Full Length Research Paper

Towards an analytical information system of traffic accidents in the function of traffic safety monitoring

Ognjen Pantelic1*, Dalibor Pesic2, Milan Vujanic2 and Dragana B. Vujaklija1

1University of Belgrade, Faculty of Organizational Science, Jove Illica 154, 11000 Belgrade, Serbia.
2University of Belgrade, Faculty of Transport and Traffic Engineering, Vojvode Stepe 305, 11000 Belgrade, Serbia.

Accepted 23 December, 2011

Quality monitoring, analysis and evaluation of traffic safety situation is the first and most important strategic step in establishing a traffic safety protection system. With the goal of establishing the protection system and traffic safety management system, it is necessary to use quality data from traffic accident databases. Apart from quality, with respect to different quality criteria, it is also necessary for the database to be both available and up-to-date. Database availability and up-to-datedness is extremely important for scientific and professional research, because that research helps define certain dependencies that can affect the choice of traffic safety improvement measures. The quality of the data which are in the database also has to be very high. This can be achieved by establishing an information system and information flow by relevant subjects that possess information on traffic accidents towards the central database. For that purpose, the data coming from the police, insurance companies, health authorities and other relevant authorities need to be processed and adapted to the form of the data from the central database. This method produces comprehensive and quality data on traffic accidents, compared to when the database is ‘filled’ by only one source, for example the police. The existing traffic accident database in Serbia is owned by the Ministry of Interior, and that database is not suitable for research and analysis, it is not public and the procedure for obtaining data is long and uncertain. The new Road Traffic Safety Law in Serbia gives different users the opportunity to access traffic safety data, but it is somewhat restricted or limited. With that in mind, there was a need to define and establish a new traffic accident database. This paper presents the possibilities of traffic safety analysis based on the integrated traffic accident database model, which was primarily compiled for the needs of the City of Belgrade within the BERTAAD Project (Belgrade Road and Traffic Accident and Analytical Database). Possibilities of the database are reflected in different levels of access to the database, depending on the user, different levels of data detail that can be accessed; they are also easier to work with because the database is user-oriented.

Key words: traffic safety, analysis, estimate, database model, access.

INTRODUCTION

For a description of a phenomenon and its behavior, usually data that refer to that phenomenon are needed. By analyzing the data on the phenomenon it is possible to analyze the phenomenon and possibly define some dependencies between the phenomenon and other indicators. The established dependencies can point to possible measures with the goal of changing the observed phenomenon.

Data analysis and processing can be primary, that is, when there is a primary reason to analyze the phenomenon. For example, we analyze traffic accidents from the aspect of spatial distribution, with the goal to define dangerous spots, that is, black spots, where there is a significant piling of traffic accidents. Secondary, tertiary and other types of analysis can point to different dependencies; for example, analysis of participants in traffic accidents at previously established black spots can
show us that pedestrians are especially endangered at those black spots. Established dependencies can, directly or indirectly, point to the subject of action and the measures that need to be used. In the given example of traffic accident analysis it can be concluded that primary database analysis can, for example, show the spots that need an action, secondary type analysis show who needs the action and so on.

With traffic accidents, as a phenomenon, traffic accident data can be analyzed and the data usually can be found in appropriate databases. The main reason to analyze traffic accidents is to establish certain dependencies of traffic accidents and characteristics of traffic accidents; by changing those relations and dependencies we can actually affect the change in the number and consequences of traffic accidents. In that way, we create the possibility for traffic safety improvement.

For quality analysis of traffic accidents it is especially important to have quality data on traffic accidents. However, in practice, data on traffic accidents that would enable good and quality traffic accident analysis are not always available. The quality of a database is observed in the sense of data structure, integrity, redundancy, data normalization, etc. Also, up-to-datedness of the data is important, as well as, its availability to users; it is especially important that they are available to researchers and scientists who can, by analyzing the data from databases, ultimately improve traffic safety.

