



DESIGNING A QGIS-BASED MULTIMODAL DATABASE FOR LEVEL CROSSING MAPPING: A CROATIAN CASE STUDY

Danijela Barić, Silvestar Grabušić

University of Zagreb, Faculty of Transport and Traffic Sciences, Croatia

Abstract

Considering the wide range of applications and possibilities for using an open-source Geographic Information System in transport infrastructure planning, this study examines the use of Quantum Geographic Information System to map level crossings on Croatia's railway network. By defining a relevant set of parameters, a comprehensive dataset was assembled, incorporating both official data from Croatian railway and road infrastructure managers (such as technical specifications, traffic accident and incident statistics, train frequencies, etc.) and field-collected data (including spatial visualisations, photo and video documentation). The main challenge was to identify and process missing data across multiple databases to create a unified multimodal database compatible with the Quantum Geographic Information System. The research resulted in the creation of a Quantum Geographic Information System database of level crossings with an extensive dataset for each level crossing, which could serve as a basis for in-depth level crossing analyses, a database for risk assessments, and for the development of future accident-prevention measures. This is the first reported use of the Quantum Geographic Information System for mapping level crossings in Croatia.

Keywords: level crossing, QGIS, railway, Croatia

1 Introduction

A level crossing is a point where road and rail intersect, posing a high risk of accidents and incidents. According to Eurostat, 25.6% of all railway transport accidents in 2024 occurred at level crossings [1]. Analysing only accident and incident statistics is insufficient to fully understand the underlying causes and patterns of accidents and incidents at LCs. A comprehensive understanding requires the collection of a wide range of basic, traffic, infrastructure, environmental, safety, time, and documentation data and integration into a unified database. Recent advances in Geographic Information Systems (GIS) enable the creation of integrated spatial databases that combine diverse data types into a coherent analytical framework. Modern GIS are powerful tools for integrating diverse datasets, allowing transportation planners and authorities to map, monitor, and analyse key points on the transport network. GIS is often used to map various data related to road infrastructure [2], the geographical distribution of road traffic accidents [3, 4], noise [5-7], and rail traffic for analysis of trespassing over a more extended period of time [8], detection of object on rail [9], assessing the risk of floods to rail [10], quality control [11], maintenance and planning [12], analysis of historical activities on rail (WebGIS) [13] safety analysis [14], and driver behaviour at level crossing [15]. There are several software programs available for mapping and spatial analysis, with each offering different benefits. QGIS (Quantum GIS) is an open-source program that provides a wide range of analysis, data editing, and cartography capabilities. ArcGIS Pro and GRASS GIS are commercial programs focused on advanced geospatial and raster analysis, while

MapInfo Pro is used in business and urban planning. Other GIS programs include Felt, SAGA GIS, GeoDa, etc. [16]. In this research, QGIS was used for its open-source status, robust capabilities for managing spatial and non-spatial datasets, support for multi-format integration, and a wide range of plugins for advanced spatial analysis.

This study focuses on developing a QGIS-based multimodal database for level crossings that integrates statistical indicators, technical attributes, spatial geometry, environmental attributes, inspection records, and visual documentation. The primary objective is to design a functional, extensible geospatial database to support comprehensive mapping and analysis of level crossings. At the national level in Croatia, there is no officially implemented risk assessment tool for level crossings that would provide practical steps to ensure consistency and reliability when combining data sources with different structures and time scales. However, there are well-established international risk assessment models that can serve as examples of good practice [17, 18]. The developed database is tested through a case study of LC Jankovci II in Croatia, providing insight into the main characteristics of LC on the Croatian railway network and opportunities associated with multimodal geospatial data integration for railway safety management.

2 Methods and materials

The research methodology consists of five phases (figure 1). The methodology combines the collection of secondary and primary data to utilise their benefits.

