Experience report

Data Integration Process – An information management framework for multiple road crash databases in Brazil

doi: 10.5123/S1679-49742018000200018

Lisandro Iusry Abulatif[1] – orcid.org/0000-0002-4530-2817
Technical group of partners of the Life in the Traffic Project

[1]Global Road Safety Partnership, Geneva, Switzerland

Abstract

Objective: to describe the management process for integrating multiple road crashes databases applied in five state capital cities where the Life in the Traffic Project was piloted. **Methodology**: This is a descriptive study which presents the process of integrating road crashes data from multiple institutions for the year 2012, including linkage with data from Mortality Information System (SIM) and National Hospital Information System (SIH). **Results**: Integration of multiple institution data, creation of a single list of road crash victims and reclassification of accidents as being serious or fatal. There was an average increase of 87% in relation to the total number of victims counted before the use of this process. **Conclusion**: The results suggest that Data Integration Process provides good results and can be adopted by other cities that need reliable information about road crashes to inform the planning, monitoring and evaluation of road safety actions.

Keywords: Information Management; Information Systems; Data Sources; Accidents, Traffic; Mortality Registries.

Correspondence:
Lisandro Iusry Abulatif – Av. Assis Brasil, No. 289, Florida, Guaíba, RS, Brasil, CEP: 92500-000
E-mail: labulatif@gmail.com

Introduction

The consequences of road crashes worldwide go beyond transit and transport fields and are considered to be an global health issue. In 2015, approximately 1.25 million people died globally as a result of road crashes. In middle- and low-income countries, the cost of accidents involving victims exceeded US$100 billion per year.[1]

In 2010, Road Safety in Ten Countries (RS-10), an international consortium for road safety, involving institutions such as the World Health Organization (WHO), Global Road Safety Partnership, Association for Safe International Road Travel, Embarq and Johns Hopkins University, received funding from Bloomberg Philanthropies to perform activities intended to improve road safety in the ten most critical coutriescountries in terms of traffic deaths, namely: Brazil, Cambodia, China, Egypt, India, Kenya, Mexico, Russia, Turkey and Vietnam.[2] In 2010, the death rate in Brazil related to road crashes was 22.5 deaths per 100,000 inhabitants.[3]

> *Data Integration Process is a management model, developed with the aim of systematizing the integration of multiple data sources on road crashes in order to build road safety indicators.*

A National Interministerial Commission was created in Brazil in 2010 to coordinate the implementation and development of the RS-10 Program, which came to be called the Life in the Traffic Project (Projeto Vida no Trânsito - PVT).[4] Five pilot state capital cities were chosen to begin PVT implementation: Belo Horizonte/MG, Campo Grande/MS, Curitiba/PR, Palmas/TO and Teresina/PI.[5,6]

The Data Integration Process (DIP) is an information management framework for the integration of multiple data sources, developed in order to support the retrieval and integration of multisectoral data on road crashes, as well as the management of these activities, based on the continuous improvement approach - data produced by means of the process are used to support road safety action planning and execution in cities implementing PVT.[7]

DIP was developed by the author of this article. It was initially implemented in Brazilian city of Guaíba/RS, using the first version of this methodology to integrate data on road safety from multiple sources.[8] In order to implement DIP, five state capital cities received training and support in the methodology through training workshops and follow-up visits on a quarterly basis. This approach can be considered to be a monitoring process, similar to that proposed by Hsieh and Hsu.[9]

The objective of this study was to describe DIP as a new and replicable process,[10] its main stages and its implementation, as well as to present its results.

Methods

The process was carried out in the municipalities to which this study refers and is based on the description of the DIP implementation methodology and the description of its stages, followed by the presentation of its results. At the DIP implementation stage, local consensus had already been reached on the need to build an integrated database on severe and fatal road crash victims.