The problem of traffic accident data collection into appropriate databases dates back to mid-20th century. Namely, perceiving the problem of traffic safety and growing numbers of traffic accidents, United Nations defined a traffic accident as 'Accident that occurred at a place open for public traffic or that started at such a place, in which one or more persons died or were injured, with participation of at least one moving vehicle' (Vujanic et al., 2005). It can be concluded from the definition that property damaged by only accidents are not, according to that definition, road traffic accidents, which is due to the fact that undeveloped countries of the world did not record or create any kind of traffic accident databases. Even traffic accidents with fatal consequences were not recorded. Apart from the problem with data collection, for the needs of quality analysis of traffic accidents, there is also the problem of the so-called 'shadow accidents'. These accidents are accidents that usually occurred with very little property damage and the Law allows the participants in the accident to come to an agreement about the guilt and damages. Since these accidents are with very little property damage for some serious analysis and actions with the goal of improving road traffic safety, they are not interesting on the global level, compared to, for example, accidents with fatalities.

In this paper we used definition of traffic accident from the Serbian Road Traffic Safety Law as ‘accident on road, and involves at least one moving vehicle, and resulting in one or more persons killed or injured or material damage only’.

Keeping in mind what has been previously said, the subject of this paper are road traffic accidents in Serbia that were recorded in a database and possibility of using those databases for quality analysis of road traffic safety. The paper analyses the existing state of traffic accident data collection in Serbia. In other words, traffic accident databases by Ministry of Interior (Mol), health authorities, insurance companies and other relevant external data sources. Structure and scope of the data, possibility of connections and use of data from different data sources for analysis of road traffic safety were analyzed. Keeping that in mind, the paper presents the model of integrated traffic accident database that can be widely applied in the sense of road traffic safety analysis in different ways down to the lowest data structure. While defining the method of traffic accident data collection, international databases IRTAD (International Road and Traffic Accidents Database) and CARE (Community database on Accidents on the Roads in Europe) has the biggest influence on the model (Vujanic et al., 2005); their basis is mainly comprised of the data coming from the police and health authorities, and as such, a similar concept is adopted in the database model presented in this paper as well and the data from other external sources are included, which can be significant for later traffic safety analysis (for example, insurance companies, road directorate, etc.).

Quality traffic accident data analysis are enabled in the proposed model by application of the so-called OLAP (On Line Analytical Processing) database concept, which represents the possibility of creating a multidimensional database with a large number of options (Vujanic et al., 2005). OLAP concept is developed over a certain DBMS (Database Management System) as a set of programs which give the user an integrated tool for adding, deleting, access and analysis of data that are stored at a location (Turban, 2009). Namely, OLAP concept actually represents the warehousing approach, which enables fast manipulation, aggregation and local calculations for trend analysis (Barry, 1997; Inmon, 1996). It can also be defined as a wide set of analytics software for collection, consolidation, analysis and access to information, with the goal of enabling business users to make better business decisions (Olson, 2010). This is especially important for strategic decision-making, because strategic decisions require predictions, statistics, simultaneous functions, time series analysis, etc. Advantages of OLAP are: high performances in execution of complex enquiries, competitive processing, ability to include data from different sources, use of a language especially designed for data analysis, etc. Another very important advantage of OLAP is that it is user-oriented, because no special computer knowledge is needed for working with OLAP: OLAP users are usually experts exclusively in their own field.
MATERIALS AND METHODS

Availability of data on a phenomenon enables the analysis of that phenomenon, and quality data on a phenomenon enable quality analysis, conclusions and definition of appropriate measures with the goal of changing the phenomenon. For that reason, with the goal of defining methods for creation of analytical traffic accident database, we conduct analysis of experiences with traffic accident databases. International databases IRTAD (http://internationaltransportforum.org/irtad/index.html, last visited on 25th June, 2011.) and CARE (http://ec.europa.eu/transport/road_safety/specialist/statistics/care_reports_graphics/index_en.htm, last visited on 25th June, 2011.) are analyzed, as well as national databases (http://www.transportstyrelsen.se/en/road/STRADA, last visited on 25th June, 2011. and http://www.vicroads.vic.gov.au/Home/SafetyAndRules/AboutRoadSafety/StatisticsAndResearch/CrashStats.htm, last visited on 25th June, 2011.). We concluded that the analyzed national databases are at a very high level in the sense of traffic safety analysis. In other words, the analyzed databases enable a variety of different traffic safety analysis, which makes them an excellent base for traffic safety improvement. On the other hand, IRTAD and CARE databases are international databases created, among other things, with the goal of unifying data from national databases. In other words, 'step-by-step' IRTAD and CARE databases raise the standard and define the structure and scope of data stored in them. Member countries of these databases pledged to obey the demands of international databases and send their data on traffic accidents in the form of a previously defined report. The existence of such international traffic accident databases enables, among other things, mutual comparison of traffic safety conditions in different countries.

