CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

Cross-Analysis of Hazmat Road Accidents Using Multiple Databases

Martin Trépanier
Marie-Hélène Leroux
Nathalie de Marcellis-Warin

November 2007

CIRRELT-2007-47

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
Pavillon Palasis-Prince, local 2642
Québec (Québec)
Canada G1K 7P4
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Université Laval · UQÀM · HEC Montréal · École Polytechnique Montréal · Université de Montréal

Cross-Analysis of Hazmat Road Accidents Using Multiple Databases

Martin Trépanier[1,2,*], Marie-Hélène Leroux[2], Nathalie de Marcellis-Warin[1,2]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7

[2] Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, C.P. 6079, succursale Centre-ville, Montréal, Canada H3C 3A7

Abstract. The usefulness of accident database is widely recognized in the literature. Hazardous material (hazmat) transportation risk assessment can hardly be done without them. It seems however that official accident databases alone are not providing enough information on all circumstances on accidents. In Canada, for example, data on workers and road conditions are not provided. Fortunately, this information could be available in other non-hazmat accident databases. This paper presents a methodology and a tool that were developed to integrate many data sources to analyze hazmat accidents in the province of Quebec, Canada. Databases on dangerous goods accidents, road accidents and work accidents were cross-analyzed to retrieve multiple informations on the same events. Results show that accidents are hardly matchable between databases (only 3 out of 140 hazmat accidents were formally identified), and that some major accidents are not reported in one or another database.

Keywords. Road accident, hazmat, accident database, spatial analysis, cross-analysis.

Acknowledgements. The authors would like to acknowledge the contribution of the CIRANO research center's partners of the project, especially those who provided the databases for this study: Transport Canada, the Quebec Ministry of Transport, and the Institut de recherche Robert-Sauvé en santé et sécurité du travail (IRSST). Special thanks to the Institut National de l'Environnement Industriel et des Risques (INERIS) from France, that initiated this study through the GLOBAL project.

---

* Corresponding author: Martin.Trepanier@cirrelt.ca

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 2007

1. Introduction

Risk assessment of hazmat road transportation can hardly be done without having sufficient information on accidents that occurred during road transportation activities. Several researchers report the lack of valuable data in order to conduct appropriate research work, especially regarding the circumstances of the accidents. Several databases need to be integrated in order to obtain a good portrait of the situation: using specific hazmat databases, like accident databases, is simply not enough. Other sources of circumstantial data must be examined (general transportation data, road network, weather and census data, workers insurance data, etc.).

In this paper, we present the methodology and the tool we developed in order to integrate multiple sources of data to analyze hazmat road accidents in the province of Quebec, Canada. The focus is mainly on road accidents as currently, more independent sources of data are available for that type of hazmat event.

The first part of the document briefly reviews the literature relevant to road accidents analysis. Most of the literature presented involves civil security aspects. Then, the description of the Object-Oriented approach used in this study is provided. Next, the paper describes the four-steps methodology used: data gathering, data formatting, the development of the Hazmat Event Cross-Observer Tool (HECOT) and results analysis. This methodology was applied to hazmat road transportation accidents in the province of Quebec (Canada). We compiled several statistics from the three accident databases used (Canadian dangerous goods transportation accidents database, Quebec road accident database and Quebec work accidents database) in addition to other sources of circumstantial data.

2. Literature review

Data analysis is an important part of most risk assessment studies dealing with road accidents. It is a time consuming task, but it can greatly improve result accuracy and bring a better understanding of the situation, especially if more than one source of data is considered. However, researchers are often limited in their attempts as data availability depends greatly on previous efforts made.

In this section, we present a few studies highlighting this complex situation. First, we present three studies dealing with road accidents in general. Then, we present studies dedicated to the hazardous material transportation field.

2.1. Data collection and road accidents analysis

Stevens and Minton (2001) studied the role of in-vehicle distraction (wireless phones, entertainment or navigation systems) in road accidents. To do so, they examined the recently built England and Wales database which stores information on fatal road accidents. Data come from police reports and include possible contributory factors. Although they found that these factors have an influence in 2% of the cases, their study shows that relying on a single source of information can be tricky as police reports do not necessarily include all the information needed.