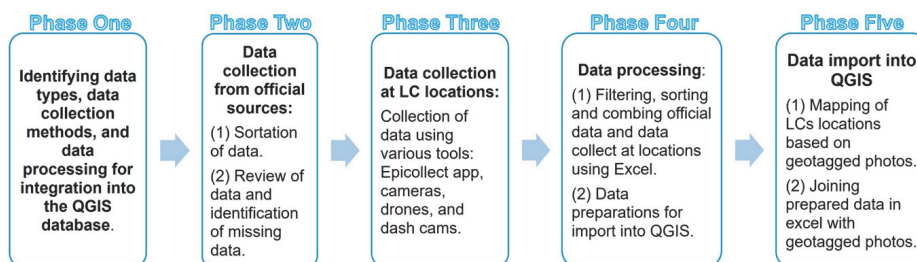


Figure 1 Phases of research methodology

The first phase involves identifying data types (which data will be collected to form variables in QGIS), choosing data collection methods, and determining how the data will be processed, structured, and integrated into the QGIS database. In phase two, data from officially available sources will be provided by the railway and road infrastructure managers. The collected data will be reviewed to identify any missing data in the official databases. In the third phase, data will be collected on-site at the LC locations. The data collection form will be designed in the Epicollect application, and data will be collected through it, which will speed up the process. The application will be complemented by recording LCs locations with cameras and drones, while a dash cam will capture LCs from the driver's perspective. In phase four, data collected during the previous phases will be filtered, sorted, and consolidated into a single unified database that encompasses all datasets. The combined database will be prepared for correct import into QGIS. This includes preparing the variables and recordings in the shapefile format and exporting the database to csv. Finally, in phase five, the mapping of LC locations will be generated from geotagged photos as points on the QGIS map. The geotagged photos (LC locations) will be linked to information about each LC via the same-named joined variables (geotagged photo IDs = the same IDs in the dataset). Examples of required tools and software that facilitate easier data collection and processing are shown in figure 2.

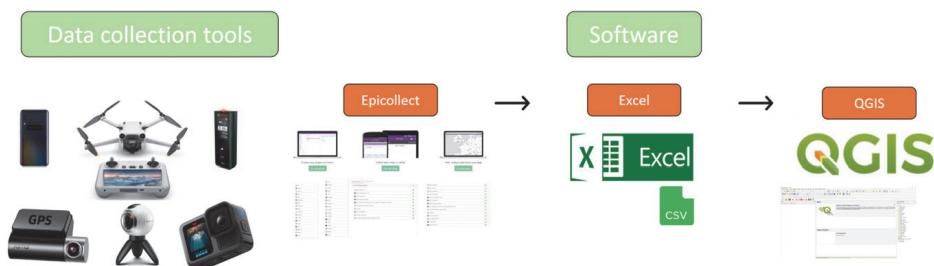


Figure 2 Tools and software for collecting and processing research data

3 Results and discussion

3.1 Data collection from secondary sources

The railway and road infrastructure managers provided data on level crossings in Croatia from official secondary sources. The variables consisted of the railway line ID, kilometre position of LC on the railway line, LC name, road ID, county, city, average annual daily traffic (AADT), speed limit on approach to LC, frequency of trains, train speed on the section of railway line, existing protection and equipment at LC (classifications), and the final technical solution for LC (FTS). These variables were used to define the spatial, administrative, and technical characteristics of each level crossing. Statistical accident data were provided for each LC. They included information on the number of accidents, incidents, and collisions with barriers and half-barriers, the number of injured and fatally injured persons, and material damage. The accident data refer to the period from 2010 to 2024 [19, 20].