The definition of deaths and serious victims was established based on the World Health Organization standard, which considers a "fatal victim" to be a person killed immediately or within 30 days, as a result of a road crash, and a "seriously injured" to be someone hospitalized for at least 24 hours due to injuries sustained in a road crash.[11]

DIP is a management model, developed with the aim of systematizing the integration of multiple data sources on road crashes in order to build road safety indicators. This approach provides reliable and comprehensive data, thus allowing better planning and intervention processes.[12]

DIP is a cyclical process of continuous improvement, the main stages of which are: (I) People Integration, (ii) Situational Awareness, (iii) Workflow Definition and Implementation, (iv) Results Delivery, and (v) Adjustments and Improvements.

Stage 1 – People Integration

People integration is the first stage, because, in terms of road safety, it is expected that a variety of agencies will have data on road crashes.[13,14,15] The first step is to establish a Data Management Team (DMT), comprised of professionals who work with data collected by institutions which deal with road crashes or their victims. DMT are basically comprised of traffic/transport and health practitioners.[14,15]

This team structure also aims to increase the trustiness and cooperation between institutions that collect road

crashes data in order to inform the planning of integrated road safety actions.[16,17] The following variables were used to characterize the DMT: organization to which the team member belongs; institution name; member's position/function; member's qualifications; if the member already works with an information system; municipality name; and if the member has experience handling any information system.

Step 2 - Situational awareness

Situational awareness is performed in Step 2 to provide a better understanding of each city's road crash information systems. This provides a clear vision of the possibilities, obstacles and possible misconceptions regarding information systems that will be used by DMT.[11,18,19]

DMT identifies all the institutions that have data on road crashes (Table 1). The information gathered is: organization name; whether it provides services at the scene of the accident; whether it uses physical or electronic forms; whether it stores data in a physical or electronic format; whether it provides data from its database; whether a DMT member has access to the database; the nature of the data (whether it is data on the scene of the accident or hospital/mortality data); whether it is considered critical for the monitoring of

Table 1 – Identified sources of road crash data: composition by area

City/FU	Health area	Traffic/Transport Area	Public Safety Area
Belo Horizonte/MG	• SAMU[a] • Mortality Information System (SIM)[b] • Hospital Information System (SIH)[b] • Information System for Notifiable Diseases (SINAN) • João XXIII Hospital • Risoleta Tancredo Neves Hospital • Odilon Behrens Hospital • Local Media	• Belo Horizonte Transport and Traffic Company[a] • BHTrans - Inspection	• Military Police • Civil Police • Fire Department • State Highway Police • Federal Highway Police • Institute of Forensic Medicine
Campo Grande/MS	• SAMU[a] • SIM[b] • SIH[b] • Santa Casa Hospital • University Hospital • Regional Hospital	• Detran-MS • Municipal Agency Transport and Transit[a]	• Military Police - Ciptran • Civil Police • Fire Department • Criminalistics Institute • Institute of Legal Medicine and Dentistry
Curitiba/PR	• SAMU[a] • SIM[b] • SIH[b] • Sinan • Cajuru Hospital • Evangelical Hospital • Worker's Hospital	• Highway Concessionaires	• Transit battalion of the Military Police[a] • Integrated Service of Trauma Care in Emergency[a] • Federal Highway Police[a] • State Highway Police • Traffic Police • Institute of Forensic Medicine • Criminalistics Institute
Palmas/TO	• SAMU[a] • SIM[b] • SIH[b]	• Transit Agency, Transport and Mobility[a]	• Military Police[a] • Civil Police • Fire Department
Teresina/PI	• SAMU[a] • SIM[b] • SIH[b]	• Municipal Superintendence Transport and transit	• Independent Company of Transit (Military Police)[a] • Fire Department • State Highway Police • Federal Highway Police[a] • Police Station for accidents • Criminalistics Institute • Institute of Forensic Medicine

[a]Source in electronic format used to build Single List of Victims.
[b]Source in electronic format for classification of the victims as severe or fatal injury on Single List of Victims.
Note: Other data sources used as complementary sources during the methodological process implementation period.
Source: Author's elaboration.

accidents (meaning that the team cannot continue with its work without access to this database); and whether it is already in use by DMT.