Laberge-Nadeau & al. (2003) conducted a similar study. They tried to assess the risk of using a cell-phone while driving. However, instead of using a single source of information, they used data coming from police accidents reports but also from a questionnaire sent to a significant sample of drivers, and data from wireless telephones companies' cell-phone activity files. Linking all of these data together has allowed them to have a more in-depth understanding of the situation.

Lee and Mannering's study (2002) dealt with a different kind of problem. They tried to determine the role of roadside features in road accidents by linking a roadside features database with an accident database and roadway geometric data. However, collecting roadside features being a costly operation, they could only obtain data concerning a 96,6 km section of highway.

2.2. Data collection and hazmat road accidents analysis

As stated above, researchers are often limited because data availability depends of previous efforts made. This is particularly true in the hazardous material transportation field. In some countries, extensive databases have been built and maintained, but it is not always the case. Ohtani and Kobayashi (2005) tried to understand the causes of an increased number of hazardous materials accidents in Japan. Few studies had been made on the subject in that country for a simple reason: the non-uniformity of accident reports. They had to rearrange the database and categorize the accidents in order to be able to conduct statistical analysis of the situation.

Shorten et al. (2002) were luckier. United States laws triggered the creation of a national database of hazardous materials releases. Furthermore, Chester County in

Pennsylvania kept detailed records of every local release. In their study, the authors could easily analyze data coming from Chester County and identify hazardous chemicals that were the most often involved in accidents. They even compared their results with data coming from other databases and stated that the situation they studied was similar to what was experienced elsewhere.

In Italy, a few studies have analyzed the impacts of hazardous substances on a given territory by linking information coming from various sources. These studies have led to the development of software programs facilitating transport risk analysis and helping managers' decision making. Among them, let's mention Spadoni & al. (2000) who introduced the ARIPAR-GIS software and Milazzo & al. (2002) who introduced the TRAT2 software. However, as showed in the latter, data collection for such means can be long and complex: four types of data coming from seven different sources had been used.

Finally, as shown in Dobbins and Abkowitz (2003), collecting data, combining data sets and building easy-to-use software tools can serve other means. By linking all of the databases and information already available, these researchers were able to build a centralized database helping first responders in case of emergency. The authors were able to demonstrate that their tool could significantly improve decision making in these situations.

Various authors have noticed the lack of statistical reliability of hazmat accident databases due to the systematic under-reporting of such events (Erkut and Verter, 1995). Furthermore, in an early study, Hobeika and Kim (1993) have attempted to cross-examine various accident databases covering Pennsylvania's territory and have obtained poor matches between databases.

2.3 Transportation object-oriented approach for database integration

The approach used in this study is inspired by recent works related to the integration of data from several sources in the case of public transportation and drinking water distribution systems.

Every five years, there is a large telephone survey in the Greater Montreal area. The survey is conducted with a 5% sample rate. A typical survey collects data on about 70,000 households to obtain a better knowledge of the transportation behaviours of families and individuals (trip origin and destination, modes, purposes, car ownership, etc.). Chapleau et al. (1997) have developed a method based on a Totally Disaggregate Approach (TDA) to disseminate and analyze data. Over the years, the TDA lead to a Transportation Object-Oriented Modeling approach (TOOM) allowing the individual analysis of survey data in relation with other transportation sources (road network, public transportation network, census, etc.). In this approach, every system is structured around four metaclasses of objects that are used for data structuring and software tool development (Trépanier et al. 2001):

- *Dynamic objects* are elements which move within the system and are at the heart of transportation activities (ex: cars, trucks, boats, planes, goods).
- *Kinetic objects* that describe the movement itself (ex: goods itinerary, trip chains and individual trips).
- *Static objects* refer to all the supporting elements of the transportation system that do not move (ex: depot, warehouse, trip generator).

– *System objects* are groups of objects embedded in the general transportation system (ex: road network, logistics network).