Examining the existing conditions in traffic accident data collection, and especially, the existing traffic accident databases in Serbia, we analyze databases by: Mol (Ministry of Interior), health authorities and insurance companies from the aspect of data structure, possibilities of use of existing data, integration into one unit, as well as, possibilities for improvement of existing ways of data collection. Previous state of information system we have analyzed is not designed for analysis. There is a solid data integrity, they are unable to achieve high level of database normalization. Access to data for a reports is carried out directly through SQL. Such an approach required a good knowledge of database structure and query language. We keep in mind that, for example, the Mol database is 'filled' from the so-called SN1 questionnaire. SN1 questionnaire is filled in by a traffic policeman who performs or participates in traffic accident investigation. After that, from the so-called ‘paper’ version of the questionnaire, the traffic accident data are entered into the Mol electronic database. SN1 questionnaires are categorized according to the severity of traffic accident consequences into: accidents with fatalities, accidents with injuries and accidents with property damage only. SN1 questionnaire consists of three units:

1. General data on the traffic accident (date and time of the accident; data on accident location – street and number, road kilometre, a closer determination; type of accident; road characteristics; ...)
2. Data on vehicles that participated in the accident (number; type of vehicle; registration; year of manufacture; ...)
3. Data on traffic accident participants (number; category – driver, pedestrian ..., sex; birth date; birth place; living address; consequences; ...).

During the analysis of traffic accident data from the Mol database certain shortcomings are defined; those shortcomings could be overcome by creating a new, integrated base and redirecting the actions taken by policemen. Some of the shortcomings of the existing Mol database are:

1. Data on accident location are not precise enough because accident location is defined based on street and house number, and streets can change names; definitions based on road kilometre are also not appropriate because roads can change length for different reasons, for example road reconstruction, etc. Due to these shortcomings, it is necessary for the new database to introduce a new system for definition and marking of accident location, and in that sense it would be best to use GIS (Geographical Information System) data and digital maps.
2. Accident description is not entered into traffic accident, and there are situations where it is impossible to determine the cause of the accident; the description of the traffic accident and situation prior to the accident could help define the cause of the accident more precisely.
3. Vehicle data do not give enough information, because no data on vehicle condition, in the sense of the vehicle’s general condition, kilometers travelled, pneumatics, etc. are recorded.
4. Consequences to participants do not have information on injury location and its severity.

Similarly to the previous analysis, an analysis of other significant databases are conducted (health authorities and insurance companies) and it is concluded that existing traffic accident databases do not enable quality traffic safety analysis due to their shortcomings. The defined shortcomings, especially the opportunity to create a quality integrated traffic accident database which would enable different types of analysis, are an additional incentive for a more detailed analysis of the existing condition and definition and creation of a new traffic accident database.

Through analysis of present experiences in collection and distribution of traffic accident data it can be concluded that:

1. The need for reliable traffic accident data is constantly growing.
2. Comparisons of traffic safety levels among countries are becoming more and more important, so it is necessary to coordinate national traffic accident databases on international level.
3. Existing traffic accident databases are not unique and do not quite enable application of traffic safety improvement measures in an efficient manner.
4. New traffic accident database model has to enable quality analysis and reliable basis for decision-making and implementation of measures for traffic safety increase, and, at the same time, provide the needed information prescribed by international standards of report, in the first place at the European Union (EU) level.