3.2 Data collected on the locations of level crossings

Data collection at LC locations was conducted to complement and verify portions of the secondary data obtained from official sources, to capture the actual conditions at the LCs, and to identify any discrepancies between official records and on-site conditions. The variables included are the kilometre position of LC at the location and the name of LC at the location, both of which could differ from official sources, traffic counting at location, existing protection level and equipment at the location of LC (may vary from official sources), material surface between road and rail, number of surface and dimensions, conditions of surface, width of road 3 meters from the closest rail, lighting at LC, vegetation at LC, triangle of visibility at LC, conditions of road and road markings, conditions of traffic signs, date and time of collection, photos of LC and proposition for safety improvements. The dash camera was used to collect recordings from the driver's perspective, enabling better identification of traffic, infrastructure, safety, and environmental variables (e.g. how vegetation affects drivers' perception and visibility of LC). At the LC location, a camera was installed to record traffic (volume, structure, and potentially risky behaviour by road users). The road surface was visually analysed, and any wear and tear was recorded, as were the traffic signs on approach and at LC. The type of road markings was recorded and visually analysed for wear (worn or in good condition). No tools or devices were used to check the condition (e.g. a retroreflector). The existing equipment at LC sometimes differed from official sources. The date and time of data collection and photo documentation were recorded to analyse future changes and could be used for the LC design maintenance schedule.

2.3 Processed and integrated multimodal LC database

Variables from secondary sources and data collected at LC locations were sorted and categorised to form a single, integrated multimodal LC database. The variables were systematically categorised into six main groups: basic, traffic, infrastructure, environmental, safety, time, and documentation variables (figure 3). Basic variables include main information about LC, such as the railway line ID, the kilometre position of LC on the railway line according to official sources, and the kilometre position at the location. For some level crossings, the kilometre position didn't match the official sources. The road ID intersecting with the railway line at the level crossing was checked on the same principle as the kilometre position. The LC's geographical location was sorted by county (an administrative division in Croatia) and city (or municipality). The traffic variable included AADT data for the road section and traffic counts at the location to analyse the traffic structure of LC. The road speed limit on the approach to LC was also recorded. The train frequency was extracted from official sources, as was the train's speed limit on the section of the railway line. Infrastructure variables consist of LC protection (passive or active), equipment (traffic signals, barriers, half-barriers, sound-flashing devices, etc.), and infrastructure characteristics (condition and dimensions). Some discrepancies between the official and on-site classifications were noted. FTS is a proposal, according to official strategy documents, to increase the safety of level crossings, published by the Ministry of the Sea, Transport and Infrastructure [21]. The number of railway tracks and the surface material between the rail and the road (rubber, wood, or asphalt), with descriptions of the dimensions and conditions of the surface material. The width of the road surface was measured three meters from the closest rail, and lighting at the location included public lighting, dedicated lighting, or both. Environmental variables included information on vegetation along roads and rail, which is closely linked to ensuring the visibility triangle at LC. This information can be crucial for vegetation maintenance and can be complemented by details on the last maintenance date and time. Safety variables include historical accident data at LCs, such as the number of accidents and incidents, records of collisions with barriers and half-barriers, the number of injured and fatally injured persons, and material damage. Additionally, these variables include analysis of road conditions (cracks, bumps, sharp bends, or straight approaches to LC, slopes, etc.) and traffic sign conditions (missing signs, damaged signs, and other deficiencies). The proposal for safety improvements includes the authors' input and proposed preventive measures to increase LC safety. The variables time and documentation include the date and time of data collection at the location and the LC's photo documentation.