Workflow mapping was performed for each institution, in order to identify the physical and electronic documents created (such as accident statements and alcohol testing reports) and to characterize how and where the documents were stored, as well as the database used for this purpose. At this point, DMT identified where and how the data could be obtained.

Stage 3 – Workflow definition and implementation

At this stage the variables from each database that should be shared are defined, in order to build an integrated database of road crash victims. Only electronic databases were used containing data on victim identification and location of the accident, such as the victim's name and age/date of birth, address and date of occurrence of the accident. A standardization phase was then begun to identify common variables in the databases and standardize them with regard to same type of data, field size and format codes and values. The databases were then combined, thus creating an integrated database of road crash victims by using records that had information about the location of the accident (marked as on-site records). This procedure was performed on a quarterly basis. Each database has its primary keys preserved. When multiple records (rows) related to the same victim were identified, the row containing the fullest name of the victim was maintained and received primary key values (referring to the data base to which it belonged) of the rows to be removed, so that only one record of the victim remained, now containing the primary key of each database holding a record of the victim in question (Table 2).

This procedure enabled the removal of duplicate records related to the same victim of a particular traffic incident, whilst also preserving all sources of identification data, thus creating the Integrated Casualties List (ICL). ICL structure is shown in Table 2. The variables shown in this Table are considered to be the minimum variables for preparing the List. Each DMT possessed autonomy to add other variables that best met the needs of its work.

The records of victims on ICL were joined with each municipality's hospitalization database, with the aim of identifying victims who were hospitalized for at least 24 hours, in order to classify them as "seriously injured". In order to classify "fatal victims", ICL data were cross-referenced with the death certificate database. Hospitalization data were obtained from the Hospital Information System of the Brazilian Unified Health System (SIH/SUS). Death certificate data were obtained from the Mortality Information System (SIM). Variables held in common on ICL, SIH/SUS and SIM, such as name, age (date of birth, when available), date of occurrence and sex were mapped. However, none of the variables in common had "primary key" characteristics (either alone or combined) which would allow deterministic relationship (comparison between two or more data tables using variables with the same values in both data tables). The non-deterministic relationship (comparison of similarity between two or more data tables using variables that should have the same values in both data tables) was the method adopted by the Brazilian Ministry of Health of Brazil to cross-reference the databases.

The status of "seriously injured" was assigned to victims on ICL who had been hospitalized for more than 24 hours.[11] The status of "fatal victim" was assigned to

Table 2 – Structure of the Integrated Casualties List of road crashes

Id. ICL[a]	Id. source1[b]	Id. source2[c]	Date of occurrence	Victim's name	Birthdate	Age
S00001	DS1S0001		02/02/2011	Victim 1	01/01/1977	34
S00002	DS1S0002	DS2S0001	02/02/2011	Victim 2	28/01/1968	43
S00003		DS2S0002	02/03/2011	Victim 3	02/01/1935	76
S00004	DS1S0003	DS2S0003	02/04/2011	Victim 4		45
S00005		DS2S0004	02/05/2011	Victim 5	04/01/1930	80
S00006		DS2S0005	02/06/2011	Victim 6		30

a) ICL sequential numeric identifier.
b) Scene of accident data source identification code (e.g.: Military Police).
c) Scene of accident second data source identification code (e.g.: SAMU).

victims with a corresponding record on the mortality database when date of death was up to 30 days after the date of the accident;[11] victims who died after 30 days were classified as "seriously injured". The status of "slight injured" was attributed to all other victims.

Variables of matched records (ICL records also found on the SIH/SUS or SIM databases), considered as the "primary key" on SIH/SUS (hospital admission authorization document number), and the "primary key" on SIM (death certificate number), were added to the ICL variables.

At this point, ICL was updated with the new variables obtained from the SIH/SUS and SIM databases, thus enabling the identification of severe and fatal victims as per the WHO standard.

