be possible for every statistics user to download parts of the original FARS disaggregate data.

The **road accident statistics' yearbook** uses both FARS and NASS-GES data and constitutes a comprehensive overview of the year's data, including some data of previous years with comparative analyses. This yearbook is issued nine months after the end of the reference year.

NCSA responds to over **3,000 requests for information** and sends out more than 50 computer tapes of FARS data each year. FARS data are used extensively within NHTSA, and requests are received from sources such as state and local governments, research organisations, private citizens, the auto and insurance industries, congress, and the media. FARS data can be used to answer many questions on the safety of vehicles, drivers, traffic situations, and roadways. FARS data can also be accessed at the state level by the FARS analyst to respond to state safety issues. To protect individual privacy,
no personal information, such as names, addresses, or specific crash locations, is coded.

4. Discussion

General characteristics

**Federal integration** in the EU is under process whereas in the USA it is established many decades ago. The central power is more important and more efficient in the USA than in the EU. It is worth mentioning that the central government in the USA disposes a considerable central budget whereas in the European Commission of the EU such budgets are limited. However, this gap in central power is diminishing due to recent developments in EU integration. On the other hand, the participation of the states in the central level decision making is direct in the EU whereas it is more indirect in the USA.

The fact that the central Government in the USA disposes more power than the European Commission in the EU is also reflected in many areas of transportation and road safety policy. Agreement on an all-states common approach for several transportation and road safety issues is easier to be reached in the USA than in the EU. This fact is also reflected in the homogeneity of **transportation legislation** as well as of road network and traffic characteristics (standards, etc.).

In most of the cases, **road safety policy** (traffic legislation, road infrastructure, vehicles, research, etc.) at USA federal level has presented more and more important initiatives already for some decades, whereas the EU road safety policy is very "young" and not very strong yet. Furthermore, there is a number of difficulties for implementing a EU level road safety policy as no culture of such a common approach exists in the various EU states. However, a number of exceptions exist where the EU has already reached a greater degree of homogeneity, such as the European driving licence, etc.

In both USA and EU there exist **road accident data bases with disaggregate data at state and central level**. The USA-FARS system is already operational since 1975, whereas the EU-CARE system started its pilot operation twenty years after. FARS system can show a considerable 25-years effort since its launching, whereas the 10-year CARE development effort can only be considered as a good start. It is worth mentioning that, in 1996, the annual FARS budget (~19m$) was almost ten times more important than the equivalent CARE budget and the personnel working for the Federal US accident data systems (~360) is a significant multiple of the personnel working in the EU accident data systems (less than 10).

Differences in the structure of the two political systems as outlined above can greatly explain differences of the two road accident data bases with disaggregate data. **History of road safety policy** at state and federal level at
the USA and the EU is also the reason for the different approaches used in FARS and CARE systems.

The basic characteristics of the two systems have been put together in the following comparative Table 1, allowing for a general direct comparison.

Table 1. CARE- FARS, comparative description of basic characteristics

| GENERAL INFORMATION | | | |
|--------------------|-------|-------|
| 1 Name | CARE | FARS |
| 2 Owner | European Union | United States of America |
| 3 Competent service | European Commission - DG Transport/Eurostat | Department of Transportation - NHTSA |
| 4 City | Luxembourg | Washington |
| 5 Starting decade | late 80s | early 70s |

| DATA INPUT & PROCESSING | | | |
|------------------------|-------|-------|
| 6 Type of data collection | transfer of national files in disks/tapes | electronic transfer based on a unified form |
| 7 Number of sources per state | 1 | several |
| 8 Accident data | disaggregate | disaggregate |
| 9 Other data | aggregate | aggregate |
| 10 Available time series | 1991- | 1975- |
| 11 Number of states | 15 | 52 |
| 12 Number of accident variables used | ~ 30 | ~100 |
| 13 Number of accident values used | ~ 600 | ~3200 |
| 14 Number of units used | 4 | 4 |
| 15 Use of correction factors | Yes | No |
| 16 Software used | ACUMEN(VM) | FoxPro (MS-DOS), Pascal (Windows 95) |
| 17 Hardware used | Main frame | PC |

