Modelling Railways in the Context of Interoperable Geospatial Data

Dissertation
submitted to and approved by the

Faculty of Architecture, Civil Engineering and Environmental Sciences
Technische Universität Braunschweig

and the

Department of Civil and Environmental Engineering
University of Florence

in candidacy for the degree of a
Doktor-Ingenieurin (Dr.-Ing.) /
Dottore di Ricerca in Civil and Environmental Engineering[*)]

by
Manuela Corongiu
born 12/03/1972
from Cernusco sul Naviglio, Italia

Submitted on	16 April 2021
Oral examination on	17 June 2021
Professorial advisors	Prof. Dr.-Ing. Markus Gerke Prof. Dr. Grazia Tucci

2021

---

*) Either the German or the Italian form of the title may be used.

Ai miei figli

Acknowledgements

First and foremost, I am deeply grateful to the LaMMA Consortium in the person of the sole director Dr. Bernardo Gozzini who granted a leave of absence from a researcher's job to pursue this doctoral degree.

I want to express my sincere gratitude to my supervisors Prof. Grazia Tucci (University of Florence) and Prof. Dr.-Ing. Markus Gerke (University of Braunschweig) for their assistance at every stage of the research project. I have immensely appreciated the guide they have given me continuously and the great degree of freedom to explore the topic to fulfil my scientific curiosity. Additionally, I would like to extend my sincere gratitude to Prof. Andrea Masiero for his technical support and research contribution. Thank you all for creating such an excellent working environment.

I want to thank the University of Florence for giving me the PhD opportunity, the coordinator Prof. Claudio Borri and the secretary Margherita Mellini for their remarkable support during this hard pandemic time. I also would like to thank the University of Braunschweig, Prof. Thiele and Stefanie Theil for their coordination and administrative support.

A special thanks to who is or used to be part of the Schema and Geco Laboratories (University of Florence, Civil and Environmental Engineering Department) guided by Prof. Grazia Tucci: Valentina Bonora, Michele Brusutti, Andrea Comparini, Alessandro Conti, Irene Cortesi, Alessio Ferroni, Filippo Fiaschi, Lidia Fiorini, Franco Flamigni, Andrea Masiero, Adele Meucci, Franco Montanari, Francesco Mugnai, Francesca Panighini, Erica Isabella Parisi, Eugenio Pellis, and last but not least Enzo Santoro. I want to extend my sincere thanks to all other colleagues I have collaborated with. I really enjoyed being a part of the group. I am also grateful to the Institute of Geodesy and Photogrammetry of the University of Braunschweig to be very lovely and welcoming in Braunschweig's period, particularly to Yahya Ghassoun for his inputs and fruitful discussions.

I am deeply grateful to Tiziana De Filippis, Maurizio Fenn, and Francesco Mugnai for their insightful comments and suggestions in reviewing the thesis.

I had the privilege to collaborate with Mauro Rossi, and I would like to offer my special thanks to him for being my research inspiration.

On a personal level, I would like to express my gratitude to my family: my husband Lorenzo and my children Cristiano and Elena, for their continued love and support during my PhD research. I have also immensely appreciated all the encouragement I received from relatives and friends. I thank you for everything.

Abstract

In geospatial information, the interoperability term must be defined at different levels to fully consider the design of complex spatial infrastructures: sematic, schematic, syntax, and, above all, on processes and steps required to be shared in a common framework. The interoperability issue is the keystone of the research topic and is analysed through different aspects and points of view, with a focus on three relevant aspects. First of all, the 3D information: from the cartographic point of view (2.5D) to fully 3D models. Then, the link between reference geoinformation and geospatial thematic applications applied in the context of railway infrastructures. Finally, multi-source information in an integrated spatial database is analysed in management, validation, and update over time.

The proposed approach starts from the reference data based on 3D geotopographic information. The research aims to devise a prototype process of a 3D data model able to describe firstly geospatial databases derived from cartography maps, then a spatial model shareable among different territorial applications and analysis. 3D city models and Building Information Model (BIM) connection has been considered. The case study refers to railway infrastructure contents.

Consequently, the research objectives touch the following aspects: the evolution of base cartography toward spatial databases, the connection between a 3D geospatial database and a 3D city modelling, the connection between 3D city modelling and BIM, the connection between geo- reference and geo-thematic applications in the context of railways, the role of point clouds data within spatial databases, and the multi-source geospatial information management.

To summarise, the thesis focuses on outlining a road map to keep interoperability using geographical standards and formal steps. Each step runs as a liaison point between different spatial data applications. Independence from technological platforms or application formats has been one of the mandatory requirements.

Contents

1. CHAPTER 1 – INTRODUCTION, RESEARCH FRAMEWORK

1.1 Glossary and terms

3DmFV	3D modified Fisher vectors
ADE	Application Domain Extension
AI	Artificial Intelligence
ATP	Automatic Train Protection
BIM	Building Information Modelling
CRS	Coordinate Reference System
CM	Conceptual Model
CMMS	Computerized Maintenance Management System
CNN	Convolutional Neural Network
CoS	Content Specification
CS	Conceptual Schema
CTR	Carta Tecnica Regionale - alias Technical Regional Map
DB	DataBase
DBMS	DataBase Management System
DDTM	Dense Digital Terrain Model
DDSM	Dense Digital Surface Model
DEM	Digital Elevation Model
DP	Data Product
DQE	Data Quality Element
DQM	Data Quality Measures
DTP	Territorial Production Department

FME	Feature Manipulation Engine
GIS	Geographic Information System
GCP	Ground Control Point
GDF	Geographic Data File
GML	Geographic Markup Language
GMM	Gaussian Mixture Model
GNM	Generic Network Model
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSD	Ground Sample Distance
GTDB	Geo Topographic DataBase
IFC	Industry Foundation Classes
IM	Implementation Model
IMU	Intertial Measurement Unit
INSPIRE	Infrastructure for Spatial Information in the European Community
INTESAGIS	Intesa Stato Regioni, Enti locali per la realizzazione dei Database Topografici di interesse generale – alias Agreement among National, Regional, Local Authorities for topographic databases implementation
IRS	International Railway Standard
ISO/TC 211	International Standard Organisation Technical Committee 211
LAM	Linear Asset Management
LBS	Location Based Service
LIDAR	Laser Imaging Detection and Ranging
LOD	Level Of Detail
LRS	Linear Reference System
OGC	Open Geospatial Consortium
MMS	Mobile Mapping Survey
MUIF	Modello Unico Infrastruttura Ferroviaria – alias Unique Model of the Physical Infrastructure

OCL	Object Constraint Language
PS	Physical Schema
RTM	Rail Topological Model
RFI	Rete Ferroviaria italiana – alias Italian Railway Network Enterprise
RGB	Red Green Blue
SDB	Spatial database
SDBMS	Spatial DataBase Management System
SDI	Spatial Data Infrastructure
SFM	Simple Feature Model
SienaGTDB	Geo Topographic DataBase of the Siena Old Town
SRS	Spatial Reference System
SVM	Support Vector Machine
TB	Tera Bytes
UIC	Union Internationale the Chemins de fer - alias International Union of Railways
UML	Unified Modeling Language
XML	eXtensible Markup Language
WebGIS	GIS on WEB
WS	Web Service

1.2 Introduction

In the past, the representation of geospatial information has been based on planimetric cartography and elevation geometric elements (contour lines and elevation points). For this reason, traditional cartography has been shortly defined as a 2.5-dimensional map (2.5D). In the last few decades, technological developments have made it possible to create real 3D city models. In 3D city models, geometries are modelled using triple coordinates (x, y, z) where the third dimension (z) is modelled on a par with the other two planimetric dimensions (x and y). The 3D city models faithfully represent the real world of the urban context in a digital system, more completely and efficiently than traditional cartography at the same scale. Nowadays, many in different domains such as civil engineering, urban planning, hazard mapping, smart city applications 3D city models have been applied for visualization and data analysis(Zlatanova et al., 2012; Li et al., 2020).

The semantic modelling of cities requires the appropriate expression of 3D data. This can be done more and more by automated processes (Aleksandrov et al., 2019; Wendel et al., 2017; Biljecki et al., 2016b) and in some cases by manual interpretation (Kolbe, 2009).

From a geomatic point of view, the 3D datasets have recently achieved a quantitatively significant impact thanks to the technological evolution that allows to survey and manage 3D data in specific models, even if the primary efforts have been addressed to visualisation matters.

