# THE CONSTRUCTION OF ROAD ACCIDENT ANALYSIS AND DATABASE SYSTEM IN MALAYSIA

Hizal Hanis H.[1], Sharifah Allyana S.M.R.[2]
Accident Database System and Analysis Unit,
Malaysian Institute of Road Safety Research (MIROS)
Lot 125-135, Jalan TKS 1, Taman Kajang Sentral, 43000 Kajang, Selangor, Malaysia

[1]hizalhanis@miros.gov.my, [2]allyana@miros.gov.my

## ABSTRACT

The number of fatalities due to road accidents in Malaysia has been consistently above 6,000 over the past few years. Although the fatality index and rates are declining, the actual figure is still at an alarming stage. In order to help in reducing the number of road accident and fatalities, the availability of a reliable and comprehensive road accident database and analysis system is critical to understand the pattern of accident causation. The Royal Malaysian Police (RMP) plays a major role in road accident data collection and throughout the years, there has been an abundance of data available for analysis. However, the existing road accident system in Malaysia is more focused onto record management and basic data analysis. A new advanced data analysis system has been developed to take advantage of the available accident data. Road accident characteristics gathered by RMP were structured and segregated into a new comprehensive database system. Better structured data will create conditions for deeper analysis, aiding in the formulation of evidence-based research on road safety and enabling better road safety interventions as well as performance monitoring.

Keywords: Road accident surveillance system, Police report, Road accident database, Road safety intervention, Performance monitoring

## 1. INTRODUCTION

Road accident is a serious problem in Malaysia. In year 2008 there were 373,047 road accidents that lead to 6,527 fatalities. The number of fatalities due to road accidents in Malaysia has consistently been above 6,000 since year 2003, whilst the indices used for road safety indicator has shown only slight decline over the years. Due to this alarming figure, the government has launched the Road Safety Plan 2006-2010 in March 2006 with the objectives to reduce fatality rates to 2 fatalities per 10,000 registered vehicles, 10 fatalities per 100,000 population and 10 fatalities per billion vehicles kilometre travelled (VKT) by year 2010 (Ministry of Transport, 2006). Since then, numerous efforts have been made by the government to improve road safety condition in order to achieve the target. However, a prerequisite to improve road safety is to have a comprehensive road accident database and analysis system. Collection of quality, accurate and reliable data that are collected over a period of time are needed in order to understand the factors influencing the arising figures of road accidents and injuries (Lyons et al., 2008). In addition, advanced road accident analysis system is needed to help strategise road safety initiative as well as inculcate better understanding of road accident causation. Furthermore, accident data is critical to monitor and evaluate the effectiveness of road safety interventions introduced by the government and road authorities.

Since 1991, there is abundance of accident data available when the Royal Malaysian Police (RMP) decided to established the Microcomputer Accident Analysis Package (MAAP) to manage accident data with the use of road accident form POL27 (T.H. & Radin Sohadi). MAAP is an accident data management system based on the DOS operating system. As technology progressed, in year 1997, RMP introduced a new Windows based system called Computerized Accident Recording System (CARS) to manage the accident data while MAAP still existed in operation. In 2005, MAAP was terminated and CARS is used solely to manage the data ever since. CARS is also capable to perform record maintenance and cross-tabulation analysis to provide general road accident statistics.

Although there is already a system exists to capture road accident data, the aim to establish a comprehensive database system has not been fully realised. This is due to the current system structure where each police station in the 145 districts of the country, maintaining individual database. The compilation for all the districts' accident data are later done through the police headquarters quarterly. However, the compilation has introduced several issues especially data quality problem. Data integrity and checking probably has not been done at the district level thus making the compilation at the national level troublesome. Moreover, for variables without predetermined values, each district will have their own reporting style.

On the other hand, road accident data were not fully utilized for decision making and performance monitoring due to the existing system unable to perform extensive and detailed analysis on road safety. Therefore, accident data are often kept just for record keeping purposes rather than using it as source of intelligence. Therefore, a new system is needed to address the issue of nationwide data compilation as well as for data analysis.

## 2. METHOD

### *2.1 Requirement Analysis*

A prerequisite for improving road safety situation in Malaysia is to have a comprehensive set of road accident data. As data are collected separately per district basis, individual dataset from each district will need to be gathered and combined to form a national road accident dataset. The dataset will later need to be stored and restructured into a database where data mining and analysis will be possible.

