

ROAD SAFETY ANALYSIS USING WEB-BASED COLLABORATIVE GIS

F. Pirotti ^a *, A. Guarnieri ^a, A. Vettore ^a

^a CIRGEO, Interdepartmental Research Center on Cartography, Photogrammetry, Remote Sensing and GIS, University of Padua, 35020 Legnaro, Italy - (francesco.pirotti, alberto.guarnieri antonio.vettore)@unipd.it

Commission IV, WG IV/5

KEY WORDS: collaborative webgis, road accidents, road safety, geospatial analysis

ABSTRACT:

Road accidents have a definite impact in terms of lives, welfare and traffic congestion, especially with increase in mobility and number of overall vehicles. Prevention policies can greatly benefit from data analysis of positions and accident characteristics. This work presents a practical integration of open source software for the implementation of an instrument to insert, store, automatically analyze accident data, and consult results. As different public bodies survey accident data, a collaborative environment is necessary, and a web-based solution is ideal for permitting multi-user access and data insertion. PostGIS and UMN Mapserver with PHP server-side scripting are a classic solution for this type of instruments; this note will discuss how to organize these elements in order to build a portal for optimal data insertion, automatic statistical analysis and visualization of final result. Road center axes are processed to extract homogeneous segments which are then connected with a one-to-many spatial relationship of “closest-point-to-segment” road accident georeferenced positions. Accident points are also connected with a series of tables reporting accident-specific information. From this data an automatic procedure can extract specific indices to represent risk and danger, such as severity index, mortality index, vulnerability and other. Not only can this information be viewed as a thematic layer for fast interpretation, but change over time can be also analyzed when datasets are present over multiple years. With this information public decision makers can assess efficiency of past policies and decide future intervention.

1. INTRODUCTION

Road accidents take a significant toll on lives every year. Road accidents in industrialized and developing countries is one of the primary sources of injuries and deaths and its prevention is one of the main objectives of the World Health Organization (World Health Organization, 2002) as well as of the European Union has set the objective of reducing deadly accidents 40% by 2010 (European Council Communication, 1997). Geographic analyses can highlight danger areas by clustering techniques as well as defining risk factors and risk areas (Chini et al., 2009). A comprehensive management platform, and an effective emergency incident analysis can timely and efficiently organize emergency rescue operations. A Geographic Information System (GIS) and related spatial information technology can provide that sort of information and the tools for the analysis of spatial data and the representation of the results in spatial format (Fateh-Moghadam et al., 2005; JIugang et al., 2009). The idea behind disaster management preparedness can be also applied to road accidents if an accident is considered a perpetuated disaster with a certain risk factor which can be inferred and reported on the territory. Defining high risk “hotspots” can be of great help when planning for mitigation actions. Classic disaster spatial analysis greatly improves response effectiveness (Yang et al., 2007) and the objective of the following analysis is to apply it to road accidents to see the same effect reverberated over immediate response as well as medium to long term policy planning. The aims of the paper is to report on the methods used to implement an information system which ties together surveyed accident and traffic data with geographic data in order to carry out spatial analysis. Information of interest and statistics will

help decision makers on their assessment of ways to reduce the number of casualties from road accidents, in line with the European Community’s target to decrease accidents by 40% between 2010 and 2020 (SafetyNet, 2009), especially in light of the fact that the previous decade’s target has not been met (50% reduction between 2001 and 2010) (ECMT, 2003 2004).

2. MATERIAL AND METHODS

The project was carried out with a specific dataset regarding accidents and traffic monitoring along specific road segments. The methods aim at extracting statistic information which is of interest for planning actions to abate the number and mortality of accidents.

2.1 Material on road traffic

Traffic dataset comes from a campaign for monitoring specific road spots using sensors for recording vehicle speed and timestamp. The sensors have a known geographic position (see figure 1). The recorded data was delivered on MSEExcel spreadsheets reporting the number and average speed of vehicles detected by the sensor in a 15 minute time span. Other information recorded was temperature and moisture condition class (either “wet” or “dry”). The total number of ten sensors are planned to be installed, whereas at the time of writing three are active and their data is used. For each monitoring spot two sensors are installed, one for each side of the road. The side is defined by a directional attribute which labels the dataset, either east-west or north-south depending on the road’s main direction.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

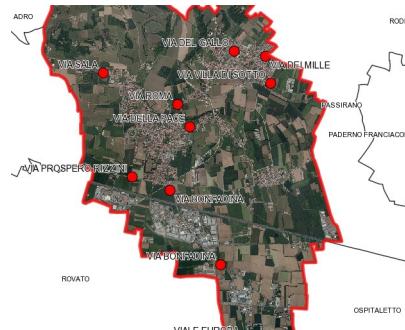


Figure 1. Location of vehicle speed sensors.