At the same time, an interesting aspect refers to the semantics of data models that manage a geospatial database as a reference for Spatial Data Infrastructure (SDI), according to standards on geomatics (such as ISO/TC 211[1], OGC[2]) and compliant with the INSPIRE European Directive[3] (Directive 2007/2/EC, 2007). Combining these 3D and SDI aspects has impacted the definition of the 3D spatial models within a Geographic Information System (GIS). For instance, processes and analyses of territorial information in a database structure could be required to achieve new environmental phenomena.

The support to decision-makers supplied by 3D city models is a part of the outputs that can be achieved once data have been natively modelled in 3D. For this reason, the semantic aspects and the generalization of modelling geospatial data are crucial requirements. About semantic aspects, it should be intended to comprise spatial characteristics including the ontological properties defined in terms of object classes or attributes or considering relationship classes among objects. Objects are decomposed into parts due to logic that follows structures that are given or observed in the real world.

---

[1] https://www.iso.org/committee/54904.html ISO/TC 211 Geographic information/Geomatics

[2] https://www.ogc.org/standards Open Geospatial Consortium Standards and Resources

[3] http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002

Thus far, in a 3D city model, the wider structure adopted to describe 3D geometries has been the Boundary Representation (B-Rep) model (Foley et al., 1995; Vebree & Zlatanova, 2004). Nevertheless, a Spatial Database Management System (SDBMS) nowadays remains the more effective best way to manage vast geospatial information over time. However, the lack of a standard process to derive a shared 3D city model, and how to manage it into an SDBMS (Gröger et al., 2004) is an open research topic, although some significant experiences have been conducted especially in urban areas (Yao et al., 2018).

1.2.1 The railway geospatial information.

Several aspects are relevant in railway geospatial modelling. Maintaining and monitoring railway systems at a national level is quite challenging: rail lines are usually thousands of kilometres long, and several objects of interest are distributed along such lines.

Given the geospatial extension of a railway network and the variety of objects to monitor, such as switches, bumper, masts, bridges, a DB structure of several objects related to the railway system is strongly recommended. The reference for modelling thematic information at the European level is represented by the INSPIRE Directive (Directive 2007/2/EC, 2007). The standard references in railways are represented by the Inspire Data Specification on Transport Networks (INSPIRE D2.8.I.7, 2014). This standard includes an integrated transport network and related features that are seamless within each national border. Moreover, transportation data includes topographic features related to road, rail, water, and air. The spatial location is defined in a conceptual schema based on a Linear Reference System where objects are georeferenced considering a one-dimensional measurement along (sometimes with a displacement from) that net. The spatial model describes the data and operations that need to use and support a Linear Reference System (LRS) (ISO/TC 211 19148, 2012). This kind of reference system is particularly useful to define objects' locations in thematic fields such as transportation networks, location-based services, and other applications based on a graph structure. On this standard, a 2D vector represents the basic element. Each domain data model represents an abstraction of a specific application of rail networks in the real world. Such application is not easily interchangeable with other rail networks based on different data models; thus, no shared model/databases are provided.

1.2.2 Model and management of multisource surveyed data

From a surveying point of view, given the considerable rail line length and the complexity of railway infrastructures, data acquisition is often carried out by using different approaches and sensors. In general, the recent development of new geospatial data acquisition tools and the consequent availability of multi-source spatial information (e.g. remote sensing, drones, laser scanners) have significantly influenced the related geodata models and structures.

A large variety of data structures are currently used, mostly depending on the acquisition systems that generated such data. Indeed, each sensor (and system) provides its outcomes according to a specific data structure: point cloud, raster data, transportation network, optical image. Hence, when dealing with a multi-source dataset, such different data structures must be integrated into a unique spatial representation, properly taking into account their peculiarities (Tucci et al., 2020) to properly manage all the available geospatial information in a unique framework.

New sensors can quickly acquire vast amounts of data. For instance, new laser scanners can easily acquire millions of 3D points per second, leading to the generation of datasets composed of billions or even trillions of points, e.g. in monitoring applications, where the need for detecting any timely variation can be achieved only by repeatedly scanning the same area. On the one hand, this approach promotes to implement monitoring procedures to detect unexpected events at high spatial and temporal resolutions, however, on the other hand, this opportunity drives the need for developing efficient ways for storing and processing such massive amount of data. Indeed, such Big Data hardly are handled efficiently (Govardanan & Gnanapandithan, 2020).

Multisource surveying data need to be organising in a conceptual geo-model design first, then it should be profiled in a specific application schema. Afterwards, the implementation of a spatial database should be carried out. For this reason, the implementation of an integrated Spatial Database Management System following interoperable rules could help to overcome the heterogeneity of current geospatial geometric and topological data (Li et al., 2020).

1.3 Research problems

Based on the general context previously explained, some research problems across the PhD topic should be considered crucial questions to achieve research objectives and take better choices in the roadmap.

1.3.1 Interoperability and involved standards

Interoperability is the fundamental key to obtain integration. The interoperability concept can be expressed at different levels: technical, syntactic, semantic (Fig. 1). These can be associated with different level of integration of geographical datasets: conceptual, logical, and physical levels.

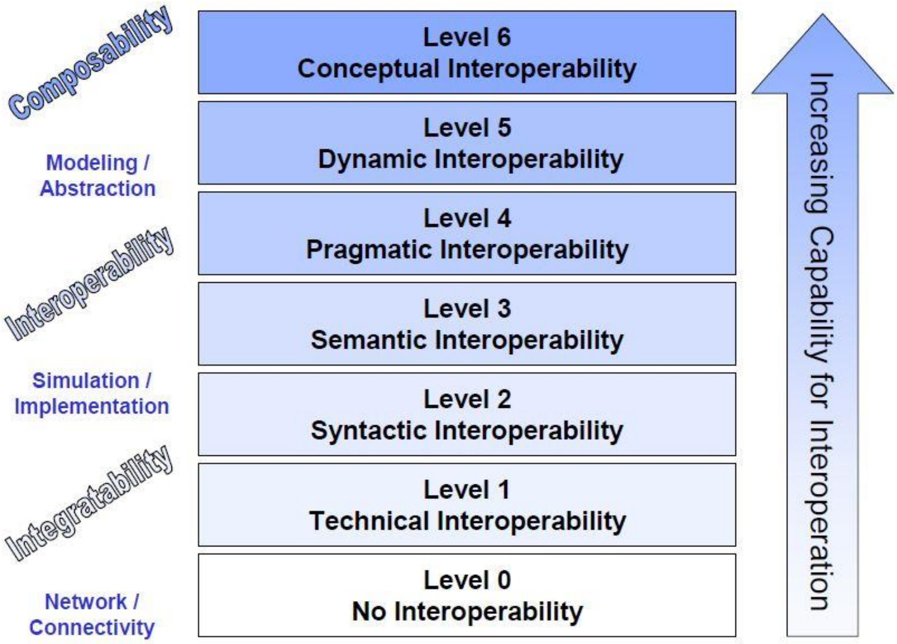


Figure 1 - levels of interoperability (Tolk et al., 2006)

The higher the level of interoperability, the greater the integration capability, but data from different domains easier belong to different conceptual models and meanings. In turn, according to specific requirements, standards on geographic information have been developed, such as abstract modelling, formal languages, semantic aspects.

Standardisation universe includes a huge variety of standards, both de jure and the facto. According to technological evolution, some of them change quickly over time, while others can be considered milestones in the standardisation processes. Some of them have been adopted in national regulations; others emerged mainly guided by industrial needs.

A growing number of standards such as CityGML (OGC CityGML, 2012) and InfraGML (OGC InfraGML, 2017) defines geospatial implemented models in dependent-platform solutions. The interoperability choice should take care of content composability at the conceptual level, according to abstract modelling of ISO TC/211 19xxx standards[4], while at the thematic level there exist mainly standards adopting specific implementation models according to OGC standards[5]. The primary approach should focus on the use of open standards, platform-independent, enabling longer-term support, but their genericity makes not immediately point-to-point solutions at physical levels (Noardo et al., 2020). Therefore, in the transition from an abstract to a physical model of geospatial datasets, attention must be paid to interoperability choices.

---

[4] https://www.iso.org/committee/54904.html

[5] https://www.ogc.org/standards

1.3.2 The evolution of reference geographic information between databases and 3D

The technological evolution has made available a significant increase of source datasets based on 3D natively. Even if for new acquisitions 3D seems to be feasible and advantageous, from the practical point of view, GIS topological operators, as well as existing datasets, are mostly based on a 2.5 D modelling. Despite a full 3D representation of geospatial objects in 3D City modelling, in geotopographic field, the elevation (third dimension) is still widely described as an attribute associated with each spatial object. So, in a 2.5D system, it's possible to associate one elevation to each feature geometry, which for geometries other than points (lines, polygons, etc.) doesn't allow B-Rep modelling of each object, because of the missed information at each vertex.