In addition, a new system for data analysis is also needed. The system will use the road accident database as the source for intelligence, to help determine accident causation and provide a clearer picture of the issues and potential intervention to improve the road safety condition. Besides encouraging the use of accident data in planning and decision making, the development of the new system is also aimed at providing a centralized accident database system for nationwide record management.

Tools for the system must be designed to give the easiest displaying mode for the most frequent information usually to be used by road safety researchers and practitioners to achieve effective road safety research and intervention planning (Mahmud et al., 1993). Outputs, either in the form of figures, graphs, maps or diagram sketches should be able to be retrieved without having to issue a long list of commands to the database system.

### *2.2 Sources of data*

The source of data was primarily based from road accident form POL27 used by RMP for all road accidents throughout the country. There are a total of 91 variables being collected for each accident case. Factors influencing accident encompassing the road environment, vehicle and road users are all collected using the form.

The POL27 form was designed for easy completion by simply having a relevant value for most of the accident parameter (Mahmud et al., 1993). Traffic officers filling out the form will simply need to circle the relevant value for each type of accident variable, which has been conveniently classified into various sections. In few cases where absolute figures are required such as age, road shoulder width or property damage cost, the officer will need to fill out the boxes with the exact figure. The accident data from the hardcopy form are then manually keyed-in into CARS.

Despite the abundance of accident data dated back to 1991, the accident data from the years 2005 to 2008 were gathered for the purpose of new application development. The resultant dataset contained a total of 995,313 accident records (22,425 fatalities and 105,728 injuries). Data were obtained from RMP headquarters in the format of text files. Nevertheless, the number of accident records totalled to about 86% from the official figure reported by RMP. The balance of the records was not available in electronic copy and was therefore not included in the dataset.

Received road accident data was then stored to create a comprehensive road accident database. The creation of a road accident database will enable various researches on road safety to be carried out (Ferrante et al., 1993). The data were structured into different types of tables to make the data more manageable and easy to retrieve.

One of the variables in POL27 form is location data. Location data includes route number, kilometre post, name of the road and the corresponding coordinate number. In addition, accident sketch diagrams from the hardcopy of POL27 were scanned and saved as electronic copy. Besides providing better understanding on the accident, a sketch diagram will also provide detail information on the location such as the final position of the vehicle on the road.

### *2.2.1 Information not incorporated into the database*

Data on injury for victims involved in accidents throughout the period of study were not incorporated into the database. At the time of the development, most of the hospitals do not maintain the patient's records in electronic media. Only a few hospitals maintain a Hospital Information System (HIS) for managing their patient's records. However, hospitals do not have a standard patient forms such as what the RMP have. Therefore, the captured data varies between hospitals. Restricting the hospital data would eliminate the problem of inconsistency of injury data format between hospitals. In addition, data linkage between police and hospital data would not be required, thus simplifying the process.

### *2.3 Application development*

In identifying and evaluating road safety, researchers and professionals are sometimes ill-equipped with the necessary specific information in relation to accident pattern and causation (Mahmud et al., 1998). As a result, some assessments are rather subjective and may be drawn from uncertainty due to lack of evidence. Hence, accident data is seemed to be the most appropriate source for evidence. To facilitate the use of accident data, an application that can analyze road accident data is needed to provide road safety intelligence.

Therefore, a new system will be designed to suit the custom and specific needs of road safety researchers and practitioners to perform road safety analysis. In addition, the system will also have the capability to manage road accident data. To facilitate the use, a user friendly graphical interface is essential to expedite analysis and planning. Capability to perform an in-depth analysis will be beneficial to road safety researchers and professionals in identifying the appropriate research or remedial work needed (Mahmud et al., 1998). Data may be presented and visualized in several ways thus making it easier for users to understand the accident information and statistics. This can be achieved by displaying the data in the form of statistical figures or graphs. In addition, accident locations should be presented through the use of built-in mapping function for better visualization.

## 3. RESULTS

Features incorporated in this section includes a critical examination of data quality; design of database based on available road accident data; the development of new road accident analysis application and an overview of intervention programs introduced resulting from the new system.

### *3.1 Data quality*

Checks for data inconsistencies within the road accident data were made where possible. Some of the variables that don't have predefined values, such as route number and coordinate number have become one of the problems for data inconsistencies. For location variables, inconsistencies in data entry will be affected when performing location ranking (Wu & Harbird, 1996). For example, route number 'F0001' will

be treated as a different route from 'F001' or 'F1'. Although they are actually referring to the same route, the database system will regard them as different since the database will read based on exact characters. For year 2008, 32% of the records needed to be corrected for its accident location. The corrections included identifying all varying coding and subsequently updating the data to the standard format.