2.2 Material on road accidents

The accident data included spatial position and a set of standard accident-related alpha-numeric attributes (Amato R., 2005). The list of attributes to be recorded are regulated by Italy's Istituto Nazionale di Statistica – ISTAT, the National Statistics Institute. All public agencies which survey the accident site are bound to record the accident's characteristics and to communicate the data on a monthly basis to ISTAT. The location of the accident can be done on toponymy and on a coordinate basis; the former uses the names of the region, town, street and progressive kilometric distance, while the latter uses spatial coordinates in Italy's cartographic reference systems (Rome40 datum). ISTAT requires only toponymy-based geolocation. Our data had coordinate information and was thus inserted on the webgis portal along with attributes.

2.3 Geographic database

The original informative layers used concurrently with the data collected from roads described above are the following: the town's boundaries, the road axes and intersections, the orthophotos at 1:5000 nominal scale in a wavelet compressed format (ECW). The last two elements come from a 2008 cartographic survey. The road axes represent the center-line of the road and contain related information on road type, width class and material of road surface and current state of the surface (damaged or consumed). These attributes come from the Topographic Database guidelines which give indications on what attributes and elements to extract from photogrammetric interpretation (IntesaGIS, 2006). The guidelines provide an answer to European guidelines for geographic database integration (INSPIRE, 2007). The availability of such data is a clear result of success of the efforts towards improvement of land surveying by means of specifications at European, national and regional level (Dell'Orto and Guzzetti, 2009).

2.4 Web Geographic Information System

All the materials regarding the surveyed data have been collected in an information system composed solely of open source components. Figure 2 provides with an overview of the various components and how these provide a seamless data flow for analysis and publication on the web portal. The data from accidents and traffic monitoring is stored in Postgresql, a relational Database Management System (DBMS), as well as related spatial data which is stored in Postgis format inside Postgresql tables. Postgis also provides several spatial functions which are used to correlate data with spatial objects as illustrated in the next sections.

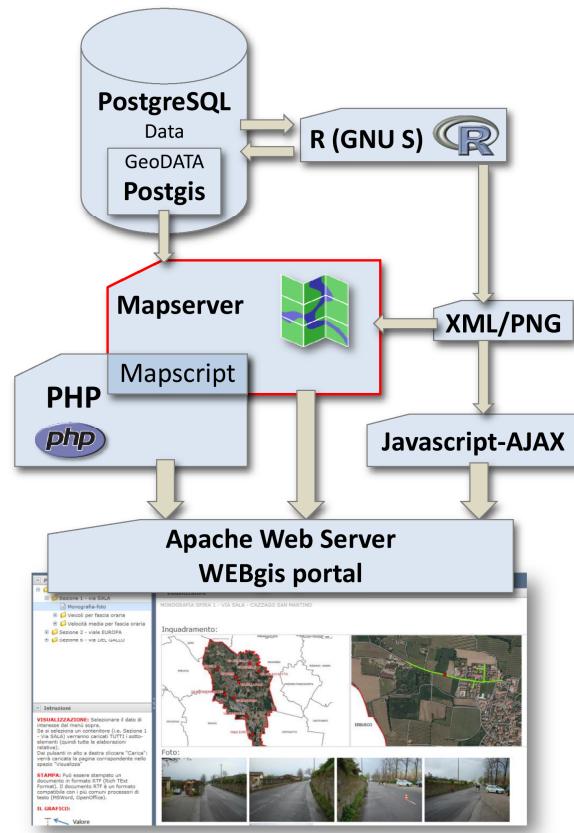


Figure 2. Outlook on the data flow and overall information system structure.

The UMN Mapserver suite, an open-source mapping engine, provides with the means to process all geographic objects along with a thematic mapping derived from association with analysis results, to be passed to the web server for publication in the webgis portal. A server-side scripting language, PHP integrates with Mapserver via the Mapscript module to increase mapping functionality and concur to map publication on the web. AJAX (Asynchronous JavaScript and XML) takes care of layout and navigation of resulting data. The latter comes from analysis and will consist in thematic maps and data plots.