It should be also considered that in common GIS, spatial topology tools provide two-dimensional results according to Clementini-Egenhofer's approach (Egenhofer et al., 1994, Clementini et al., 1996). Hence geospatial analysis takes advantage of using the third information as an attribute rather than an effective spatial dimension.

Consequently, large scale reference cartography evolution is progressing both with contents organisation in databases (transition from maps to GIS) and with primary acquisition of 3D city models (better data organisation in urban areas). GML standard (OGC GML, 2016) refers to the former case (database), while the CityGML standard (OGC CityGML, 2012) refers to the latter (3D City Model). Recent experiences integrate the CityGML standard and SDBMS in 3DCityDB (Yao et al., 2018) but in general 3DCity models and SDBMS are designed following different requirements that do not make them automatically interoperable.

1.3.3 BIM and GIS connection

Each built spatial object can be described considering geolocation and relationships with surrounding territorial objects following a GIS point of view. Otherwise, objects could be described by their parametrical parts, which are related to functional or life cycle aspects, following a BIM point of view. GIS and BIM allow to define the same geospatial context but following different point of view. Recently, the standardisation organisations are focusing on multi-disciplinary open standard development to integrate concepts from different domains (ISO ISO/TC 211 19166, 2018).

However significant differences characterise the two approaches (BIM and GIS). Moreover, focusing on GIS aspects, it should be considered that problems like loss of information, improper conversion, loss of relationships, and topological inconsistencies arise while converting and combining 3D city models from different formats (Aleksandrov et al., 2019). For instance, the CityGML standard is the most widespread standard for 3D City modelling but follows a prominent perspective in buildings. Otherwise LandInfra standard (OGC LandInfra, 2016) considers infrastructures in 3D GIS so it is better appropriate to describe transportation infrastructures such as the railway. In turn, also the BIM approach predominantly focusing on buildings, is going toward infrastructure parametric content description, not wholly implemented in its most popular open interchange format, namely the Industry Foundation Classes (IFC) (ISO/TC 59/SC 13, 2018). To summarise, the BIM-GIS connection is an early stage of research analysis, characterised by the lack of standardised interchange processes.

1.3.4 Reference GIS Vs Thematic GIS of Railways

The geographic information in the context of railways can be described following two essential points of view: the first one aims to implement services and application considering railways as one of the main network transportation systems, the other one is oriented to the management of infrastructures in a vast territory, requiring for that to organise asset information in a database structure. These two aspects refer to different ways to describe “the universe of discourse” that refer to different modelling types of the real world. In reference GIS traditional cartography geotopographic objects are organised in spatial databases where topological relationships are based on spatial reference systems while transportation networks are modelled according to graph topology where geolocation is based on Linear Reference Systems (LRS). Spatial and network topology are not compatible with the same spatial data model, so it is necessary to describe the same dataset in different data models and interrelating dependencies between them in an ad-hoc solution.

1.3.5 Integrated SDBMS

When an SDBMS is implemented for the first time, traditional sources commonly have been considered to completely survey the entire content. Accordingly, continuous updates of such SDBMS must be managed over time. Often, thanks to its quick acquisition, new data come from a laser scanner source. Such data source types offer to acquire natively more information apart from metric ones (as it happens for traditional sources in survey data). Hence, there is an implicit greater interest to include 3D point clouds as content in an SDBMS beyond the fully operational process to classify, segment, and relate point clouds. However, there is no standardised way to connect source data such as point clouds into an SDBMS where basic object classes are described through vector geospatial components.

1.4 Objectives

The interoperability is the leitmotif of the research topic, analysed through different aspects and point of view. The interoperability term must be defined at different levels, to fully consider the design of complex spatial infrastructures: sematic, schematic, syntax, and, above all, on processes and steps required to be shared in a common framework. Taking care of the general interoperability objective, the thesis focuses on the following relevant aspects:

1. The 3D city modelling in SDBMS.
2. The BIM-GIS connection
3. The link between a reference geoinformation and geospatial thematic applications, considering railway infrastructure context.
4. The management of multi-sources survey data in terms of SDBMS.

In particular, the thesis aims to structure a 3D GeoTopographic Database (GTDB) addressed not only to visualization but toward the geographical infrastructure data management above all. Nevertheless, the evolution toward a 3D city model in an SDBMS for the INSPIRE-compliant interchange data must be considered.

The research addresses to devise an interoperable (to solve problem 1 of 1.3.2. paragraph) process prototype of a 3D data model able to describe:

A. the 3D GTDB derived from a cartography map, implemented in an SDBMS to solve problem 2 (1.3.2. paragraph). The application of the 3D city model derivation strategy from geo-topographical cartography (footprint and roof map) will be considered.

B. the 3D spatial model for different territorial applications and analysis to solve problem 3 (1.3.3. paragraph). 3D City model and the BIM connection related both to buildings and infrastructures will be considered.

C. the case of railway infrastructure content to solve problem 4 (1.3.4. paragraph). The interoperability of data models based on 3D spatial topology and network topology will be considered.

D. the updates of the SDBMS over time to solve problem 5 (1.3.5. paragraph). Frequently laser scanners supply source data and a massive new data batch needs to be integrated. An intermediate step that automatically classifies and segment point clouds will be implemented. Hence it will be possible to relate new datasets to each geographical object. This process could be the evolution step from a repository toward a manageable geographical database over time.

To summarise, this thesis focuses on outlining a road map to keep interoperability using geographical standards and formal steps.

For each step, the independence from technological platforms or application formats will be investigated to gain conceptual interoperability (problem 1, 1.3.1 paragraph).

The general target will research a procedure to model integrated spatial information, rather than a single-step solution, defining data models as a connection point between different standards, applications, and thematic point of view.

The starting point of the research road map is represented by the reference base cartography (the 3D GTDB) integrated with the railway asset.

The ending point is the definition of an integrated 3D SDBMS as a reference base to be continuously updated in near real-time. Consequently, the research steps touch in the road map (Fig. 2) the following aspects:

- the evolution of base cartography to a GTDB;
- the connection between a 3D Spatial Database and a 3D city modelling;
- the connection between 3D city modelling and Building Information Modelling (BIM);
- the connection between a geo reference base and a geo thematic application (considering geographic information in railway infrastructures);
- the contribution of point clouds data in the integrated SDBMS;
  - from the survey to the 3D SDBMS: management, multi-sources, and continuous updates.

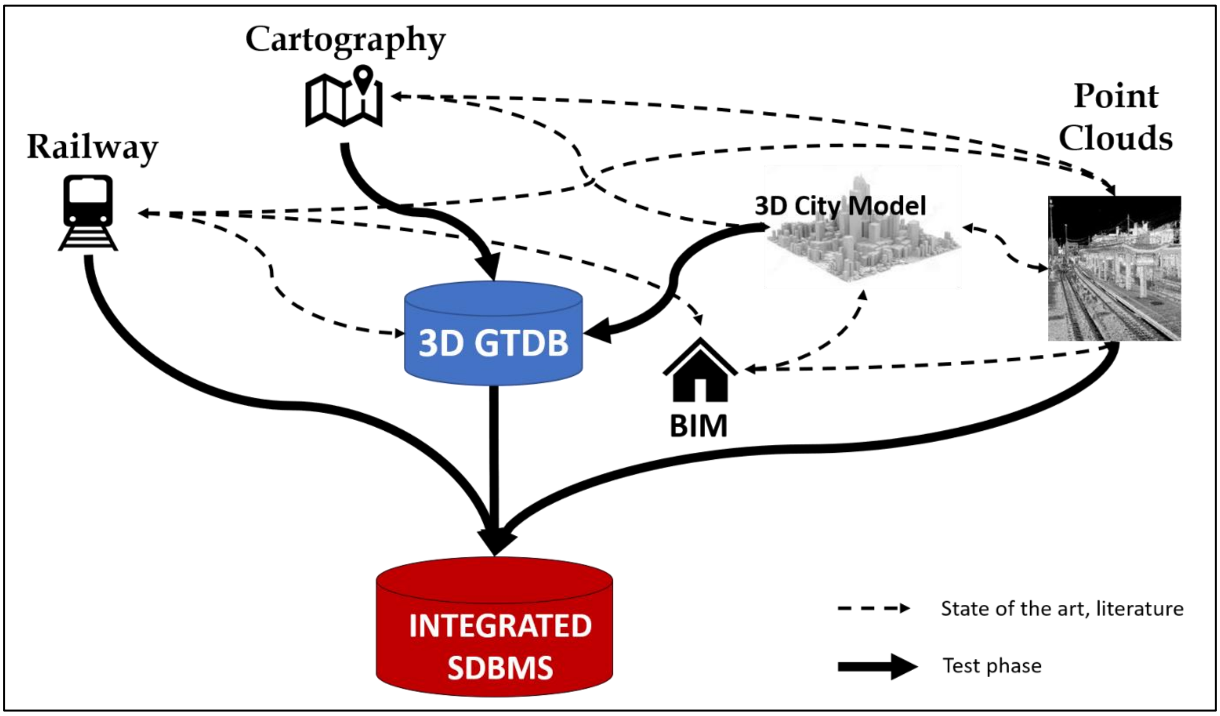


Figure 2 - the road map methodology of the research. Continuous arrows: the 3D GeoTopographic Database (3D GTDB) is obtained considering the evolution of reference cartography toward 3D city modelling, then an SDBMS has been designed to include railway about contents and point clouds about additional spatial components. Dotted arrows: all the bidirectional connections have been considered in the state of art.