The approach has been successfully applied to the detection and analysis of low quality events in drinking water distribution systems in five cities in North America and Europe (Besner et al. 2005). This project involved the integration of several databases from different sources, similar to this present study.

3. Methodology

The first step is to gather data from different sources, without discrimination. Then, data are documented and formatted into a centralized database. Next, data are parsed within the cross-observer tool we developed.

3.1. Data formatting

It can be quite difficult to integrate data from various sources in a single database as the structures and formats are usually incompatible. Figure 1 illustrates the structure of the integrated database that has been developed for the project. First, a metadata table references the sources of data available: table source, description, key fields (fields that are used for identification, temporal and spatial purposes) and the documentation of all fields in the event table.

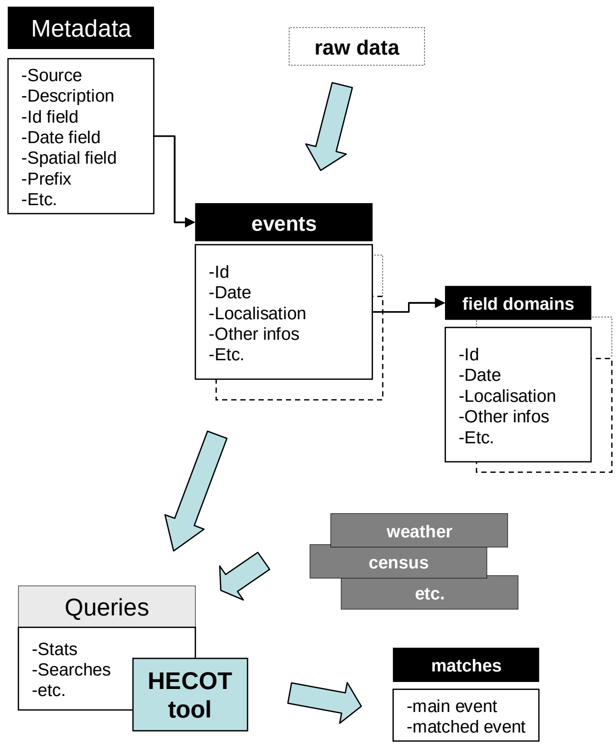

Figure 1: integrated database structure

Then, a master table is built for each event database, respecting the original format of each source. Finally, a table is created to store the domain values of each field of the master table, with a special naming scheme made from a concatenation of the prefix representing the source and the field name. This allows the direct identification of each value's source of data. Tables containing other observations like weather, census, etc. are stored aside in the same database. These tables are also identified as sources in the metadata.

This structure is useful for queries and cross-observations (the HECOT tool is explained hereafter). For example, the structure allows building groups of values for each field to facilitate the production of statistics across the master tables. The usage of key fields enables the representation of data from multiple sources on the same map or the same temporal scale.

3.2. Hazmat Event Cross-Observer Tool (HECOT)

Integrating data within Database Management System software (DBMS) was not enough to perform data analysis for this study. The use of classical SQL queries can be sufficient to obtain general statistics, but the analysis of single events in relation with others necessitates the use of an interactive tool as we need to “observe” data from several sources simultaneously. This has been demonstrated in past projects (Chapleau et al. 1997, Besner et al. 2005).

The Hazmat Event Cross-Observer Tool (HECOT) is a Microsoft Access lightweight tool aimed: 1) at identifying the same event through different databases and 2) at identifying the possible causes for the event. The tool contains the integrated database structure shown in Figure 1. However, some of the event tables are attached to, and not embedded in, the Access database.

HECOT’s main screen (Figure 2) is divided in three parts. The left part presents data on the main event analyzed. This event can come from any of the sources that are described in the metadata. For example, we can choose an accident or a weather event as the main event. The upper right part of the screen is a map displaying the main event’s location and the other events that were found in its surrounding during a spatial search. Each event can be identified on the map (Microsoft MapPoint component). The lower right part is the search results display section. That section is updated each time a main event is selected. More information on that section will be given later.