2.5 Data Analysis Methods

The data flow starts from the Postgresql database which is accessed by Mapserver for what concerns vector objects, and by R for data statistical analysis (figure 2). The results from statistical analysis are then shown in the web portal either directly from graphics created with R (plots – see figure 3) or by thematic classification of vector objects when R produces attribute data.

Spatial objects: both road accidents and traffic characteristics can be spatially associated either directly to cartographic coordinates (absolute position of the accident/sensor) or to related vector objects such as road axes. Road axes are multi-line objects which break at intersections, thus lines are made of segments of different lengths. The objective is to spatially associate accident positions to the nearest line segment. To do a first move towards normalization, a pre-processing step was taken to divide all road axes into segments with lengths of 50 m or less. Result is a collection of line segments which are successively a spatial function of nearest-points-to-line was used to aggregate all accidents which fall nearest to a line

segment. SQL language construct was used to process Postgis data and aggregate functions were used to calculate statistics from detected nearest multi-point collections (Obe and Hsu, 2010). Aggregate statistics were linked as attributes to segments.

Accident data: this data is processed to extract indicators and to associate results to the nearest road segments by calculating aggregate statistics as indicated in the preceding section. The attributes of interest where the number of deceased persons in 24 hour, the number of deceased persons after 30 days and the number of injured persons in the accident. The first two are grouped together to get the total number of deceased persons. Four indicators are calculated from the n number of accidents which are associated to a specific road segment by the above-illustrated nearest neighbor procedure; fatality rate, injury rate, hazard (Amato, 2005) and risk. The fatality rate (FR):

$$FR = \left(\frac{D}{A} \right) \cdot 100 \quad (1)$$

where D is the number of deaths,
A is the number of accidents

the injury rate (IR):

$$IR = \left(\frac{I}{A} \right) \cdot 100 \quad (2)$$

where I is the number of injured persons,
A is the number of accidents

the hazard (H):

$$H = \left(\frac{D}{D+I} \right) \cdot 100 \quad (3)$$

The risk (R):

$$R = \left(\frac{A}{L} \right) \cdot 100 \quad (4)$$

where A is the number of accidents,
N is the number of accidents.

The risk therefore gives an idea of the chance of having an accident when traveling over a certain road segment, and the hazard tells the reader what is the amount of adverse effect that can be suffered by the people in the accident. In the case of figure 3 two accidents are present near a certain segment, and their data will be associated to that segment.

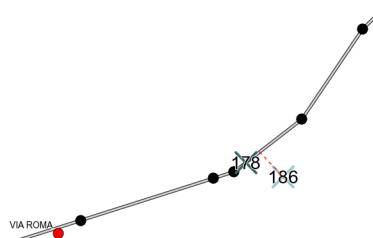


Figure 3. Road segments with accident positions.

Traffic data: the recorded log of the traffic sensors is downloaded at constant time intervals (every three days or more) and all data is processed for extracting statistical

indicators. The information which is of interest is then appended in the webgis portal for open consultation for stakeholders. The two main distributions which are analyzed are the number of vehicles and their speed measured and assigned to a time interval. The daily set of information is divided on a per-hour basis to extract mean, standard deviation, minimum and maximum value of the subset of the total distribution. These statistics form the basis for graphical representations via boxplots (see figure 5) and for further analysis to investigate on significant differences from the overall mean and from the expected – for example if the mean hourly average speed is significantly above the speed limit. All calculations are carried out using R spatial and non-spatial modules - rgdal, spatstats and RODBC. The latter is necessary to read the data source which is given in MSEExcel2007 format. Cells with data are read and processed automatically by the R script, which can be launched also from the webgis portal. This is a further example of integration with the information system (figure 2) which is structured to be operational also remotely.

3. RESULTS AND ANALYSIS

The project was carried out with a specific dataset regarding accidents and traffic monitoring along specific road segments. The methods resulted in thematic maps which clearly show which areas are more dangerous and more at risk than others.

3.1 Road accidents

The figure 4 below shows some map extracts where road segments are mapped with thematic colors which grow from white to dark red as the rate value increases. It is interesting to detect areas which have a higher risk of accident, but which do not necessarily have a higher hazard value. Some spots can be more dangerous in terms of probability of accidents happening, but the accidents do not necessarily correlate with injuries or deaths. For example at intersections the risk of road accident is quite high, but because most crashes happen at limited speed, the injury rates and fatality rates are low; especially the latter rate is low at intersections as clearly seen in figure 4 where the intersection at the lower left of each quadrant is dark red relating to injury (A) and risk (D), but white in scale, meaning without any fatalities, for fatality (B) and hazard (C).