1.5 Research questions

Based on the above research problems and objectives some questions have to be solved during the research road map.

About interoperability (leitmotif research and general Objective):

1. *For any step of the roadmap, what compliancy must be considered concerning the implementation level, type of modelling, and standards to guarantee the best increase of interoperability?*
2. *What is the better level of abstraction of the present research case study to choose the congruency with standards?*

About 2D-3D transaction (Objective A):

1. *How to combine geographic information in the transition phase from 2.5 and 3D?*
2. *How to combine implemented topological operators together with visual analysis?*

About BIM-GIS connection (Objective 2):

1. *Which tools or connecting models are best suited to manage BIM-GIS interrelationships focusing on the infrastructure topic?*

About Reference vs Thematic (Railway) GIS application (Objective 3):

1. *Does a shared and interoperable point of view exist?*
2. *What are the interactions between topographic objects and railway objects?*

Integrated SDBMS overtime (Objective 3, Objective D):

1. *How to manage multi-sources acquisition phases?*
2. *How to manage point clouds within an SDBMS?*
3. *Which validation will be implemented to maintain quality parameters?*

1.6 Outline

The thesis is organised into nine Chapters, in some of them the specific declaration of PhD contribution is declared.

1. CHAPTER 1 – INTRODUCTION, RESEARCH FRAMEWORK

This chapter is a general overview of the state of the art, the open issues, and the current debate and the objectives on research topics. Interoperability is the main task considered in the context of geographic information systems. Requirements and questions here expressed are then developed across the next chapters and resumed in chapter 9 as conclusions.

2. CHAPTER 2 - FROM THE CARTOGRAPHY TO THE GTDB IN ITALY

The chapter describes the state of the art in Italy about the organization of cartographic contents in GeoTopographic databases, according to national laws and based on a spatial conceptual model formalised in the so-called GeoUML methodology. The detailed explanation of this part is due to previous experience in editing such contents in the geotopographic Italian specification working group (IntesaGIS project, Corongiu et al., 2004). This chapter anticipates some of the requirements to obtain interoperability applied to the specific case study (Chapter 5) about rules, relationships and constraints to define in a conceptual geospatial model by a formal language (UML), and at the same time, maintaining compliancy with the Italian national Specifications.

3. CHAPTER 3 - GIS (GTDB) TOWARDS 3D CITY MODELS

This chapter defines the evolution of spatial models, in terms of all the dimensions that define geometrical coordinates (3D) and time dimension (4D). In a spatial database, the 3rd dimension is addressed for visualisation tasks for the 3D city models and as a fundamental component of geolocation and description of topological constraints in a Geographic Information System. As the previous one, this chapter anticipates some requirements on 3D aspects to be guaranteed in the specific use case (chapter 5). Particularly, this part explains which 3D model could be adopted according to the GeoUML methodology (chapter 2) as an intermediate model toward a full B-rep (Foley et al., 1995) of 3d city models. The analysis of the 3D models has been developed using the case of the Siena geotopographic database as a term of comparison with the spatial model adopted in the railway case study. The technical specifications and the supervision of works of Siena database example were carried out by me before these PhD studies.

4. CHAPTER 4 – THE BIM-GIS CONNECTION

This chapter compares 3D city model contents and the BIM classification approach. Semantic in object classifications and spatial/georeferencing aspects have been linked from a conceptual perspective, avoiding any physical transformation. Some futures scenarios have been proposed according to geographic standards. The analysis focuses on a vice-versa approach: from GIS to BIM because the research topic starts from a cartography reference base point of view. On the contrary, both literature and standards mainly consider the unidirectional approach, from BIM to GIS. The proposed approach is oriented to link different data models instead of transforming them from one format to another. Moreover, the connection GIS-BIM is analysed in terms of an intersection metamodel design.

5. CHAPTER 5 – THE CASE STUDY

The case study refers to datasets of an Italian project called MUIF (Unique Model of the Physical Infrastructure), aiming at generating a georeferenced spatial digital representation of the Italian railway system. Starting from MUIF datasets many characteristics of interoperability issues have been considered: territorial extension of data, multi-sources multi-accuracies of geographic information, compliancy with standards-national regulation, 2D-3D, BIM-GIS connection, and finally management in the integrated SDBMS. Profiling of technical specifications according to the national one (MD, 2012), modelling geospatial classes attributes integrated with railway assets are the core of my PhD contribution. This Chapter mainly refers to (Corongiu et al., 2018) scientific paper.

6. CHAPTER 6 – GEO-REFERENCE VS. GEO-THEMATIC INFORMATION SYSTEMS.

Interoperability also means to integrate different models to describe territorial aspects or phenomena in a unique geospatial context. This chapter focuses on the design of an integrated model between topographic reference information, namely the GTDB, together with thematic applicative geospatial information, namely the Railway geospatial infrastructure. From a technical point of view, a specific analysis has been developed to consider the connection between Graph Topology vs Spatial Topology. Hence about geospatial location, the first one is typical in transportation networks and refers to the Linear Reference System (LRS) while the second one, commonly used in GTDB, refers to Coordinate Reference System (CRS).

7. CHAPTER 7 - POINT CLOUDS IN SDBMS.

A physical experimental phase has been implemented to demonstrate how to manage the point cloud dataset, not only as a source of survey campaign but, above all, as geometrical components in a spatial database management system. Indeed, point clouds have not been considered as-a-whole, as it happens in the case of orthophotos, but segmenting point clouds and relating them with asset features in SDBMS. Moreover, the spatial intersection between features and point clouds has been used as ground truth in a deep learning approach addressed to automatically classify point clouds coming from the new updates of data. The workflow of this procedure and the implementation has been performed thanks to the contribution of my colleague Prof. Andrea Masiero (University of Florence) and formalized in (Corongiu et al., 2000) scientific paper. The integration of such aspects into this modelling thesis is only aimed to demonstrate the feasibility of adopting the train MMS as a cyclical source to update the DB, as explained in Chapter 8.

8. CHAPTER 8 - AN INTEGRATED SPATIAL MODEL FOR HETEROGENEOUS DATA: MANAGEMENT, UPDATE AND VALIDATION

All the aspects analysed in previous chapters became here requirements to guarantee interoperability and updates over time. The following characteristics have been thoroughly considered: maintenance processes such as continuous updates and multi-source validation processes. These aspects have been carried out using compliance with standards. All the validation processes and DB model integration aspects have been designed as my PhD contribution. Then I’ve personally tested data about the GTDB part, while the rest of the MUIF case study validation tests have been applied with the contribution of the University of Florence Schema Lab (geomatic Lab guided by Prof. Grazia Tucci) which is in charge to validate overall MUIF cartographic products. This chapter is mainly based on (Corongiu et al., 2018) and (Tucci et al., 2019) scientific papers.

9. CHAPTER 9 – CONCLUSIONS

Finally, this chapter recaps the results and critical aspects of this research. Furthermore, it focuses on future work and perspectives with the awareness that some of the proposed choices and solutions need to be tested over time or optimised according to technological evolution.

2. CHAPTER 2 – FROM THE CARTOGRAPHY TO THE GTDB IN ITALY

2.1 Technical cartography vs spatial databases

The reference cartography is traditionally drawn up by a specific institutional entity.

The main scope refers to a carryout reference information for the territory described as-is without thematic photointerpretation. In Italy, reference cartography provided by Regional Authorities as technical Regional Maps (CTR alias Carta Tecnica Regionale) represents the standard reference map at larger scales (1:2.000, 1:5.000, 1:10.000). To manage infrastructure like bridges, railways, streets, etc., private enterprises or public bodies define their requirements to obtain spatial data accurately, even covering large areas across several regions.