Furthermore, variables with pre-defined values were also checked for their integrity. For example, the driver's injured body part were tabulated against accident severity and 2.7% of the records show as damage only accident while the driver part of body injured were filled in with the injury parts instead of none. Similarly, road surface conditions were tabulated against weather. The result shows 0.8% of the records were dry roads on rainy conditions. These are all the discrepancies that need to be amended and verified.

### *3.2 Database design & structure*

The accident database contains all road accident records throughout the country from year 2005 to 2008. Accident data were separated into their year of occurrence and are stored in a different database set. Each set will maintain accident records for one year. This will help to expedite the query process since the number of accidents records is not very large. Moreover, most of the road safety statistics produced is on yearly basis (Royal Malaysian Police, Traffic Division, Bukit Aman, 2007). MySQL was chosen to be the Database Management System (DBMS) for the accident data.

Table 1: Accident Variables List

| | | | | |
|---|---|---|---|---|
| General | General Accident Info | | | Location Info |
| | Report number | No pedestrian injured | Collision type | Kilometre post |
| | State | No pedestrian injured | Weather | Nearest KM post |
| | District code | Accident severity | Light condition | Map series |
| | Police station number | Road surface type | Road type | Map code |
| | Month | Road geometry | Location type | Latitude |
| | Hour | Quality of surface | Area type | Longitude |
| | Day of week | Road condition | Vehicle damage cost | Route no. |
| | No. vehicle involved | Lane marking | Property lost | Node 1 |
| | No. vehicle damaged | Hit run | Accident sketch | Node 2 |
| | No. drivers killed | Control type | Road defect | Direction of travel |
| | No. drivers injured | Road width | Speed limit | Section no |
| | No. passengers killed | Shoulder width | Road surface condition | Animal Info |
| | No passengers injured | Shoulder type | Traffic system | Animal fault |
| Vehicle / Driver | Vehicle Info | | Driver Info | |
| | Vehicle damage | Vehicle defect | Driver sex | Driver qualification |
| | Vehicle model | Vehicle modification | Driver age | Drinking drive |
| | Vehicle year | Length of skid marks | Driver race | Driver occupation |
| | Registration no. | Tyre burst | License status | License type |
| | Vehicle type | Foreign vehicle | Driver injury | Driver part of body injured |
| | Vehicle ownership | Vehicle movement | Driver belt wearing | Driver errors |
| Injury | Passenger Info | | Pedestrian Info | |
| | Passenger injury | Passenger race | Pedestrian action | Pedestrian age |
| | Passenger belt wearing | Vehicle Code | Pedestrian injury | Pedestrian part of body injured |
| | Passenger position | Passenger part of body injured | Pedestrian sex | School pupil |
| | Passenger gender | Passenger age | Pedestrian race | Pedestrian location |

Each of the database has three tables (Table 1), i.e. the general information table (with 51 headings); vehicle and driver information table (with 24 headings); and injury information (with 16 headings). Each table maintains the report number variable. The variable will act as the key to link between the tables in the database. Structure of the database is kept to the minimum number of tables to enable query to be executed easier.

Information retrieved from the database is often subject to general inquiry or specific inquiry. For general inquiry, the processing is straight forward, using a few variables at a time. However, for specific type of inquiry, information retrieved from the database will consist of two or more tables hence will require tables to be linked. The process of linking the tables in this database structure is done by using the report number. This database is structured to employ 'one-to-many' type of relationship. If a query made to one table and need information from the other tables, the report number value will be used to find the related records from the other tables.

Most of the variable options are stored in the database in numeric code instead of full text. This will greatly help the process of querying data as database process number faster than text. In addition, coding the variable option in numbers will also eliminate the problem of spelling error or non-standard data entry.

Data structuring is important as better structured data will aid in analyzing process thus providing better information presentation (Mahmud et al., 1998). Hence, the structure of the database is interesting whereby it naturally covers the three factors that influence accidents that are road user, vehicle and road environment. It is not therefore surprising that this structure is used all over the world to manage road accident records (Breton & Vervialle, 2005).