Considering the previously given results of analysis, a platform where traffic accident database will be realized is chosen, and that is OLAP, based on warehousing data storage concept as a useful way of data analysis. We chose the OLAP taking into consideration the necessary characteristics of the system, which requires read only type of data processing. Also as mentioned earlier, OLAP is user-oriented and no special computer knowledge is needed for working with OLAP. Furthermore, OLAP users are usually experts exclusively in their own field. The goals that need to be fulfilled by traffic accident database are (Vujanic et al., 2006):

1. To be an efficient service for all interested subjects, to enable them to quickly obtain and efficiently process all traffic accident data of specific interest to them.
2. To help decision-makers (local city authorities) to implement measures for traffic safety improvement at the city level, that is, it needs to enable timely decision-making based on reliable data.
3. To represent a unique source of traffic safety data; that is, to
incorporate the existing data from different sources into this database, as well as research results from scientific and higher education institutions.

The concept of traffic accident database development, considering goals and needs that a database needs to fulfill are (Vujanic et al., 2006):

1. Integrality – the database should enable unique treatment of all data, archiving and use of the entire database fund.
2. Modularity – in development, especially in implementation, to enable a modular approach.
3. Flexibility – respecting diversity of data in form, structure and sources.
4. Transparency – authorised users need to be enabled to access all data within their competence, including Web access.

By analyzing possibilities for establishing a system of data collection into a database, it is determined that the database basis should be made of data obtained from the Ministry of Interior database. Those data are collected by traffic policemen in the field, based on a previously defined form, so-called SN1 form (Figure 1).

Database construction in the form of an n-dimensional polyhedron (so-called OLAP cube) is presented in Figure 2.

Database also includes data obtained from health authorities and insurance companies, as well as all other available data related to traffic accidents (for example, road directorate data, statistical office, etc). Considering all that has been said previously, a database model is suggested; it has all relevant data from the Mol database, and it can accept a number of data from different sources. Data on individual traffic accidents are grouped into levels: accident level, vehicle level, participant level. Each level contains data that describe certain levels in detail (Figure 3). Such database enables:

1. Pre-aggregation of frequently accessed data – for quick ad hoc enquiries.
2. Multidimensional data model suitable for selection, navigation and display.
3. User can go down to elementary data.
4. Creation of infinite number of dimensions.

Considering the current condition and development of traffic accident data collection process in Serbia, it is suggested that introduction and operation of the new integral database be implemented in two phases. Namely, due to current development of automatization and computer equipment of external institutions, the first phase of development and implementation of traffic accident database would include the basis of the database from the police database. Other external institutions (health institutions, insurance, etc.) would submit their traffic accident data on previously defined paper forms at a collection point for further processing and entering of the data into the database. The model for data collection and distribution in phase I is presented in Figure 4. The represented model anticipates and makes reports and analysis possible on several levels, based on data from BERTAAD database:

1. Standard reports submit to previously defined users (management organs at different levels (city, municipality...), Ministry of Interior, Roads Directorate, IRTAD...). These reports have a known form and frequency of generation. There are previously defined generation procedures.
2. 'Ad hoc' analysis and reports – analytical procedures are prepared for them, based on which it is possible to perform an unlimited number of different analyses. Other users will be able to

![Figure 1. SN1 form – basic element of BERTAAD database (Vujanic et al., 2005).](image-url)
Figure 2. Consolidation of data tables from SN1 form into BERTAAD base (Vujanic et al., 2005).

Figure 3. Simplified 3D view of OLAP base (Vujanic et al, 2005).
access analytical procedures, with a limited access to data, depending on their competence (researchers, insurance, professional institutions).

3. Web access – a part of accident data, respecting the principle of anonymity and valid regulations will be available to the public, via the Internet.