For this reason, there is a problem of harmonisation between CTRs with different levels of detail in time and spatial accuracy.

The reference content structure has been evolving over the years, passing in a basic map to complex databases of 3D geometries few decades combining topological constraints, interrelationships, numerical models, etc.

In Italy, the evolution from cartography toward databases start with the *"Intesa Stato Regioni, Enti locali: Specifiche per la realizzazione dei Database Topografici di interesse generale"* (alias: Agreement among National, Regional, Local Authorities for topographic databases implementation) project, summarised hereafter IntesaGIS. The project was set up to define the general structure and contents of a national reference GeoTopographic DataBase (GTDB), a National Core (Corongiu et al., 2004). IntesaGIS originated in 1996 following an agreement among all the state cartographic offices such as the IGM, (the Italian Military Geographic Institute acting as National Mapping Agency), the Agenzia del Territorio, which runs the Cadastre, and regional and local authorities such as Regions, Provinces and Municipalities. The Italian large and medium scale cartography is mainly delivered by these national Government and Local Authorities. The Conferenza Stato Regioni (namely the permanent organization between national and local Government) approved the IntesaGIS agreement to define geotopographic database specifications. , the Ministry of the Environment provided most of the funding.

The reference guidelines written in that project addressed the following (Fig. 3):

- organise data in a DB oriented structure;
- share data among:
    - all level of the Public Administration;
    - different applications;
- manage data with different precisions and level of details;
- derive the DB25 (1:25000 national DB of IGM) from databases at a larger scale;
- carry out the cartographic representation from the database information;
- be compliant with international standards and national laws;
- design adequate metadata information;
- pay attention to the evolution from 2D to 3D.

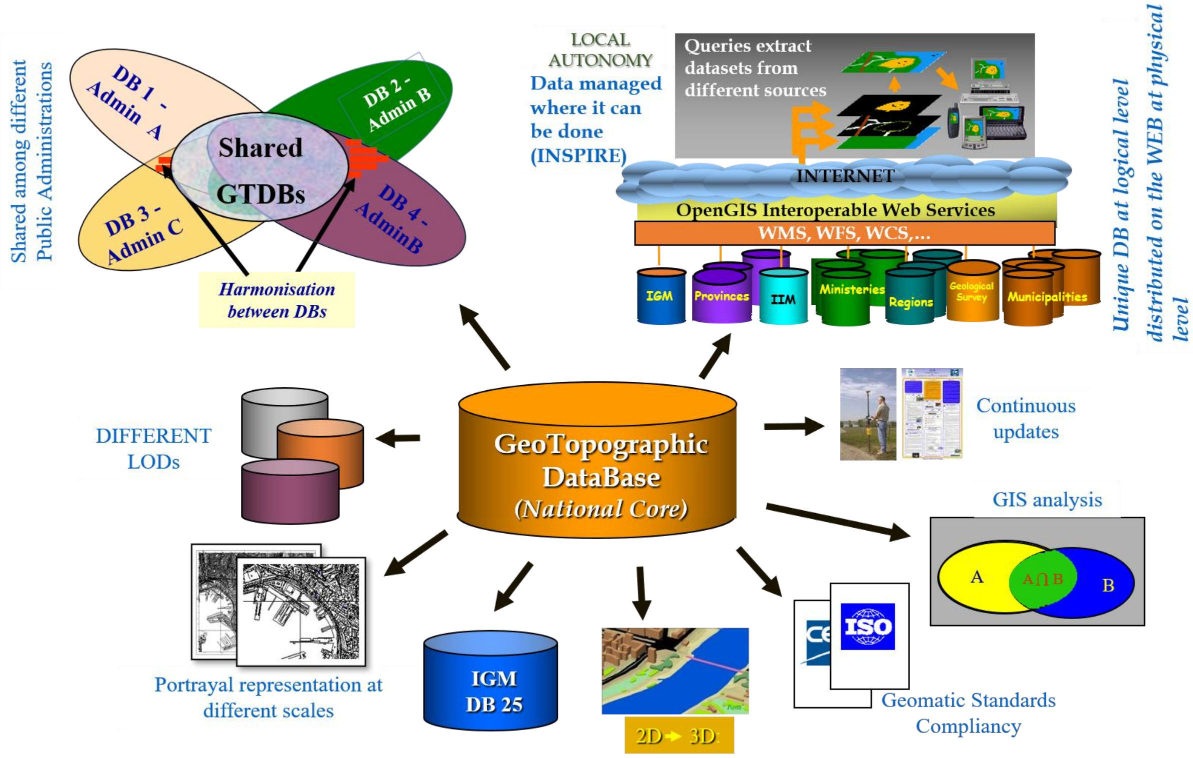


Figure 3 - IntesaGIS project requirements: the GTDB is shared among different public administrations throughout WEB Services, but maintaining local Autonomy (Inspire Directive principle), it must be continuously updated, GIS oriented and taking into account the 2D-3D transaction, standard-compliant, able to manage different LOD, allowing to deliver IGM DB25 and portrayal maps.

2.1.1 The data acquisition: sensors and processes evolution

Geospatial contents coming from the interaction among different sensors are integrated thanks to embedded software procedures, not explicitly outlined in the final products. This means that such intermediate aspects aren’t ignorable from the quality certification point of view, even if becoming implicitly in the final products (Tucci et al., 2020).

New issues in quality certification activities require starting the analysis from the cartographic product evolution. Even as far back as the end of the last century, cartographic products were made on a simple structure based on consolidated mapping rules. About The topographic network worth noticing that Laser sensors were not widely used, and the GNSS (Global Navigation Satellite System) receivers requiring a GPS constellation availability must be integrated with a master-rover system to complete the topographic network. Moreover, the aerial triangulation phase could not always be replaced by inertial systems. It could say that they had achieved an excellent quality standard about photogrammetric cameras when equipped with some advanced tools such as sophisticated dragging. The brand could only distinguish Distortion-free optics smoothing and stabilisation systems. Almost identical cameras because they were based on the same standard. Consequently, simple and standardised rules characterise camera processes. Finally, the final product represented by the Numerical Cartography could be validated considering the specific accuracy related to the level of detail of contents. To summarising, phases necessary to achieve the final product are generally based on static steps as follows:

1. Flight planning, according to required scale cartography;
2. topographic network design survey area depending on;
3. Aerial triangulation;
4. Acquisition;
5. Editing;
6. Reconnaissance.

Specific validation processes related to constraints and thresholds are significantly consolidated for each step. So a progressive certification of each step starts once the previous one is done, guaranteeing the quality of the final product. For example, it is assumed that the topographical network of stereo-models is corrected when triangulation evaluation based on new surveyed points has been done. Even though it appears too expensive, this methodology allows to absolutely certify the quality of the results, given that a triangulation calculation complexity makes difficult error propagation evaluation. Therefore, anomalies can be highlighted only by measurement comparison as a deductible result in calculating dense mesh points arranged according to orography. Then, thematic (classification) accuracy, geometrical accuracy, and content completeness must be validated in the acquisition step.

Nowadays, where cartography contents have been organised in spatial databases, data models must be robust enough to withstand the technological evolution, meantime flexible sufficient to be updated over time.

2.2 The GTDB and the GEOUML Methodology

In the last decades, a progressive transition between technical cartography toward spatial databases has been carried out. From a standardisation point of view, at the governmental level, several authorities have defined their own standard data model. In Italy, the traditional base cartography contents have been organized in a GeoTopographic Database (GTDB).

The GTDB Italian Specs were defined in 2012 according to the "Decreto sulle regole tecniche per la definizione delle specifiche di contenuto dei database geotopografici" ("Decree on the Technical Rules for the definition of the Content Specifications for the Geo-Topographical databases") in a decree of the "Public Administration and the Environment" ministry (MD, 2012).

The decree defines the so-called "National Core", which is the set of geo-topographical data that needs to be shared at the national level (Corongiu et al., 2018). Indeed, a GTDB shared core has been specified to support different applicative or thematic maps among all concerned government administrations.

The GTDB is oriented to the creation of the Spatial Data Infrastructure (SDI) according to the INSPIRE Directive (Directive 2007/2/EC, 2007). It simplifies the implementation of databases that share common contents and support the definition of datasets suitable for the exchange of those contents. Specifically, some reference principles of the INSPIRE have been taken into account: data have to be collected only once, and managed where it can be done more efficiently; it should be feasible to combine data coming from different sources and to share them among several users and applications; it should be feasible to share information that has been collected at different levels of details; spatial data that are necessary for the territory government should exist and be widely accessible.