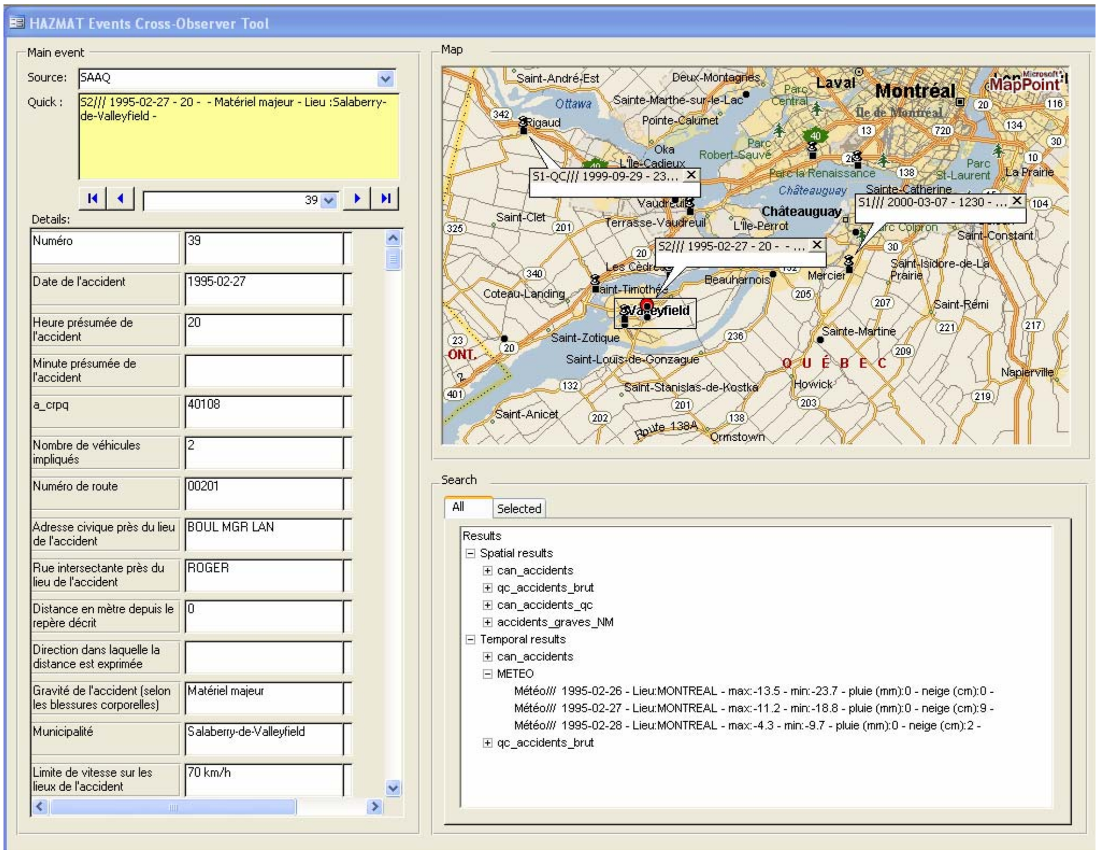

Figure 2: HECOT main screen

3.3. HECOT search process

The HECOT search process is based on the assumption that all hazmat events can be localized in space and time, which is usually the case in accident databases.

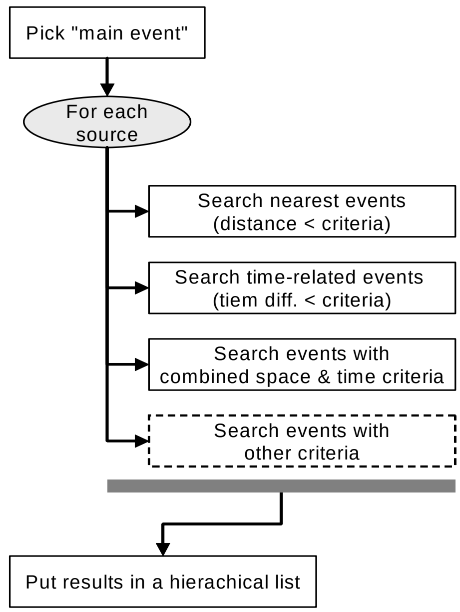

Figure 3: HECOT search process

After selecting an event that will be the center of analysis (main event), the software repeats the following searches using all the data sources available in the database (Figure 3):

- events that happened anytime within a given distance around the main event's location;
- events that happened within a given time frame anywhere before or after the main event's date (within a given time gap);
- events that happened within a given time and space gaps around the main event's location and time (intersection of the two previous searches);
- events that match other criteria not based on space and time.