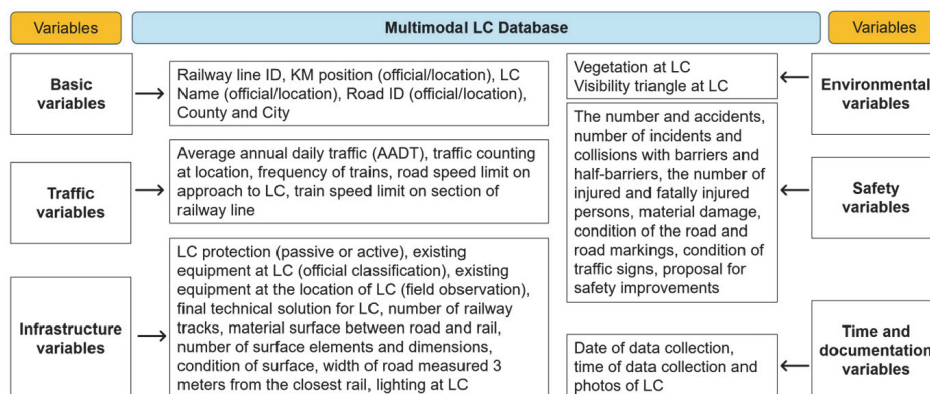


Figure 3 Categorisation of variables in multimodal LC database

2.5 Mapping and creation of the QGIS multimodal database

The final data collection step was to map and import a multimodal database into QGIS. The example is shown at the level crossing in Jankovci II, Vukovarsko-srijemska County, municipality Stari Jankovci (figure 4). For mapping, geotagged photos (geographical coordinates) were recorded at the location and imported into QGIS. These locations represent the LC location on the map. The only information stored in the points is the level crossing's geographical location. Numerical and textual data have to be assigned to each LC. The form design in the Epicollect application was used to collect most of the information at the LC location. The entire dataset collected through the Epicollect form (primary data) was exported to CSV format and imported into an Excel file, where it was joined with secondary data. The next step was to clean and prepare the data for import into QGIS. The shapefile format stores geographic features, and it has a limited number of available symbols, which is why the variable names had to be shorter in QGIS (e.g. official kilometre position of LC -> KM_off). It is also recommended to clean the text data by removing special characters, depending on the database's language. The points on the map (geotagged photos) had only an ID as a feature. The same ID was created in the database to link the points to information about each LC via the "join data feature". To access the data for each LC in QGIS, the "identify feature" option must be checked, and a single mouse click opens detailed information about each LC.

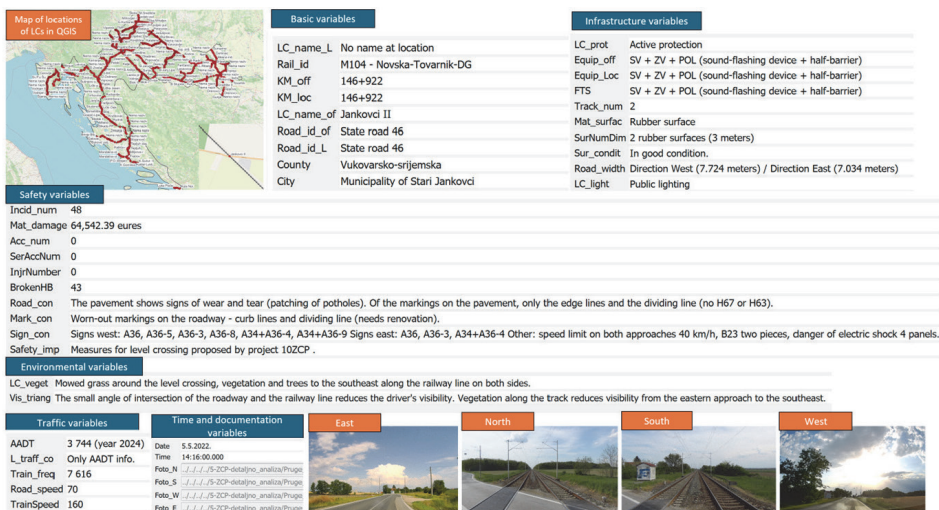


Figure 4 Example of multimodal QGIS database - level crossing Jankovci II

4 Conclusion

The level crossing database was structured as a multimodal database implemented in QGIS, integrating spatial, numerical, textual, and unstructured data (photos and recordings). The objective of this research was to systematically map, document, and visualise level crossings, including their technical, spatial, and safety characteristics. The first step was to define and select variables to include in the database, which involved identifying variables missing from official sources and screening prior literature. It was noted that the official sources lacked descriptions of the infrastructure's condition and dimensions. This included discrepancies between official records of rail and road infrastructure managers and actual LC conditions.