Stage 4 – Results delivery

Results delivery takes place in two main ways. Firstly, the variables are selected and organized in one or more tables to enable accident data tabulation for different purposes, such as identification of the place where road crashes occurred, time of day, day of week etc.[11,22] Secondly, the data are delivered for use by the accident analysis team in identifying risk factors and behaviors, health service users who have contributed to the occurrence of accidents and groups of victims.[7]

The data produced are made available to DMT on a quarterly basis, a well as being shared with the institutions that provided data for DIP. Given that there are data that identify victims, the information produced can be manipulated by technicians of each DMT, so that when external requests for access to data are met they do not contain the elements that identify victims.

Stage 5 - Adjustments and Improvements

Adjustments and improvements are made to data structure and usage, whenever the need or opportunity to do so is perceived. This procedure leads back to the first and second stages of DIP, restarting the process as a cycle of systematic continuous improvement, which leads to an active strategic review of the management of road crash information.[23]

Results

In the People Integration stage, multisectoral teams were established in each capital city, thus creating the municipal DMT, as shown in Table 3. The number of information sources identified each city is as follows: Belo Horizonte/MG used 16 different data sources; Campo Grande/MS 13; Curitiba/PR, 15; Palmas/TO, 7; and Teresina/PI, 11. Table 2 details the sources of information of each municipality, per area, and how they were used.

Of the data sources identified, 53.2% came from institutions that provide services at the scene of the accident. The average number of data sources for each city was 12. Of the sources considered critical for accident monitoring, 94.4% were effectively used (total of sources in use divided by the total number of sources considered critical).

The data of 90.0% of information sources are stored in electronic format. In Campo Grande/MS and Palmas/TO, all sources of information on accidents stored their records in an electronic format. Curitiba/PR had 12 of its databases in an electronic format, Teresina/PI had 4 and Belo Horizonte/MG had 10.

Throughout the entire Workflow Definition and Implementation stage, during which multiple on-site data sources and the ICL production methodology were combined, Curitiba/PR, Palmas/TO and Teresina/PI achieved increases of 3,054, 756 and 5,054 records, respectively, in the total number of identified victims on local roads. Campo Grande/MS and Belo Horizonte/MG cities decided to work only with traffic/transport data.

Table 3 – Composition of Data Management Teams regarding the number of professionals in each area

City/Federation Unit	Health area	Traffic/Transport Area	Public Safety Area
Belo Horizonte/MG	3	3	1
Campo Grande/MS	5	3	-
Curitiba/PR	4	3	5
Palmas/TO	3	4	3
Teresina/PI	5	-	-

Using the non-deterministic relationship methodology, the cities were able to classify their road crashes victims as "fatal", "severe" or "mild" victims, in accordance with WHO standards.[11] Table 4 describes the increase in the number of fatal victims identified through DIP.

Using the Adjustments and Improvements stage, the Teresina/PI DMT created an electronic database to record information on cases dealt with by its Mobile Urgent Care Service (SAMU), which is now used to record care provided to road crash victims. This database has been in use since 2011.

Discussion

DIP implementation produced relevant results: the potential for aligning of multisectoral data management teams; integration of road crash databases obtained from multiple institutions; creation and development of ICL, with a mean increase of 87.5% in the total of road crash records in the capital cities that used the process and which, without this process, would not have been identified; reclassification of severe and fatal accidents in accordance with WHO standards,[4] with an average increase of 124.3% in total number of fatal victims identified in relation to local data on deaths; and continuous improvement of road crash data integration and management. Such benefits are in line with the WHO recommendation on reliable road crash data, providing a sound basis for road safety action planning and decision-making.[24]

It is noteworthy that although it is not explicit in the results, throughout the DIP implementation process it was perceived , that one of the major challenges for the sustainability of qualified ICL data production classified into severe and fatal accidents is political support on the part of local service managers. This is a key element, since the turnover of DMT members and difficulty in accessing databases are situations in which service managers play a critical role in solving obstacles. This also applies to the interlocutions between the municipal, state and federal levels, with regard to providing access to databases, as is the case of obtaining data from military police databases. These are managed at state level, and access to them depends on articulation between municipal and state authorities in order to access to maintained in a continuous manner.