All results are put in an easy to navigate and interrogate hierarchical list (the content of the search result section mentioned in section 4.2). The software uses the spatial fields (usually longitude and latitude in geographic coordinate system) and the date field of the source tables for this purpose. Of course, no event is found for a spatial search for sources without spatial fields; the same goes for time searches.

When one of the events from the search result section is selected, its full information is displayed and the user can confirm that there is a match between the main event and this event. Comments are added to qualify the match. Matches are kept in a log to be used in the cross-observation analysis.

4. Application

The methodology has been applied to hazmat accidents in the province of Quebec, Canada. Three databases were used for the study, in addition to other sources of circumstantial data:

- Dangerous Goods Accident Information System (DGAIS) from Transport Canada (S1);
- Road Accident Database from the *Société de l'assurance-automobile du Québec* (Quebec Automobile Insurance Board) (S2);
- Community Database on Work Accidents from the *Commission de la santé et de la sécurité du travail du Québec* (Quebec Work Health and Safety Board) (S3).

The Dangerous Goods Accident Information System (DGAIS) contains information on Canada's hazmat transportation accidents involving either: spills, injuries, death counts or other parameters related to hazmats (Transport Canada, 2002). Information is provided by the reports filled by the accident respondents (usually the personnel of the company involved). The database is computerized since 1988 and contains 6314 observations for the 1988-2004 periods. For the territory of Quebec, the DGAIS contains 713 events. This only includes the events that must be declared following the Canadian laws. Other events of the database were omitted.

Data of the Road Accident Database are collected by the police officers of Quebec for each accident involving injuries, deaths, trucks, large financial damages, etc. This paper examines hazmat accidents, which represent a subset of the whole database. In 2004, there were 188 recorded hazmat accidents compared to 40,368 total accidents in the database. Hazmat accidents are identified by the police officers with a special code, but no further information is available in the database.

The Community Database on Work Accidents contains the declaration of accidents where workers were injured or killed on duty. Only the events related to road transportation were studied. Unfortunately, there is no clear identification of hazmat accidents in this database that contains 5062 events. The industry code of the company was used to deduct possible involvement of dangerous goods in the accident. Moreover, the exact date of the event is known only for years 2000-2005 (2748 events).

4.1. Accident events object-model

An object-model related to hazmat road accidents helps visualize the scope of information available in event databases used in this study. Figure 4 also shows which data source (S1, S2 and S3) gives information on each object defined.

From the model, we observe that not all objects are described in all the source databases, due to their specific role. For example, the workers' accident database (S3) does not contain data on accident site but describes the carrier company of the truck driver. Also, the model separates the vehicle itinerary (on the road network) from the hazmat itinerary. As we know, the vehicle itinerary could be only a fraction of the hazmat itinerary within the logistics supply chain.

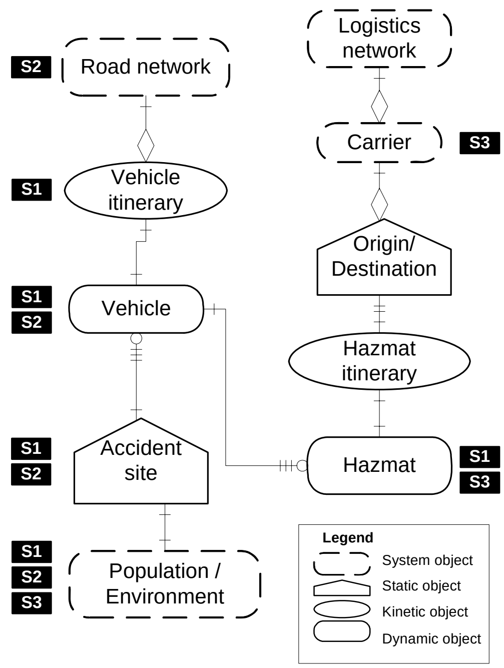

Figure 4: Hazmat road accident object-model

4.2. General statistics

We can compile several statistics from the sources, given the large number of characteristics that are reported for each event. The goal here is to demonstrate which kind of results can be obtained. Of course, several other results are expected.