| OUTPUT SERVICES | | | |
|-----------------|-------|-------|
| 18 Access | National admin. only | Everybody |
| 19 Publication | Test reports | Annual |
| 20 Completed after X months of the reference year | 12 - 15 months | 9 - 10 months |
| 21 Number of tables | 22 | 117 |
| 22 Notes on road safety | No | Yes |
| 23 Definitions | No | No |
| 24 Languages | English (more in the future) | English |
| 25 | | |
| 26 Off-line electronic form (disks) | No | Yes |
| 27 On-line electronic form | Yes | Yes |
| 29 Accident data available on the www | No | Yes |

1: Other data refer to additional data used in accident rates (vehicles, vehicle-kilometres, etc.)
2: With common label for all countries considered in the data file
3: As units the number of accidents, killed persons, injured persons and vehicles involved are considered

Data collection and processing

A basic common characteristic of the collection of disaggregate road accident data is the fact that it is a state competence in both the USA and the EU. The central systems (FARS and CARE) have been developed in parallel and on the basis of the data from the state systems already operational for decades.
A basic difference of the two systems, responsible also for a number of additional differences, is the fact that CARE contains data on fatal and injury accidents whereas FARS contains data on fatal accidents only. This basic difference explains why the computer memory required for the storage of the information for all road accidents data (fatal and injury) in 15 states for one year inside CARE is more than 1 Gigabyte whereas the equivalent memory for 52 states inside FARS (fatal only accidents) is approximately 30 Megabytes.

Another basic difference between the two systems is the fact that FARS requires further processing of the state data at state level (by special agencies) whereas in CARE further processing of the state data takes place at central level. In the FARS case, extra work is foreseen at state level but then the complete data set is homogeneous, whereas in the CARE case, no extra work is asked from the states but the central level data processing cannot reach fully homogeneous data sets. FARS people processing data at state level are numerous (about 300) in comparison with the very few CARE people working on the homogenisation of state data.

The different volume of data contained in the two systems is the main reason that CARE has been developed and is operational today in main frame machines whereas FARS is operational today in PC. FARS uses a more decentralised system architecture as there exist two copies of the FARS data base [at central level and in each state (only with state data)] which are simultaneously updated whereas CARE uses a centralised structure with the complete data base available only at central level.

Special attention is given to data quality control in both systems. In FARS, data quality control is carried out in two stages, first by the state data collection agencies which use several sources to cross-check the data and then by the central system with the use of mainly automatic cross-checking procedures. Equivalent detailed data quality control does not take place at CARE central level as data are considered as correct; quality control has been thoroughly carried out in the original data files at state level. However, a number of further data processing is necessary for CARE data in order to enhance inter-state data comparability.

Links with other data (traffic, vehicle fleet, etc.) and data files is possible for some data sets in FARS whereas it is a priority for implementation in CARE system. In the case of CARE these links with other data present similar comparability problems between states as those found in the accident data base. A basic problem common for both systems is the fact that the degree of disaggregation of these other data cannot be in most of the cases similar to that of the road accident data, limiting thus disaggregate data analysis.

Both in the USA and the EU, road accident data are collected under the principle of confidentiality and all personal information is deleted from the disaggregate data set to which external people have access (directly or indirectly). The CARE system is also examining alternative ways for a higher degree of data confidentiality which can be ensured by a system feature.
according to which CARE reports refering to total number of accidents less than 3 will not be possible to be extracted, ensuring thus that no single accident can be identified.