In the second phase of the database development, conditions for higher level of automatization in data collection and providing more comfortable conditions for data access for analysis are expected. The model for data collection and distribution in phase II is shown in Figure 5. In the second phase of database development the Ministry of Interior is expected to semantically enrich its database and migrate to a more modern hardware platform; direct download of data from MoI database is anticipated; the data will go through simpler transformation procedures and BERTAAD updating. It is expected that MoI will, among other things, submit data on GIS coordinates for each traffic accident. Since there is a possibility for the database to be connected to appropriate digital maps, GIS coordinates and data clustering would enable different analysis of spatial distribution of traffic accidents. Namely, the fact that all the data would be able to be ‘put down on the map’ is extremely
important, because in that case the traffic accident analysis considers the so-called 'complete set of events'. The importance of this possibility is even greater when experiences in Serbia show that with the traditional way of 'putting accidents down on the map' it can be done with about 54% of accidents (Vujanic et al., 2009). The reasons for such low percentage of precisely defined location of accidents are that current analysis mostly use MoI database, which was, apart from being outdated and unavailable, also incomplete.

Improving the equipment of external institutions and communication infrastructure, external institutions will, in the second phase, submit traffic accident data electronically, and that data will be stored in BERTAAD through previously defined and automatized procedures. The interest of institutions to participate in the traffic accident data collection process should be in the possibility of obtaining complete data on accidents, for own analytical needs, etc. That opportunity will be provided through electronic distribution of data, with the ability of remote access, while respecting access limits, depending on competencies of each individual institution.

RESULTS

Possibilities of traffic safety analysis by application of the presented integrated traffic accident database model are extremely numerous. However, exit from an integrated database depends on the model development phase and quality of data that are stored in the database. Since at the time of creation of this paper phase I was still current, we present here the results obtained with possibilities of traffic safety analysis on several examples based on data analysis from BERTAAD database for phase I.

Abilities of the database are 'export' of data to some of the data analysis programs, e.g. Microsoft Excel, which enables table and graphic traffic accident data analysis. Apart from that, report formatting enables different appearances and variations of visual representation of traffic accident data analysis. An analyst or decision-maker is now able to obtain various reports by simply dragging the data into, for example, a Microsoft Excel table. Appearance of one of the report forms is given in Figure 6. Some of the possible output results of traffic accident data analysis in tables are presented further in the paper:

1. Total number of accidents and persons, classified by months of a year (Figure 7).
2. Total number of accidents and vehicles, classified by type of vehicle and consequences (Figure 8).
3. Total number of accidents, vehicles and persons, classified by sex and driver's license category (Figure 9).

The presented numerical examples of output results of possible traffic accident data analysis refer to phase I, so the possibilities of database analysis are limited, on one hand, but they are much more advanced compared to current traffic accident data processing. Namely, the new model of traffic accident database enables faster and simpler access and faster and simpler analysis and calculations that enable decision-makers to make timely decisions and apply appropriate interventions with the goal of traffic safety analysis and management.

Since the beginning of phase II mostly depends of external factors, that is, on the automatization speed of data collection by external institutions (health institutions, insurance, etc.), it is difficult to anticipate when it would start, but the important thing is that the model anticipated the contents, pre-conditions and operation mode for phase II. Once phase II starts, greater degree of process automatization, as well as faster access and more quality traffic accident data analysis are expected (Vujanic et al., 2006). Proposed database model is only a trial implementation in practice in Serbia, because of lack of funding.

DISCUSSION

Familiarization with the current condition is the precondition for definition of goals of the desired condition, that is, for direction of measures towards reaching those goals. Only precisely defined current condition in traffic safety enables precise definition of the desired state, i.e. the goals of the desired state. The goals need to be defined both quantitatively and qualitatively, and experiences in the world suggest that highly set goals motivate the society and interested parties to achieve them and, in that way, better results in traffic safety improvement are achieved (Elvik, 1993, 2001, 2008; Broughton et al., 2000; Rumar, 1999). An inevitable way to analyze current conditions is traffic accident analysis. Traffic accident analysis results in the so-called objective risks. In other words, based on traffic accidents it is possible to define the so-called spatial objective risk, that is, locations of increased vulnerability in traffic, the so-called 'black spots'; then, it is possible to define the so-called temporal objective risk, that is, time and period of increased traffic vulnerability, and one of the most important dimensions of traffic accident analysis is reflected in the possibility to define vulnerable categories of participants in traffic (for example, pedestrians, children, young drivers, etc.), while other traffic accident analysis can also contribute to problem perception and definition. Only if the problem is properly defined there is a possibility that it can be solved, and that is the goal of the analysis of the current state.