First of all, Table 1summarizes the number of records available in each database, and shows the availability of data for each year. We can see that numbers are quite stable throughout the years although there are some variations. For S1, only accidents located in

Quebec's territory (11,3% of Canadian accidents for about 25% of the population) are considered.

Table 1: Number of events in source databases

Year	S1	S2	S3
1988	49		
1989	62		
1990	51		
1991	51		
1992	53		
1993	35		
1994	31		
1995	33	195	510
1996	57	185	438
1997	38	186	444
1998	49	145	429
1999	64	183	493
2000	57	191	480
2001	42	154	469
2002	41	184	484
2003		182	449
2004		188	450
2005			416
Total	713	1793	5062

Several statistics can be compiled from the S1 source (Dangerous Goods Accident Information System). On this paper, we focus on the type of hazmat involved in accidents.

Figure 5 shows the distribution of road accidents according to hazmat types. In Quebec, corrosive materials (class 8) are more likely to be involved in a road accident as the province is an important producer of sulphuric acid. On the other hand, in Canada, gases and flammable liquids are related to more road accidents than in Quebec.

The S2 source (Quebec Automobile Insurance Board) does not contain data on hazmat types. The information can sometimes be found on police accident reports, but it is not compiled in the database. Our interest will be on accidents circumstances such as weather or road conditions. Figure 6 shows that snowy and icy conditions are found in some of the accidents, but most of them happen with a clear sky and a dry road.

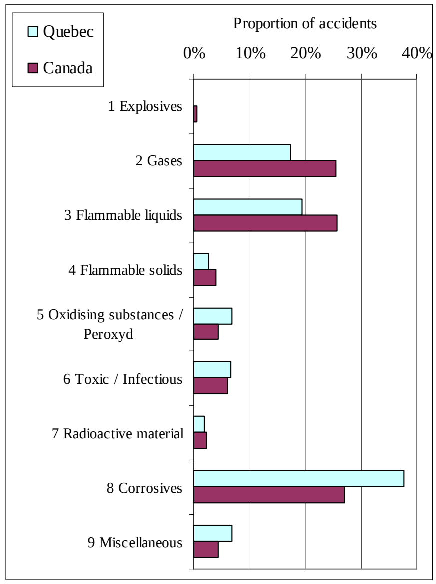

Figure 5: Distribution of hazmat types for Quebec and Canada road accidents, 1988-2002 (S1) shows the kind of information found in S3 database (workers accidents).

We can see that a large portion of road accidents is due to a loss of control or a situation where no other vehicles are involved. Also, for collisions between vehicles moving in the same direction, the average age of workers involved is slightly lower than the average. For collisions between vehicles moving in opposite directions and for collisions involving pedestrians, the workers involved had to spend more time off work after the event. We observe the opposite situation for accident with fixed vehicles.