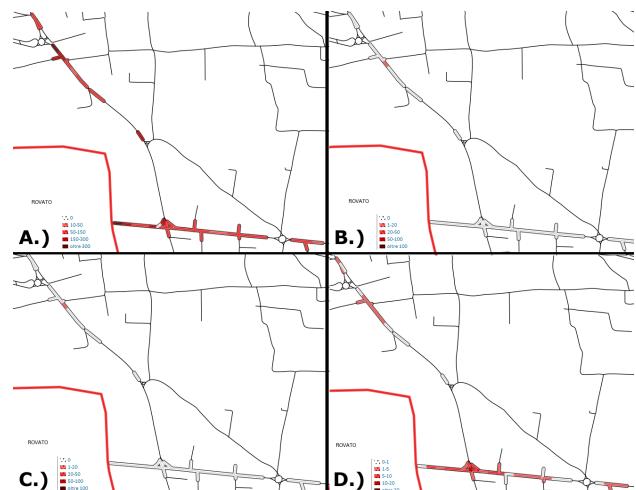


Figure 4. Map extracts from the webgis portal depicting an area with A.) – Injury rate, B.) - Fatality rate, C.) – Hazard, D.) – Risk. Colors from white to dark red.

The webgis portal puts all results online and lets end-users evaluate what road segments and what situations present the higher risk and hazard. Having set the spatial analysis as an SQL construct future data updating will automatically update all statistical analysis. Collaboration with local police forces, which are responsible for surveying road accidents, will create a considerable added value to the system. This is done by a module which allows to upload a table with the relative data – x and y columns and the attributes dictated by ISTAT. It is important to educate public administrations with correct procedures for importing data and for the use of the webgis portal as knowledge transfer is a critical aspect for the distribution of new technology to a wide spectrum of users.

3.2 Traffic monitoring

The other aspect of the project is the analysis of traffic in terms of vehicle numbers and average speeds. This undoubtedly is related to road accidents and give an added focus to driver's behavior. The webgis portal connects the monitoring sensor positions which are visible plotted on the map (figure 1) with specific information pertaining to the sensor – from respective analytics of surveyed data (figure 5) to site description (figure 7) as well as printing utilities to export in Rich Text Format for documentation and distribution of data.

What is interesting to note here is that in all cases of the three days which were monitored, the number of vehicles is obviously lower at very early morning times (figure 5), but the maximum speeds are quite high (figure 6). This denotes that a few drivers count on the fact that no control is present and speed.

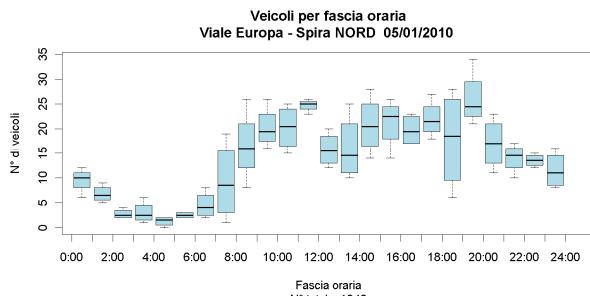


Figure 5. Hourly number of vehicle distribution analysis: x axis has the hour time bins and y axis has the number of vehicles.

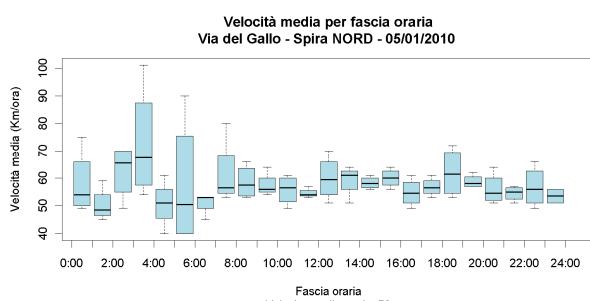


Figure 6. Hourly average speed distribution analysis plot: x axis has the hour time bins and y axis has the average speed of vehicles.

3.3 Considerations

The integration between geographical information, road accidents positions surveyed by police forces, traffic monitoring stations, can lead to a very solid database for numerous analysis of different aspects. The practical implication is that end users will be able to participate actively by inserting data and by promoting other types of analysis which interests them. For example a future implementation will regard assessment of the best position for inserting enforcement of speed limit respect. The best position can be detected by creating an indicator which compares accident risk and hazard with average speed distribution. Spots which have a high value for both aspects will need enforcement solutions, whereas if a spot has a high value for the former but a lower value for the latter, it indicates that structural solutions are needed, perhaps visual alerts or speed deterrents depending on the specific case.