### *3.3 MIROS Road Accident Analysis and Database System (M-ROADS)*

A new application system called MIROS Road Accident Analysis and Database System (M-ROADS) has been developed to take advantage of the newly structured database. The objective of the system is to provide road safety stakeholders with accurate, continuous and comprehensive information on road traffic accidents. In addition, M-ROADS is also aimed to provide intelligence on road safety condition to aid researchers and decision makers in implementing interventions to improve road safety in the country.

The main task of M-ROADS is to facilitate the process of querying the accident database. Most of the relational process between the application and the database were done in the background and the users will not need to understand the Structured Query Language (SQL) that was used to query the database in order to retrieve accident data. A menu was created as an interface to make it as user friendly as possible and easy to use. By having the interface, it can protect the database by restricting faulty commands given by the user that might bring damage to the database (Mahmud et al., 1998).

Within M-ROADS, a few functions have been developed. Table 2 describes the summary of functions in M-ROADS.

Table 2: Features and Functions Build in M-ROADS

| Feature | Function |
|---|---|
| Cross-tabulation | Display two accident variables simultaneously, with set condition capability |
| Location ranking | Identify frequent accident locations. Locations can be ranked according to four different scopes which are district, route number, coordinate number and road name, with set condition capability |
| Set condition | Complement cross-tabulation analysis and location ranking by enabling the user to set their specific requirements based on all the 91 accident data variables. |
| Location mapping | Achieved via Google maps. User will have the ability to map the locations analysed from location ranking module |
| Record management | Fundamental record management such adding a new record, editing and deleting existing record as well as to view accident record |
| Database management | Add a new database or remove the existing database |
| User management | Add or remove user from the system as well as setting the user privileges |
| Data converter | Convert the accident data in the text file format received from both of RMP system either CARS or CAAS and upload them to the M-ROADS database server |
| Export | Export the result to Microsoft Excel or graph for reports preparation |
| Print | Print the displayed result |

The main function that is regularly used for road safety analysis is cross-tabulation. Cross-tabulation will display the result of two accident variables simultaneously with optional set condition capability. The set condition will enable flexibility in the data query as deeper analysis is possible to be carried out. As an example, a user may need to identify all entities in the database that meet his specifications such as to display all accidents involving motorcycle, not wearing helmet, after midnight and on rainy days. Having the ability to query with specific details will enable specific information to be retrieved thus enabling specific research or intervention to be carried out.

Displaying accident information in M-ROADS is done by translating the numeric figure that represents variable names in the database to the textual format. The process of translation is done by mapping each of the numeric value to its text and the text will be displayed to the user instead of the numeric value. The mapping index files are saved in a text file format and will be accessed each time when data retrieval is done. Each of the variables that are coded in numeric form maintains a single separate text file. Therefore, should there is a need to edit the variables; the user will be able to do it easily and instantaneously via the text files. The flexibility in editing is important as over the time the number of variable's options may increase thus this will require modification to the text files.

### *3.4 Research and intervention introduced*

M-ROADS has helped to support road safety researchers and policy makers in planning for new intervention and programmes in Malaysia. For example, in year 2005, Road Safety Education (RSE) programme was initiated in Malaysia to reduce fatality and accident among school children. Accident data were used to select the districts with high fatalities involving school children and these districts were given the priority for initial RSE implementation to obtain significant benefits and impact.

In addition, M-ROADS was also used in the Global Road Safety Partnership (GRSP) workshop for RMP and Malaysian Road Transport Department in aiding them to carry out specific enforcement strategy based on accident data as the evidence. This includes the identification of problems, location, time and day for enforcement.

For the Community Based Programme - Helmet Initiative, the system identified areas in each state with high fatalities due to head injury among motorcyclists. Helmet wearing campaign and free helmets were distributed to encourage people in that particular area to wear helmet. The performance of the programme was evaluated in terms of helmet wearing compliance and head injury after the campaign period and has shown positive result (Nusayba et al., 2007)

In addition, accident data can be used to monitor performance of the road safety programmes and interventions. One of the main indicators for the effectiveness of the programmes or interventions is via the number of road accidents reduced. M-ROADS can help to evaluate the programmes and interventions by examining the accident trend after the implementation.

Currently at MIROS, M-ROADS acts as an important tool in determining area to focus in order to give significant impact on road safety. M-ROADS provides scientific evidence in assisting the government to introduce new policy at the national level. The system is able to produce a clearer picture of the current situation pertaining to any arising issues related to road safety. The system is also capable to provide immediate information in order to answer important queries from any interested parties even including ministers and other road authorities.