The results presented suggest that DIP can be expanded and implemented in Brazil and internationally as a reliable model for road safety data management, contributing to public policies intended to reduce road crash morbidity and mortality, as well as to the definition of new practices of information recording and sharing in the areas of health, traffic/transport and public security. DIP's flexibility is currently being tested in other Brazilian cities and the results will be presented in a future article.

Acknowledgment

Technical group of partners of the Life in the Traffic Project for their contribution to Data Integration Process implementation activities.

Authors' contributions

Abulatif LI developed, drafted and reviewed the article. The author has approved the final version of the manuscript.

Table 4 – Number of road crash victims before and after the Data Integration Process (DIP) and increase after DIP, in pilot capitals of Life In the Traffic Project (PVT), 2011a and 2013

City/FU	Before DIP		After DIP					
	Total victims	Fatal victims	Total victims	Fatal victims (30 days)	Seriously injured (24 hours)	Slight injured	Increase in total number of victims[b]	Increase in total number of fatal victims
Curitiba/PR[a]	3,767	66	6,821	163	749	5,909	81.1%	147.0%
Belo Horizonte/MG	17,689	170	17,689	231	1,729	15,389	-	35.9%
Palmas/TO	1,817	29	2,573	42	329	2,022	41.6%	44.8%
Campo Grande/MS	8,284	56	8,284	116	1,065	7,103	-	107.1%
Teresina/PI	1,617	38	6,671	147	7,634	816	312.5%	286.8%

Source: Belo Horizonte/MG, Campo Grande/MS, Curitiba/PR, Palmas/TO and Teresina/PI Data Management Teams.
[a]Data related to 1st semester 2011.
[b]Belo Horizonte/MG and Campo Grande/MS did not add SAMU data.