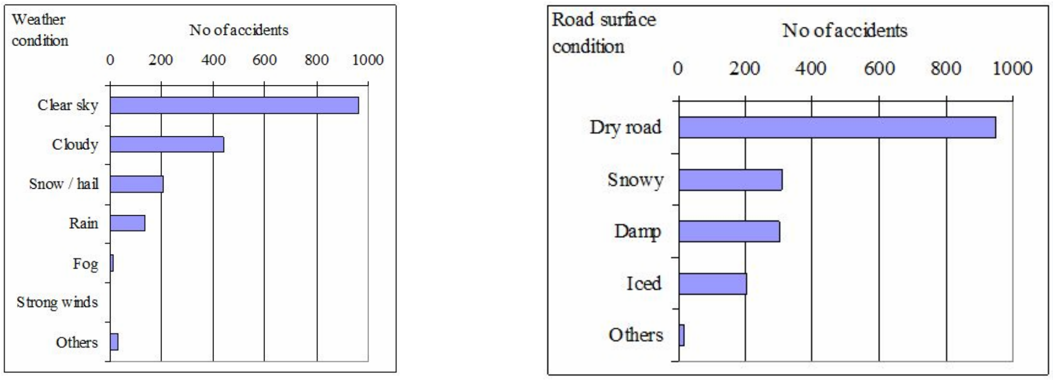

Figure 6: Distribution of weather and road conditions for Quebec road accidents, 1995-2004 (S2)

Table 2: Statistics on road accident type, S3 (1995-2005)

Accident type	%tot	AAD	ADOW
Collision between vehicles, loss of control	44.3%	38.6	115.0
Collision between moving vehicles, opposite directions	12.5%	38.3	131.7
Pedestrian hit by a vehicle	12.3%	37.9	141.7
Collision between moving and fixed vehicles	12.1%	39.0	84.6
Collision between moving vehicles, same directions	11.9%	37.6	103.6
Collision between moving vehicles, at intersection	6.9%	39.2	90.2
Collision between moving and fixed obstacle	5.9%	38.8	101.4

AAD: Average age of drivers, AODW: Average days off work

4.3. Cross-Observation

The cross-observation of the three databases is conducted with the help of HECOT. The goal is to find matches between databases, in order to evaluate the quality and the coverage of each database. The criteria for temporal and spatial searches have been widening to obtain more cross-referencing between events, due to:

- Spatial imprecision of events' location. Some accidents are reported to the municipality level. In rural area, Quebec municipalities can have an area up to several thousands square kilometers.
- Temporal imprecision of event reporting. There is often a difference between accident times across databases due to the different reporters.
- Imprecise information in general. There can be orthographic errors, unofficial location names, coding and entry errors, etc.

The preliminary results (March 2006 analysis) on the Venn diagram of the Figure 7 show that very few accidents are reported in more than one database. This could be normal as the databases do not necessarily have the same scope and use. But, for example, an accident that causes the death of a hazmat truck driver should be found in all three databases. Only 3 events were positively part of all three databases. Obviously, further developments will have to be done in HECOT to find more cross-observations (the cross-observation between S1 and S3 is not completed yet).

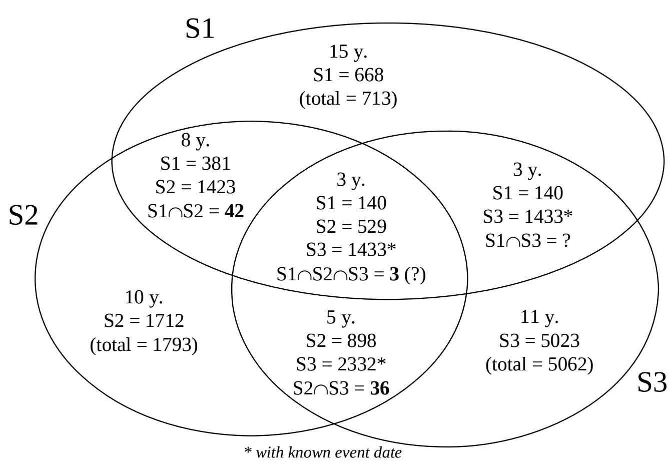

Figure 7: Cross-observations between S1, S2 and S3 sources

However, we would like to highlight the fact a total of forty-eight accidents with death counts in S2 source were not found in the S1 database, even though Canadian laws oblige the report of these accidents to Transport Canada. Also, 40 major accidents with injuries and death counts found in the S1 database were not found in the S2 database. Twenty-three of these accidents happened during the transportation phase and 13 during loading/unloading operations. In this case, it is quite possible that not all hazmat accidents are coded by police officers in the S2 database and that these accidents could be found in the Quebec Insurance Board global database, which was not available to us at the time of the study.