GTDB Contents and Conceptual Model (CM) have been described through the GeoUML Methodology and the GeoUML Tools of the Spatial DB Group[6]. The GeoUML Methodology is about the conceptual schema for DGTB, where some specific software, namely GeoUML Tools, could support the profiling of a geographical Conceptual Schema (CS) and the conformity of a Data Product (DP) related to this Conceptual Schema (CS) (SpatialDBGroup, 2011). The fundamental principles of this development have been:

- the ISO/TC 211 Geographic Information/Geomatics standards (https://www.iso.org/committee/54904.html) compliancy;
- to be implementable on current technology;

---

[6] http://geo.spatialdbgroup.polimi.it/en/

- to be independent of any specific (commercially or open) GIS product;
- to keep a clear separation between the conceptual and the implementation levels.

The GeoUML Catalogue is a tool able to defines geospatial contents according to a GeoUML model.

The GeoUML Validator is a tool used for checking if a Data Product (DP) is conformant to a specific Conceptual Schema (CS) is called.

Specification File is delivered from specifying a Conceptual Schema (CS) for different catalogues or toward the Validator tool.

The Implementation Models (IM) can be chosen to start from a Data Product (DP), translating CS into a physical structure and validating them with the GeoUML validator module.

The whole GeoUML Tools are developed in Java programming language[7].

To summarise, GeoUML Methodology is supported by GeoUML Tools.

Generalising, the GeoUML Methodology (and the GeoUML Tools) can be adopted to design any geospatial database. Topological properties and relationship among object classes could be defined as conceptual contents of any geospatial DB (Pelagatti et al., 2009). Moreover, it will be possible to derive from the conceptual specification the corresponding physical structure, namely the Implementation Model. According to a unique CS, different delivery files could be supplied, such as Shapefile[8] or GML[9] formats, to exchange data format to users or stakeholders distribution. Finally, the conformity between Data Product (DP) and related specifications could be verified.

A particularly useful application about the GeoUML Methodology could be addressed to a consistent Spatial Data Infrastructure (SDI) definition. Indeed, above in the Italian context, Local Authorities independently manage different Geospatial DBs, representing, on the whole, an integrated model of a global territory (Belussi et al., 2006). Each Local Authority adopts a specific platform for physically storing its DB. Therefore, common data contents and global spatial constraints could be conceptually shared so that databases implemented with different IMs could be checked according to common rules.

Moreover, other characteristics of the GeoUML approach are highlighted:

- clear separation between Conceptual and Physical Levels, so specification and implementation are different levels;

---

[7] https://docs.oracle.com/javase/7/docs/technotes/guides/language

8 https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

9 https://www.ogc.org/standards/gml/

- many supported IMs, mapping rules from the same CS on different PSs could be provided;
- different supported integrity constraints (including spatial ones), allowing automatic data validation at the physical level.

Finally, while the GeoUML Catalogue supports the definition of the Conceptual Schema CS in terms of Specifications, also suggesting parameters that are necessary for the generation of a Product Specification (PS) according to the chosen IM, the GeoUML Validator performs the conformity check of a DB according to Specifications and CS designed into the GeoUML Catalogue. (Tucci et al., 2020).

2.2.1 Content Specifications (CoS) and Model Implementation (IM)

A Content Specification (CoS) describes the informative content of a DB. Different kinds of definitions, depending on other purposes, could be used. Informative Elements are the basic objects to be represented in the DB (classes, alphanumeric and geometric attributes, relationships and associations, domains). Then the definition of integrity constraints, i.e. the intrinsic properties that the informative elements must meet. Intrinsic properties refer to the properties that are verifiable on the informative elements themselves, without directly looking at the real world (Belussi et al. 2011). Descriptive elements represent pieces of information to understand how to interpret DB content in a human language and describing the real world.

A formal Content Specification (CoS) is composed both by informative elements and integrity constraints.

A Conceptual Schema (CS) defines the properties that a Data Product (DP) must have at the conceptual level, which is independent of the technology that has been chosen to implement it. A set of properly defined rules allow to automatically derive the Physical Schema (PS) corresponding to a defined Conceptual Schema (CS). This set of rules is called the Implementation Model (IM). The Physical Schema (PS) defines the physical data structure of the Data Product (DP) in a specific technology (Shapefiles, SQL DB, GML, etc.).

The main motivation for separating the Conceptual Schema (CS) from the Implementation Modeling (IM) is the possibility to define different IMs and to use them to generate several Physical Schemas (PS) starting from the same Conceptual Schemas (CS). As a result, a conceptual description is independent of the technological changes.

2.2.2 The GeoUML model components

a set of formally defined constructs characterise both Conceptual Schema (CS) and Content Specification (CoS). Constructs are of two categories:

- Structural elements: required to define the data structures used for the content representation;
- Integrity Constraints: required to define properties that data must satisfy according to a consistent Data Product (DP).

Structural elements are:

- Class;
- Attribute (non-geometric);
- Cardinality;
- Enumeration;
- Hierarchical enumerations;
- Association;
- Inheritance;
- Geometric attribute;
- Attribute of geometric attribute;
- Primary key;
- Topological layer.

These structural elements properly combined allow to completely define geospatial objects and their relationships formally. Nevertheless, extensively explanation is out of the scope of this thesis. More detailed information on GeoUML model and tools could be found on the SpatialDBGroup website[10] (SpatialDBGroup, 2011).

---

[10] https://spatialdbgroup.polimi.it/

2.3 The spatial model and THE ISO standards

The main components of the GeoUML model are composed of a set of geometric types. Geometric types are used to specify spatial components of the classes (spatial attributes). Spatial integrity constraints have been defined to specify constraints between spatial attributes and object class, using Object Constraint Language (OCL[11]) Templates. The Italian National Core specs adopt the GeoUML model as the structural part to define a specific CS. Similarly, the CoS of the Italian National Core defines the reference content specification for the creation of GTDBs at different levels of public administrations. CoS must be contained in a Data Product (DP). The DP refers to the definition presented in the ISO standard (ISO TC/211 19131, 2018). Moreover, the GeoUML is a specialisation, formally named as *profile*, of the ISO standards:

- ISO 19103 Conceptual schema language (ISO/TC 211 19103, 2015).
- ISO 19109 Rules for application schema (ISO/TC 211 19109, 2019)
- ISO 19107 Spatial schema (ISO/TC 211 19107, 2002)

In turn, these standards use UML v1.3 (Unified Modeling Language)[12] and OCL languages. Congruently, the GeoUML mapping rules produces UML schemas compliant with the above-cited standards.

In particular, the GeoUML spatial schema is a profile of the ISO Spatial Schema (ISO/TC 211 19107, 2002).

The geometric model is expressed in terms of geometric attributes of a GeoUML class.

The geometric types allow the definition of two categories of geometric objects:

- primitives: elementary geometries that could not be subdivided into parts could connect and homogeneous elements of the reference space (e.g., a surface).
- geometry collections: set of geometric primitives, homogeneous (multi-points, multi-curves or multi-surfaces) or heterogeneous (geometry collection) types; sometimes, spatial integrity constraints could be required among collection components.

GeoUML geometric types are defined by UML classes that, in turn, could be hierarchically related (Fig. 4). GU_Object is the higher hierarchical abstract class (abstract type) and cannot be used as an attribute domain. Common properties are illustrated in the root class of the hierarchy, allowing an incrementing description of them. A specific coordinate reference

---

[11] https://www.omg.org/spec/OCL/2.4/PDF
[12] https://www.omg.org/spec/UML/1.3/About-UML/

system of each GeoUML spatial attribute must be defined. According to this feature, GeoUML types are classified into two categories:

- GeoUML types describing geometries without the third coordinate (Z), called 2D types. The GU_Object2D type is the root of the sub-tree in the type-hierarchy of Fig.4;
- types describing geometries in the 3D space called 3D types. The GU_Object3D type is the root of the sub-tree in the type-hierarchy of Fig. 4.