Output reports and exploitation policy

FARS output services comprise a complete set of formats (statistics' yearbook, fact sheets, CD-ROM, on-line access, etc.) meeting users’ needs in a very suitable way. FARS output services have been improved over the years as a result of a constantly increasing demand for federal level statistics. On the other hand, CARE output services policy is under development. It is expected that a policy very similar to the current EC-Eurostat policy in the field of statistics' dissemination, offering a wide range of products (yearbooks, CD-ROM, on-line access, etc.), will be initially adopted for the exploitation of CARE potential.

FARS has a well-established mechanism for the satisfaction of the various requests for information whereas CARE actual pilot operation phase can not yet cope with the various users’ requirements. It is worth mentioning that in contrast with the USA where demand for federal level statistics (not only for road accidents) is a rather predictable and stabilised phenomenon, equivalent European-level demand for statistics is still an unknown and rapidly expanding phenomenon, especially in the field of road safety. The rather young European-level road safety community is not used to work at European level in such a detail and therefore its requirements are not yet clear.

It is noted that both CARE and FARS standard interfaces do not give access to original disaggregate data as original data can only be used for the extraction of aggregate reports. Original CARE data are not available to anyone whereas parts or the complete set of the original FARS data are available to a limited number of special users (50 magnetic tapes are forwarded every year, download via Internet, etc.).
5. Conclusions

The above comparative analysis of the two large scale databases with disaggregate data, the CARE system of the European Union and the FARS system of the United States revealed not only differences and similarities of the two systems but also differences and similarities of the way road safety is treated in the two continents. In many cases, the various strengths and limitations of the two systems reflected (and were explained by) equivalent strengths and limitations of the road safety policy in the two continents.

The FARS system, concerning only fatal accidents and already operational since 1975, has reached its maturity and is a powerful tool for federal-level road accident analysis, intensively used by the various actors involved in road safety issues in the USA. On the other hand, the CARE system, in pilot operation since 1995 and with limited data homogeneity, established a new and significantly upgraded standard for European wide road accident comparisons but fails to meet its potential users’ needs yet.

The differences between the two political systems, especially as far as the interrelation between central power and states is concerned, lie behind the differences between the two road accident data systems. The strong negotiation power of the central Government in the USA vis-à-vis the states made more easy the creation and operation of a uniform central system, whereas the lack of an efficient common road safety program in the EU lies behind the way CARE has been evolved (without any intervention to the states systems) leading to a long development phase and an actual system with limited comparability.

The CARE system development can benefit a lot from the experience gained during the 25 years of development and operation of FARS system. Especially, the FARS experience on a number of technical issues concerning the selected common variables and values as well as the exploitation policy could be of special interest for CARE. For example, the evolution of CARE towards a system containing data, at least initially, only for fatal accidents (and possibly later for hospitalised persons) could lead to a better manageable system with fewer problems of data incomparability. Additionally, FARS could also benefit from some CARE experiences, such as the development of methods for overcoming incomparability problems (whenever it applies to the USA data) or from the European approach for some accident characteristics and their classification.

A future challenge for both road accident data bases with disaggregate data could be the establishment of a common platform which could allow for comparisons between the EU and the USA states at the detailed level that disaggregate data offer. Such a convergence of the two systems could open new horizons and perspectives in international road accident data analysis and allow for very useful exchange of experience on traffic safety on the two continents. The critical issue of this convergence is the establishment of common variables and values of the data to be collected in the two continents.
The future convergence of CARE and FARS requires further EU-USA common research at technical level on one hand and related political decisions on the other. Mutual cooperation for a common data structure or leading to a common structure through transformation rules could be very fruitful. The examination of such a possibility could be of great interest for the road safety community in the two continents and could also be a great opportunity for the further comparative examination of road safety in the two continents.

6. References

3. NATIONAL TECHNICAL UNIVERSITY OF ATHENS.- Overview of international road accident data files.- NTUA Department of Transportation Planning and Engineering.- Athens.- November 1996.
24. NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION.- Request for applications to support the fatal accident reporting system (FARS) program.- US NHTSA.- Washington.- 1996.