5. Conclusion

This paper is the first step to study databases dedicated to accidents resulting from hazmat road logistics activities. The examinations of accidents across different databases have provided a lot of additional information about some of them.

However, efforts still have to be done in order to correctly report road accidents involving hazardous materials. This study has demonstrated that the union of different databases is possible, but there is a need to clearly identify key attributes such as spatial location, date and time, and the presence of hazmat in each database. The study also proved that many hazmat events are missing from these databases, even though their declaration is mandatory. However, it is difficult to assess the risk of hazmat road operations if no database covers all accident's parameters and if release events are still systematically under-reported.

Further work will look at private companies' databases, where information on accident is much more detailed than in official database. These new databases will be included in the HECOT tool.

This work is part of a larger project that tries to assess the risk throughout the whole hazmat logistics supply chain. Cross-observations with HECOT will be extended to hazmat accidents related to other modes of transportation (rail, air, ship) and also to fixed locations (facilities, warehouses, plants, etc.) to match for example accidents during loading/unloading operations.

Acknowledgments

The authors would like to acknowledge the contribution of the CIRANO research center's partners of the project, especially those who provided the databases for this study: Transport Canada, the Quebec Ministry of Transport, and the *Institut de recherche Robert-Sauvé en santé et sécurité du travail* (IRSST). Special thanks to the *Institut National de l'Environnement Industriel et des Risques* (INERIS) from France, that initiated this study through the GLOBAL project.

References

Besner, M.-C., Gauthier, V., Trépanier, M., Prévost, M., Leclair, M., 2005. Interactive analyser for understanding water quality problems in distribution systems. Urban Waters 2 (2), 93-105.

Chapleau, R., Trépanier, M., Lavigueur, P., Allard, B., 1997. Origin-Destination Survey Data Dissemination in a Metropolitan Context: A Multimedia Experience. Transportation Research Record 1551, 26-36.

Dobbins, J.P., Abkowitz, M.D., 2003. Development of a centralized inland marine hazardous materials response database. Journal of Hazardous Materials 102 (2), 201-216.

Erkut, E., Verter, V., 1995. Hazardous materials logistics. In: Facility Location, Z. Dreznet, Ed. Springer, New-York.

Hobeika, A.G., Kim, S., 1993. Databases and needs for risk assessment of hazardous materials shipments by trucks. In: Transportation of hazardous materials: issues in law, social science, and engineering, L.N. Moses and D. Lindstrom, Eds. Kluwer, Boston.

Laberge-Nadeau, C., Maag, U., Bellavance, F., Lapierre, S.D., Desjardins, D., Messier, S., Saidi, A., 2003. Wireless telephones and the risk of road crashes. Accident Analysis and Prevention 35 (5), 649-660.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34 (2), 149-161.

Milazzo, M.F., Lisi, R., Maschio, G., Antonioni, G., Bonvicini, S., Spadoni, G., 2002. HazMat transport through Messina town: from risk analysis suggestions for improving territorial safety. Journal of Loss Prevention in the Process Industries 15 (5), 347-356.

Ohtani, H., Kobayashi, M., 2005. Statistical analysis of dangerous goods accidents in Japan. Safety Science 43 (5-6), 287-297.

Shorten, C.V., Galloway, J., Krebs, J.G., Fleming, R.S., 2002. A 12-year history of hazardous materials incidents in Chester County, Pennsylvania. Journal of Hazardous Materials 89 (1), 29-40.

Spadoni, G, Egidi, D., Contini, S., 2000. Through ARIPAR-GIS the quantified area risk analysis supports land-use planning activities. Journal of Hazardous Materials 71 (1-3), 423-437.

Stevens, A., Minton, R., 2001. In-vehicle distraction and fatal accidents in England and Wales. Accident Analysis and Prevention 33 (4), 539-545.

Transport Canada, 2002. Système d'information sur les accidents concernant les matières dangereuses. Database Documentation.

Trépanier, M., Chapleau, R., 2001. Linking Transit Operational Data to Road Network with a Transportation Object-Oriented GIS. Urban and Regional Information Systems Association Journal 13 (2), 23-27.