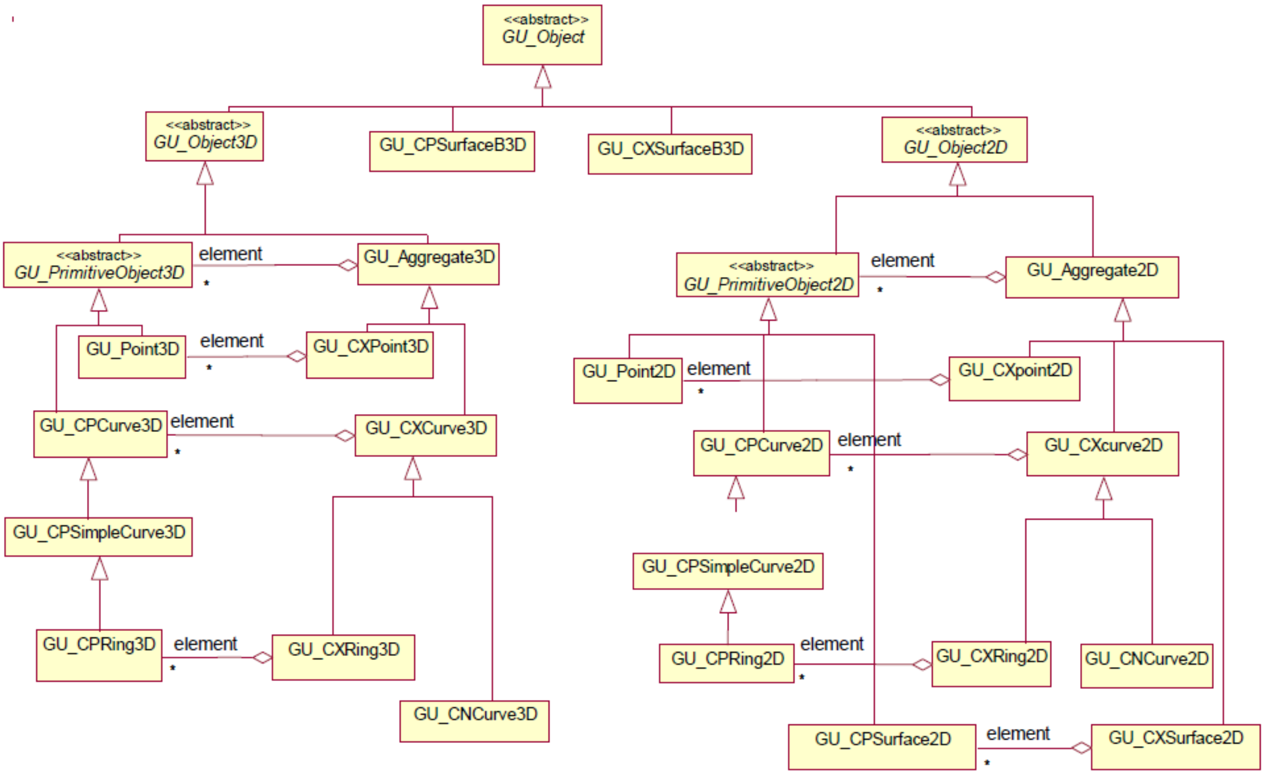


Figure 4 - UML hierarchy of the classes representing the available geometric types of the GeoUML

Summarising, the GeoUML geometric model is obtained profiling the following aspects of the Spatial Schema standard (ISO/TC 211 19107, 2002):

- 3D geometries representing points and curves compliant with operators and topological relations to be tested (the test is performed in 3D space);

- extending to the 3D space the boundary of the 2D surfaces by introducing the type “surface with 3D boundary”;
- specializing some geometric types as new ones for representing curves.

It is worth noting that the GeoUML geometric model doesn’t consider the physical level of geometry representation, such as those where interpolation methods are required. Just to give an example, the GU_CPCurve type refers to a curve in the Euclidian space at the conceptual level defined as a spatial attribute of a GeoUML class. Considering physical implementation according to the Simple Feature Model standard (ISO/TC 211 19125, 2004), it refers to a Linestring type, where a curve geometry is made by segment concatenation.

To recap, the correspondence between the Spatial Schema in GeoUML spatial classes is shown in Table 1.

GeoUML	description	Spatial Schema
GU_Point2D	Point 2D	GM_Point
GU_Point3D	Point 3D	GM_Point
GU_CPCurve2D	Composite Line 2D	GM_CompositeCurve
GU_CPCurve3D	Composite Line 3D	GM_CompositeCurve
GU_CPRing2D	Ring 2D	GM_CompositeCurve
GU_CPRing3D	Ring 3D	GM_CompositeCurve
GU_CPSurface2D	Composite Surface 2D	GM_CompositeSurface
GU_CNCurve2D	Connected curve 2D	GM_Complex
GU_CNCurve3D	Connected curve 3D	GM_Complex
GU_CXCurve2D	Complex curve 2D	GM_Complex
GU_CXCurve3D	Complex curve 3D	GM_Complex
GU_CXRing2D	Complex Ring 2D	GM_Complex
GU_CXRing3D	Complex Ring 3D	GM_Complex
GU_CXSurface2D	Complex Surface 2D	GM_Complex
GU_Complex2D	Complex 2D	GM_Complex
GU_Complex3D	Complex 3D (only Points, and Lines)	GM_Complex
GU_Aggregate2D	Aggregate 2D	GM_Aggregate
GU_Aggregate3D	Aggregate 3D	GM_Aggregate
GU_MPoint2D	Set of Points 2D	GM_MultiPoint
GU_MPoint3D	Set of Points 3D	GM_MultiPoint

GU_MCurve2D	Set of Lines 2D	GM_MultiCurve
GU_MCurve3D	Set of Lines 3D	GM_MultiCurve
GU_MSurface2D	Set of Surfaces 2D	GM_MultiSurface
GU_MRing2D	Set of Rings 2D	GM_Aggregate
GU_MRing3D	Set of Rings 3D	GM_Aggregate

Table 1 - GeoUML ISO Spatial Schema class comparison

3. CHAPTER 3 - GIS (GTDB) TOWARDS 3D CITY MODELS

3.1 The 3D City Models

With the term “3D city model”, a representation of an urban environment with three-dimensional geometries has been intended. Such a model mainly represents everyday urban objects and structures, where buildings are the most prominent features. The growing body of research and material on this subject (many software tools that support a 3D city modelling format are nowadays available) suggests that it is of great commercial and research interest.

Nowadays, a growing number of geospatial applications require a 3D representation to analyse better environmental and urban phenomena such as sustainable urban development, water flow modelling, city climate studies, soil consumption analysis, architectural design, and generally 3D visualizations. These applications share a common need for reliable 3D geospatial data, whereas their requirements in terms of accuracy, resolution and interoperability can be considerably different (Tack et al., 2012; Biljecki et al., 2015). Given the growing number of applications integrating 3D city modelling among their features, it is even difficult to keep track of them: a complete inventory of the 3D city modelling applications probably does not exist.

Recent progresses in 3D geographic data acquisition, data management and 3D visualization tools have made 3D city models available for a broad range of uses, thus providing effective solutions for numerous applications (Peters et al., 2017). Nevertheless, the automatic, fast, and cost-effective construction of 3D city models is still an ongoing research topic.

Despite the predominant use of 3D city models is related to the visualisation or graphical exploration of cityscapes, there is a progressively increasing usage on tasks far beyond visualisation (Ross, 2010).

Moreover, some specific geospatial applications need to specialize and integrate objects provided in standardised 3D city modelling representations by formal extensions (Kolbe, 2009).

An important example is represented by the use of 3D city modelling on transport networks to analyse noise caused by railways and car traffic. For instance, (Lu et al., 2017) shows that the propagation of traffic noise coming from cars, motorcycles, aeroplanes and railroads can be determined much more accurately by exploiting a 3D map, e.g. considering also the building heights, than just a 2D grid, because the sound waves propagate on all the three spatial directions.

In practice, on the one hand, virtual 3D city models need to be managed to take into account independent data sources they are based on; on the other hand, close links to existing administrative workflows and databases have to be maintained. The balance of these two aspects is a significant challenge of 3D city model integration systematically and pragmatically (Döllner et al., 2006).

3.2 The CityGML

Recently, the semantic aspect of a 3D city model based on the CityGML standard has been dedicated to different initiatives. Several cities have already created their 3D city model using CityGML (TU Delft, 2019). The CityGML (OGC, 2012) is an open standard that defines a data model to exchange and store digital 3D models in city or landscape areas. The model includes buildings, roads, bridges, urban furniture expressed as 3D interrelated features.

According to the standard that defines "Rules for application schema" (ISO/FDIS 19109,2005), all classes are derived from the basic class called 'Feature'. Moreover, according to the standard GML3 (OGC, 2016), features comprise spatial as well as non-spatial attributes that are mapped to GML3 feature properties with the corresponding data types. The CityGML standard describes spatial objects recalling feature concepts from the GML'3 standard (OGC GML, 2016), which is based on the geometry model of the ISO Spatial Schema (ISO/TC 2011 19107, 2002). 3D geometries are then represented according to the B-Rep model (Foley et al., 1995).

CityGML uses only a subset of the GML3 geometry package (Stadler and Kolbe, 2007). Moreover, the topology can be represented explicitly. Every geospatial component is defined as a part of space and is modelled only once. Then a relationship with all features referenced to that geometry could be provided, avoiding redundancy and maintaining topological relations among different parts.