4. CONCLUSIONS

The information system described and evaluated in this paper wraps together a set of procedures for extracting information of interest for road safety evaluation starting from data concerning accidents and traffic. Such webgis-oriented platform processes spatial data to provide end-users with a reliable tool to evaluate what roads have the highest accident density, the highest danger rate and any other statistical indicator which can be extracted from attributes. Another example of interest is which roads are more dangerous for pedestrians as opposed to roads which are more dangerous for vehicles.

The collaborative environment is an added value where future surveys update the database with new information. As matter of fact it is important to gather data from different sources; accident surveys are carried out by police forces and successively communicated to ISTAT, georeferencing is done successively by third party bodies, traffic monitoring is carried out by the district or by projects at different levels of organization (regional, national or European). The webgis portal is an ideal tool to store all such information and to give also a way to import or to directly georeference new upcoming statistics. All processes are stored in the information system, therefore all processing for re-calculation and analysis is automatic.

Future updating will also improve comparison of the statistics on a time basis. What is the overall trend of number of vehicles and average speed on a certain road? Have accidents decreased over time? Has excessive speeding decreased on a certain road? Are there hotspots which denote a dangerous area? What are the causes of the high danger rate on an area? What are the solutions? The objective of this work is to facilitate answers to all these questions. The updating process via web, and the distribution of results from analysis via images of plots and tables with data, are two aspects which can be well done using collaborative geographical information systems. The open source world provides with optimal tools which can be made to seamlessly integrate with one another, allowing investors to concentrate budget on training personnel surveying and for research towards improvements of the software and of the analysis procedures.

REFERENCES

- Amato R. 2005. *Statistica degli incidenti stradali*. ISTAT Ed. Sistema Statistico Nazionale-Istituto Nazionale Di Statistica-Automobile Club D'Italia.
- Chini, F., Farchi, S., Ciaramella, I., Antoniozzi, T., Rossi, P.G., Camilloni, L., Valenti, M., Borgia, P. 2009. Road traffic injuries in one local health unit in the Lazio region: results of a surveillance system integrating police and health data. *International Journal of Health Geographics*, 8:21.
- Dell'orto C., Guzzetti F., 2009. Cartographic base for territorial planning in Lombardy Region. Proceedings of 27th Urban Data Management Symposium June 24-26 2009, Ljubljana, Slovenia.
- EMCT - European Transport Safety Council 2003. *Assessing risk and setting targets in transport safety programmes*. <http://www.etsc.eu/documents/riskassess.pdf>
- EMCT - European Transport Safety Council 2004. *Road safety: implementation of the objective -50% killed by 2012*. CEMT/CM 12, ECMT, OECD, Paris.
- Fateh-Moghadam, P., Dallago, G., Piffer, S., Zanon, G., Menegon, S., Fontanari, S., Furlanello, C. 2005. Epidemiology of road traffic accidents in the province of Trento: first results of an integrated surveillance system (MITRIS). *Epidemiol Prev.*, 29(3-4):172-179.
- IntesaGIS 2006. *Il Catalogo degli Oggetti*. www.cnipa.gov.it/site/_files/1n1007_12_vers2006_33.pdf.
- INSPIRE - 2007/2/EC 14 March 2007 - http://eur-lex.europa.eu/LexUriServ/site/it/oj/2007/l_108/l_10820070425i_t00010014.pdf.
- Jiugang, L., Xinning T., Zhengjun L., Minyan D. 2009. Design and Implementation of Webgis for Government emergency management based on SOA. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXVIII-7/C4, pp 149-152.
- SafetyNet 2009. Quantitative road safety targets retrieved on 13 June 2010. http://ec.europa.eu/transport/road_safety/
- Yang, M.D., Lin, C.C., Chen, S.C., Su, T.C. 2007. A Web-GIS disaster management system applied in central Taiwan. In: *2nd International conference on urban disaster reduction*. November 27-29, 2007, pp 213-218.
- Obe, R.O., Hsu, L.S. 2010. *PostGIS in action*. Manning Publications Co. Eds. pp. 95-98.
- World Health Organization 2002. *The injury chartbook: a graphical overview of the global burden of injuries*. Geneva, World Health Organization.

ACKNOWLEDGEMENTS

The data used for analysis is part of a pilot project carried out by Cogeme Informatica s.r.l. in collaboration with the Province of Brescia, with Prof. Maurizio Tira (University of Brescia) in the town of Cazzago San Martino (BS).