References

1. World Health Organization. Global status report on road safety 2015 [Internet]. Geneva: World Health Organization; 2015 [cited 2018 Mar 5]. 323 p. Available in: http://apps.who.int/iris/bitstream/10665/189242/1/9789241565066_eng.pdf?ua=1
2. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde. Saúde Brasil 2011: Uma análise da situação de saúde e a vigilância da saúde da mulher [Internet]. Brasília: Ministério da Saúde; 2012 [citado 2018 mar 5]. 444 p. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/saude_brasil_2011.pdf
3. United Nations. Improving global road safety [Internet]. General assembly. Session 64/255. Geneva: United Nations; 2010 [cited 2018 Mar 5]. Available in: http://www.who.int/violence_injury_prevention/publications/road_traffic/UN_GA_resolution-54-255-en.pdf?ua=1
4. World Health Organization. Road safety in ten countries [Internet]. Geneva: World Health Organization; 2010 [citado 2013 May 30]. 1 p. Available in: http://www.who.int/violence_injury_prevention/road_traffic/countrywork/rs_overview.pdf
5. Brasil. Ministério da Saúde. Portaria Interministerial nº 2.268 de 10 de agosto de 2010. Institui a Comissão Nacional Interministerial para acompanhamento da implantação e implementação do Projeto Vida no Trânsito. Diário Oficial da República Federativa do Brasil, Brasília (DF), 2010 ago 12; Seção 3:154.
6. Silva MMA, Morais Neto OL, Lima CM, Malta DC, Silva Jr. JB. Projeto vida no trânsito – 2010 a 2012: uma contribuição para a década de ações para a segurança no trânsito 2011-2020 no Brasil. Epidemiol Serv Saúde. 2013 jul-set;22(3):531-6.
7. Cardita J, Di Pietro G. Proactive partnership strategy: a community participation model to addres road safety [Internet]. Geneva: Global Road Safety Partnership; 2010 [cited 2018 Mar 5]. 64 p. Available in: https://www.grsproadsafety.org/wp-content/uploads/GRSP-PPS-manual.pdf
8. Johannsen CG. Strategic issues in quality management: I. Theoretical considerations. J Inf Sci. 1996 Jun;22(3):155-64.
9. Hsieh ST, Hsu PY. Mentoring effects in the successful adaptation of information systems. Inf Dev. 2013 Dec;31(2):164-75.
10. Zook C, Allen J. O poder dos modelos replicáveis: a construção de negócios duradouros em um mundo em constante transformação. Rio de Janeiro: Elsevier; 2012.
11. World Health Organization. Data systems: a road safety manual for decision-makers and practitioners [Internet]. Geneva: World Health Organization; 2010 [cited 2018 Mar 5]. 146 p. Available in: http://apps.who.int/iris/bitstream/10665/44256/1/9789241598965_eng.pdf
12. Organisation for Economic Co-Operation and Development. International Transport Forum. Towards zero: ambitious road safety targets and the safe system approach [Internet]. Paris: OECD Publishing; 2008 [cited 2018 Mar 5]. 241 p. Available in: https://fevr.org/wp-content/uploads/2017/12/Towards-Zero-OECD-PDF-57-MB.pdf
13. Organisation for Economic Co-Operation and Development. International Transport Forum. Road safety annual report 2016 [Internet]. Paris: OECD Publishing; 2017.[cited 2017 May 12]. 584 p. Available in: http://www.oecd-ilibrary.org/transport/road-safety-annual-report_23124571
14. Pedrosa AA, Costa EM, Passos GP, Quirino GG, Sousa PG, Silva RS. Relatório do projeto vida no trânsito: 1º trimestre de 2014 [Internet]. Teresina: Prefeitura Municipal de Teresina; 2014 [citado 2018 mar 5]. 4 p. Disponível em: http://portalpmt.teresina.pi.gov.br/admin/upload/documentos/9116b82fc7.pdf
15. Bellavinha J. Projeto vida no trânsito de Belo Horizonte. In: 19º Congresso Brasileiro de Transporte e Trânsito; 2013. Brasília: Agência Nacional de Transportes Públicos; 2013.
16. Parmigiani A, Santos MR. Clearing a path through the forest: a meta-review of interorganizational relationships. J Manag. 2011 Jun;37(4):1108-36.
17. Mathieu JE, Tannenbaum SI, Donsbach JS, Alliger GM. A review and integration of team composition models: moving toward a dynamic and temporal framework. J Manag. 2014 Jan;40(1):130-60.
18. Flett A. Information management possible?: Why is information management so difficult? Bus Inf Rev. 2011 Aug;28(2):92-100.
19. Costa EAD. Gestão estratégica: da empresa que temos para a empresa que queremos. 2. ed. São Paulo: Saraiva; 2007.
20. Schnell R, Bachteler T , Reiher J. Privacy-preserving record linkage using Bloom filters. BMC Med Inform Decis Mak. 2009 Aug;25(9) 41-52.
21. Camargo KR, Coeli CM. Reclink: aplicativo para relacionamento de bases de dados, implementando

o método probabilistic record linkage. Cad Saúde Pública. 2000 abr-jun:16(2);439-47.

22. Turban E, Sharda R, Aronson JE, King D. Business intelligence: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman; 2009.

23. Kerr KA, Norris T, Stockdale R. The strategic management of data quality in health care. Health Informatics J. 2008 Dec;14(4):259-66.

24. World Health Organization. World report on road traffic injury prevention [Internet]. Geneva: World Health Organization; 2014 [cited 2018 Mar 5]. 217 p. Available in: http://apps.who.int/iris/bitstream/10665/42871/1/9241562609.pdf

25. International Transport Forum. Reporting on serious road traffic casualties: combining and using different data sources to improve understanding of non-fatal road traffica crashes [Internet]. Paris: International Transport Forum; 2013 [cited 2018 Mar 5]. 106 p. Available in: https://www.itf-oecd.org/sites/default/files/docs/road-casualties-web.pdf

26. Randall SM, Ferrante AM, Boyd JH, Semmens JB. The effect of data cleaning on record linkage quality. BMC Med Inform Decis Mak. 2013 Jun;13:64-73.

Received on 04/05/2017
Approved on 27/12/2017