It can be concluded from the previous that the main precondition for quality analysis and definition of the current state is existence of a quality and up-to-date traffic accident database. Analyzing experiences with traffic accident databases and analyzing the current state of traffic accident databases in Serbia, it is concluded that it is necessary to define a new traffic accident database which would enable more quality traffic safety analysis.

The integrated database model has data from the Ministry of Interior traffic accident database as its basis,
Figure 6. Total number of accidents (Total), vehicles (Broj Vozila) and persons (Broj Lica) by consequences (Posledice Lica) (Vujanic et al., 2006).

and apart from those, the base also contain data obtained from other external institutions (health institutions, insurance, road directorate, etc.). The proposed integrated traffic accident database model enables total functionality and use of all data, moving down even to the lowest levels in data structure. Apart from that, the presented model has the option to increase the number of external institutions, that is, it is able to accept data from different sources. The previous is enabled by including OLAP database concept into the integrated traffic accident database model. However, it needs to be emphasized that this concept of so-called data warehousing is especially convenient for systems that do not have significant oscillations in operation, because the abilities to predict a phenomenon might be faulty, which is not good for the so-called strategic management by decision makers. As we move towards strategic management, details of the needed information are rejected and we move towards aggregate data (Davies, 2009). In any case, the advantages offered by OLAP in integrated traffic accident database are: new view of data, fast trend identification, simple use of tools, expertise of new values, shorter report cycle, improved error detection and misuse, profit from better analysis of new solutions' efficiency, etc.

CONCLUSION

Considering the current degree of development of traffic accident data collection and distribution process automatization at all relevant data sources (Mol, health institutions, etc.), the integrated traffic accident database includes introduction of the database into total operation in two phases. In phase I, with lower degree of automatization, apart from traffic accident data from Mol,
which would be submitted electronically, other external institutions would submit their traffic accident data on previously defined paper forms, at previously defined time and to a place that would be a so-called data collection center. After entering the data into the base, they can immediately be analyzed and appropriate desired reports can be obtained (standard, ad hoc reports, etc.). Second phase of traffic accident database development would include a higher level of automatization in data collection and provision of more comfortable conditions for data access for analysis.

Apart from the fact that the suggested traffic accident database model enables ‘connection’ among several traffic accident data sources into one unit, the model anticipates and also enables new modern systems. Modern systems, such as GIS, also integrated into the analytical traffic accident database model, additionally simplify and enable traffic safety analysis, especially from the aspect of spatial analysis of traffic accidents, consequences of traffic accidents, as well as all other characteristics of accidents, which are stored in the database.

Furthermore, it is necessary to monitor system introduction into phase II, on the suggested database model, and, after that, monitor the operation and results provided by phase II, through more detailed operation of the database.

Practically, considering all that was previously said, it can be concluded that the suggested BERTAAD integrated database model provides:

1. Possibility to quickly obtain statistical information, vertically and horizontally, in connection to any characteristics of a traffic accident;
Figure 8. Total number of accidents (Total) and vehicles (Vozila) by type of vehicle (Vrsta Vozila) and by consequences (Posledice Lica) (Vujanic et al., 2006).
Figure 9. Total number of accidents (Broj Nezgoda), vehicles (Broj Vozila) and persons (Broj Lica) by sex (Pol) and by category of driver's licence (Kategorija VD) (Vujanic et al., 2006).
2. Access to all available data related to a specific characteristics of a traffic accident (all accidents at a location, caused by a type of vehicle, at a specific time interval and the like);
3. Simplified decision-making within the field of road directorate and decision makers;
4. Possibility of fast composition of analysis and reports on various criteria;
5. Local and remote access.

REFERENCES

Barry D (1997). Data warehouse from architecture to implementation. Reading, MA, etc: Addison-Wesley.
http://www.transportstyrelsen.se/en/road/STRADA (last visited on 25.06.2011.)
http://internationaltransportforum.org/irtad/index.html (last visited on 25.06.2011.)