## 4. DISCUSSION

The results of the development indicate that road accident data are able to provide and serve as the source of intelligence for road safety in the country. The development of M-ROADS has also resolved the issue of having a comprehensive road accident database and analysis system whereby the data source used by M-ROADS are compiled and combined from all the districts by converting road accident data received from each of the individual district. Nevertheless, issues relating to data quality will still exist if M-ROADS is not actively and extensively used for nationwide data entry.

Functionality in M-ROADS such as the enhanced data tabulation will help the road safety authorities to develop strategies in reducing the severity of accidents and ultimately reduce the number of accidents and

fatalities. Focused or targeted programme can be planned to give effective treatment to specific road safety issues. Therefore, with the new system, accident data will be fully utilised and will not be used for record keeping only. Furthermore, having an online system will help to solve the delay in data entry thus enabling fast access to the data. In addition, with M-ROADS, time taken for routine manual query are shortened and the users either law enforcement, researcher or policy maker can use the time saved to concentrate on other important part of their analysis.

Since M-ROADS was developed in-house, it is possible for its functionalities to be further expanded in the future. For example, MIROS has currently embarked on an in-depth road accident investigation to investigate the causation of high profile accident cases throughout the country. The data collected will be more extensive compared to the accident data collected in POL27. As the data accumulates, data from the in-depth investigation can later be combined with the M-ROADS database. Correlation between both data can be made and an advance study can be carried out. In addition, injury data are currently being sought from the hospitals for road accident patients. Injury data will greatly complement the accident data in the database by enabling better understanding of the post crash impact.



## 5. CONCLUSION

The result of the new development of road accident analysis and database system in Malaysia has shown that accident data can be used as the primary source for road safety authorities and professionals to understand the accident situation and its causation. Focus or targeted programmes can be planned to give effective treatment to specific road safety issues. Hence, accident data will be fully utilised and will no longer be treated for record keeping only.

In addition, the road accident data can be used as evidence in evaluating the progress of road safety intervention as well as for performance monitoring. By having a well established and comprehensive set of road accident database system, it will aid in the formulation of effective road safety research and interventions thus providing a more promising result in reducing both road accidents and fatalities.

## References

Breton, P. L., & Vervialle, F. (2005). Multivariate Analysis Applied to the French Accidents Database as A Multilevel Accidents Register. Road Safety on Four Continents, 5-7 October 2005, Warsaw, Poland .

Ferrante, A. M., Rosman, D. L., & Knuiman, M. W. (1993). The Construction of A Road Injury Database. Accident Analysis and Prevention , 659-665.

Lyons, R. A., Ward, H., Brunt, H., Macey, S., Thoreau, R., Bodger, O. G., et al. (2008). Using Multiple Datasets to Understand Trends in Serious Road Traffic Casualties. Accident Analysis and Prevention , 1406-1410.

Mahmud, A. R., Radin Sohadi, R. U., & Mansor, S. (1998, May 15). A GIS Support System for Road Safety Analysis and Management. Pertanika J. Sci. & Technology , pp. 81-93.

Mahmud, A. R., Radin Sohadi, R. U., & Salan, S. (1993, August). Persembahan Maklumat Kemalangan Dengan Sistem Maklumat Geografi. The Surveyer, The Professional Journal of Surveyors Malaysia , 34-38.

Megat Johari, N., Mohd Sirat, S. R., Subramaniam, S. N., Yellappan, K., Singh, S., & Radin Sohadi, R. U. (2007). Community Based Safety Helmet Program in Kuang. Kuala Lumpur: MIROS.

Ministry of Transport. (2006). Road Safety Plan of Malaysia 2006 - 2010. Kuala Lumpur: Road Safety Department.

Royal Malaysian Police, Traffic Division, Bukit Aman. (2007). Annual Road Accident Statistic. Kuala Lumpur.

T.H., L., & Radin Sohadi, R. U. (n.d.). The Development of Road Accident Database Management System for Road Safety Analyses and Improvement. Retrieved February 15, 2008, from Department of Works Website: http://rakan1.jkr.gov.my/csfj/editor/files/File/Dokumentasi/Nota%20Kursus%20dan%20Seminar/Malaysia%20Road%20Conference/dev_road.pdf

Wu, J., & Harbird, L. (1996). A functional database system for road accident analysis. Advances in Engineering Software , 29-43.