Currently, the databases are maintained in basic Excel spreadsheets without a centralised structure, which increases the risk of errors, duplication, and version inconsistencies. To ensure long-term accuracy, it would be necessary to establish a centralised collection database with standardised data schemas, automated validation rules, and clearly defined responsibilities for data updates and quality control. Some variables could be defined only at the LC location, so the test locations were chosen to complement the variables already selected. The main benefit of this research is the creation of a unified multimodal LC database for all deep LC analyses, which could serve as a database for risk assessments and the development of future accident-prevention measures. The multimodal database provides easy access to data on level crossings without having to check multiple sources, enabling a comprehensive understanding of the level crossing network. The combination with the open-source software tool QGIS enables visualisation, management, editing, analysis, and display of geodata. Despite these benefits, this research also has some limitations. The database does not contain all LCs; this requires additional time and resources, especially for collecting data on-site at LCs. The primary data were collected at a specific point in time. Any decisions and safety measures could prove less effective if the database is not continuously updated with new information. Another challenge in data collection is the need for multiple stakeholders to collaborate and to share data continuously. The database could also benefit from integrating other specialised data collection systems. It is also possible that secondary data may not be fully up to date, which could affect the overall quality of the database. Given this, any available information was checked for accuracy on site. Future research should focus on completing the database with all LCs using this established approach, followed by creating a web-based QGIS database and making it publicly available to other experts and the wider community. It is also possible to connect QGIS to external programming tools, such as R, to incorporate probabilistic methods (such as Bayesian networks or machine learning). The results of this research could serve as the basis for developing an LC network safety risk assessment and proposing appropriate preventive measures.

Acknowledgements

This work is funded by the European Union – NextGenerationEU as part of the institutional research project of the University of Zagreb, Faculty of Transport and Traffic Sciences (Project: Development of Human-Centred Design Solutions to Prevent Risks of Traffic Incidents and Accidents at Level Crossings (PRISK)).

References

- [1] Eurostat, Railway safety statistics in the EU, <https://ec.europa.eu/eurostat/statistics>
- [2] Ratnasari, N., Widaningrum, A.H., Pratiwi, A.H.: Geographic Information System (GIS)-Based Road Damage Mapping in Google Earth for Village Infrastructure Planning, *G-Tech: Jurnal Teknologi Terapan*, 9 (2025) 2, pp. 1033–1044, DOI: 10.70609/gtech.v9i2.6774
- [3] Shaira, H., Naik, P.R., Pracheth, R., Nirgude, A.S., Nandy, S., Hiba, M.M., Karthika, S.: Epidemiological profile and mapping geographical distribution of road traffic accidents reported to a tertiary care hospital, Mangaluru using quantum geographic information system (QGIS), *Journal of Family Medicine and Primary Care*, 9 (2020) 7, pp. 3652-3656, DOI: 10.4103/jfmprc.jfmprc_190_20
- [4] Rassu, N., Coni, M., Rombi, J., Salis, M., Maltinti, F.: Road Accidents - Study of the Evolution and Mapping of Accidents Within the Municipality of Cagliari – Italy, Gervasi, O., et al. *Computational Science and Its Applications – ICCSA 2025 Workshops, ICCSA 2025, Lecture Notes in Computer Science*, pp. 401-418, Istanbul, Turkey, 30 June - 3 July 2025.
- [5] Kumar, K., Bhartia, A., Mishra, R.K.: Monitoring, modeling, and mapping of rail-induced noise at selected stations in megacity Delhi, *International Journal of Environmental Science and Technology*, 20 (2023) 9, pp 10243–10252, DOI: 10.1007/s13762-022-04529-6

- [6] Graziuso, G., Francavilla, A.B., Mancini, S., Guarnaccia, C.: Open-source software tools for strategic noise mapping: a case study, 5th International Conference on Applied Physics, Simulation and Computing (APSAC 2021), Salerno, Italy, 3-5 September 2021.
- [7] Adza, W.K., Hursthouse, A.S., Miller, J.: Exploring the Combined Association between Road Traffic Noise and Air Quality Using QGIS, *International Journal of Environmental Research and Public Health*, 19 (2022) 24, pp. 17057, DOI: 10.3390/ijerph192417057
- [8] Oswald Beiler, M.R., Fillion, E.: Amtrak Rail Trespasser Analysis using a GIS Space-Time Approach, *Public Works Management & Policy*, 26 (2021) 4, DOI: 10.1177/1087724X211003249
- [9] Wisultschew, C., Mujica, G., Lanza-Gutierrez, J.M., Portilla, J.: 3D-LIDAR Based Object Detection and Tracking on the Edge of IoT for Railway Level Crossing, *IEEE Access*, 9 (2021), pp. 35718-35729, DOI: 10.1109/ACCESS.2021.3062220
- [10] Alabbad, Y., Beyza Cikmaz, A., Yildirim, E., Demir, I.: Flood Exposure Assessment of Railway Infrastructure: A Case Study for Iowa, *Applied Sciences*, 15 (2025) 16, pp. 8992, DOI: 10.3390/app15168992
- [11] Nathanail, E.: Framework for Monitoring and Assessing Performance Quality of Railway Network Infrastructure: Hellenic Railways Case Study, *Journal of Infrastructure Systems*, 20 (2014) 4, DOI: 10.1061/(ASCE)IS.1943-555X.0000198
- [12] Grabušić, S., Barić, D.: Effective Railway Infrastructure Planning Using GIS for Data Mapping and Analysis: Case Study Level Crossings in Zagreb, Croatia, 7th Conference on Sustainable Mobility, CSuM2024, September 4–6, 2024.
- [13] Kun, S., Piovan, S., Luigi Fontana, G.: A WebGIS for Visualizing Historical Activities Based on Photos: The Project of Yunnan–Vietnam Railway Web Map, *Sustainability*, 13 (2021) 1, DOI: 10.3390/su13010419
- [14] Keramati, A., Lu, P., Tolliver, D., Wang, X.: Geometric effect analysis of highway-rail grade crossing safety performance, *Accident Analysis & Prevention*, 138 (2020), pp. 105470, DOI: 10.1016/j.aap.2020.105470
- [15] Lee, S., Chen, T, Sze, N.N., Mao, T., Ou, Y., Mihaita, A.S., Chen, F.: Analysing driver behaviour and crash frequency at railway level crossings using connected vehicle and GIS data, *Travel Behaviour and Society*, 39 (2025), pp. 100957, DOI: 10.1016/j.tbs.2024.100957
- [16] Simon Fraser University, GIS software and mapping tools, <https://www.lib.sfu.ca/find/other-materials/data-gis/gis/software>, 12.1.2026.
- [17] Borsos, A., Gabor, M., Koren, C.: Safety Ranking of Railway Crossings in Hungary, *Transportation Research Procedia*, 14 (2016), pp. 2111-2120, DOI: 10.1016/j.trpro.2016.05.226
- [18] Ficzer, P.: The role of artificial intelligence in the development of rail transport, *Cognitive Sustainability*, 2 (2023) 4, DOI: 10.55343/CogSust.81
- [19] HŽ Infrastruktura, Internal documentation and statistics, 2025.
- [20] HŽ Infrastruktura, Safety Report 2024, 2025.
- [21] Strategic documents and projects, Program for the resolution of level crossings and pedestrian crossings for the period 2023–2027, Ministry of the Sea, Transport and Infrastructure of the Republic, <https://mmpi.gov.hr/promet/zeljeznicki-promet-129/strateski-dokumenti-i-projekti/13827>, 19.1.2026.