In practice CityGML refers to the formal geometrical constructs and languages defined in Geomatics-Geographic information standards, using them to define a semantic model oriented to describe urban areas. Indeed, it defines different standard Levels of Details (LODs) for the 3D objects (Löwner et al., 2013), thus allowing to represent them for different applications and purposes. The CityGML LOD could be intended as a pointer of spatial-semantic coherence between the real world and the adopted model to represent it or as an index of the richness of the geometry, intended as semantic granularity (Stadler and Kolbe, 2007). The LOD concept could be applied to different thematic object classes, even predominantly focused on buildings. . Five LOD instances have been related to the increasing of geometric and semantic complexity (Fig. 5) (Biljecki et al., 2016; Löwner et al., 2016).

Figure 5 - Building CityGML LODs (Biljecki et al., 2016)

3.3 3D cartography toward 3D city models: the Siena Old Town use case

From the topographic cartography point of view, some interesting experiences have been recently carried out in Italy: in (Corongiu et al., 2006), a 3D topographic model of the medieval town of Siena has been implemented paying attention to both third-dimension modelling and at a quite detailed scaling (1:500 cartographic scale). Based on a Specific GeoTopographic DataBase (GTDB), called SienaGTDB, this model preserves all selection query available in a geospatial DB, both related to geospatial features and alphanumeric attributes. Technical cartography and the roof map have been derived as outputs from this dataset.

The model refers to a mighty database, able to be used both by municipal officials as part of management procedures and by GIS user, for urban planning, spatial analysis, etc. In fact, data are modelled not only for visual scope but above all as territorial management support, thanks to their data inquiry and analysis potentialities.

The 3D information has been thoroughly extended to achieve a complete and accurate 3D reconstruction of volumetric objects. Then the overall conceptual approach of this model has been implemented in the Italian GTDB specifications (MD, 2012), even at a lower level (the biggest scale of the national core refers to a 1:1.000 scale). As a consequence, also a GTDB explained through the GeoUML methodology (as detailed in chapter 2) is compliant with the Siena GTDB spatial model. It is worth knowing that, the case study used in chapter 5 integrates contents formalised by the GeoUML methodology (explained in chapter 2) with the 3D spatial model of SienaGTDB (explained in this chapter) in a unique GTDB.

For what concerns SienaGTDB, the main task is aimed at defining each database object using one or more specific spatial components that in turn semantically characterise footprints or roofs. Geometrical footprints came from a celerimetric topographic survey while roofs have been acquired thanks to an airborne photogrammetric campaign. Then, both sources have been fully integrated into a GTDB.

The effort has been addressed to accommodate conventional geometries (foot and roof levels) within a 3D database avoiding the adoption of a specific three-dimensional model to obtain pseudo-vertical extensions (walls, buttresses, architectural foothills, etc.) in a medieval architectural context.

More precisely, each spatial component is made up of 3D objects through their primitive geometry coordinates. One of the most important issues in the development of this model was related to the correct description of void spaces below constructions (i.e., porches, lodges, galleries, etc.), and architectural details (i.e., bow-windows, overhangs, walls with battlements, etc.). To meet these requirements, the three-dimensional representation has been obtained by matching specific elementary volumes. Each elementary volume is obtained as a solid carried out extruding a reference surface along the vertical axis of a quantity q_g called extrusion elevation(Fig. 6), called "extrusion elevation". At physical level, each extrusion surface is implemented in an SHP file[13] format as PolygonZ, while the extrusion elevation is a numerical value expressed as an attribute of this spatial component.

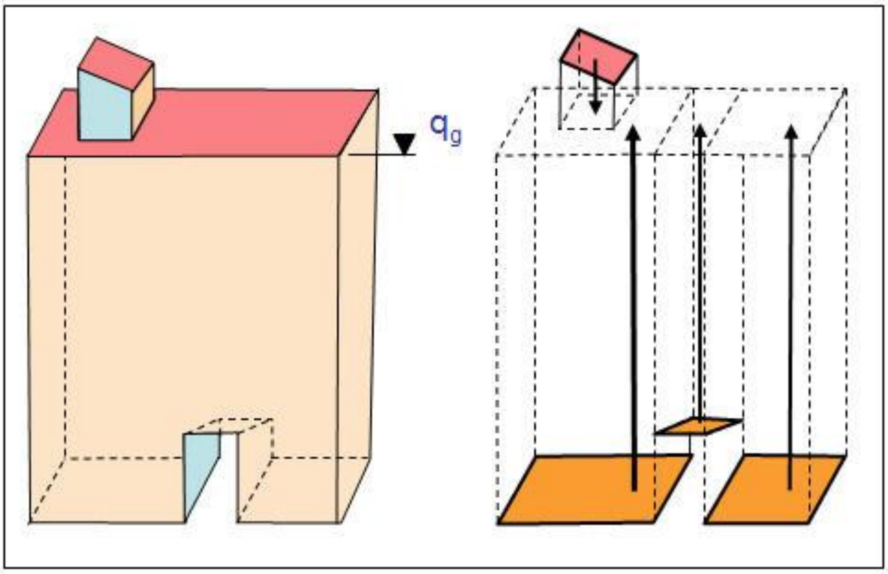


Figure 6 - extrusion mechanism

Some of the main differences between this 3D city model and a traditional 3D numerical cartography are:

- the 3D city model allows a 3D view not only for building features but for the whole urban environment, including green areas, walls, fences, transport infrastructures, etc. (Fig. 7);

---

[13] https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Figure 7 - overview of the Siena medieval old town GTDB

- The 3D city model provides different LODs because for each object one or more geometry types could be defined. For instance, a building could be defined by its volumetric units (higher LOD) as well as by its footprint or navigating its relationship with architectonical parts, roofs (Fig. 8), etc.

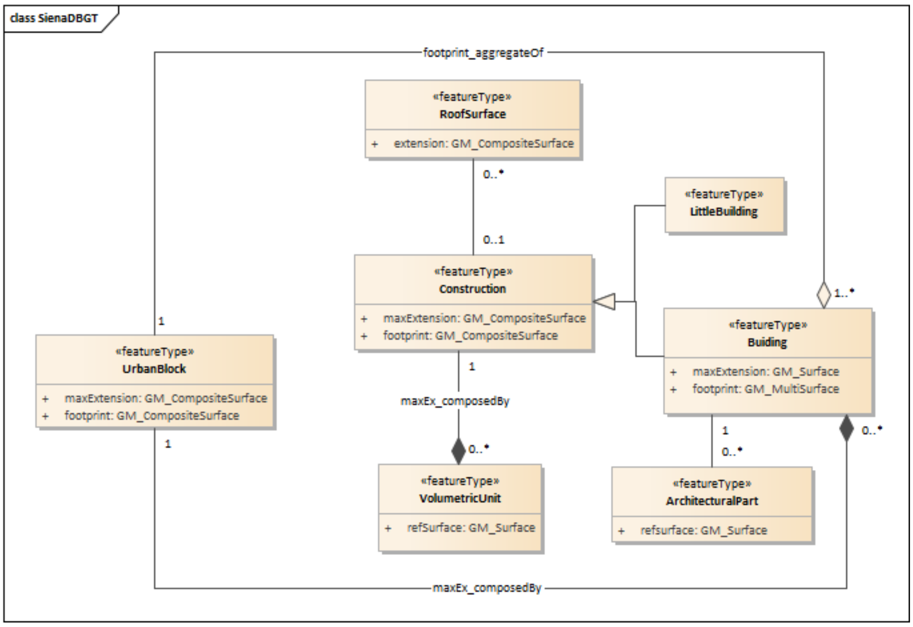


Figure 8 – UML schema of the SienaGTDB building relationships

- The topology must be verified and made geometrically congruent also taking into account the 3rd dimension. For instance, there are some constraints between objects on the ground and above it (e.g. bridges, projecting canopies…) that must be guaranteed.

Moreover, the descriptive and topological information must be semantically correspondent to their map portrayal.

Meantime some important requirements must be guaranteed:

- achieve a planimetric and altimetric precision for spatial objects according to their LOD (1:500 scale);
- ensure consistency between artificial structures (i.e. buildings) and contiguous ground surfaces;
- accommodate details not acquired by photogrammetric methods (galleries, porches, etc.).

Then the aim of this approach would be:

- avoiding the use of special/complex techniques, adopting those typically used for the production of digital maps (i.e. photogrammetric methods and topographic surveys);
- derive a real 3D visualisation by geometric components' extrusion methodology (Fig. 9, Fig. 10).

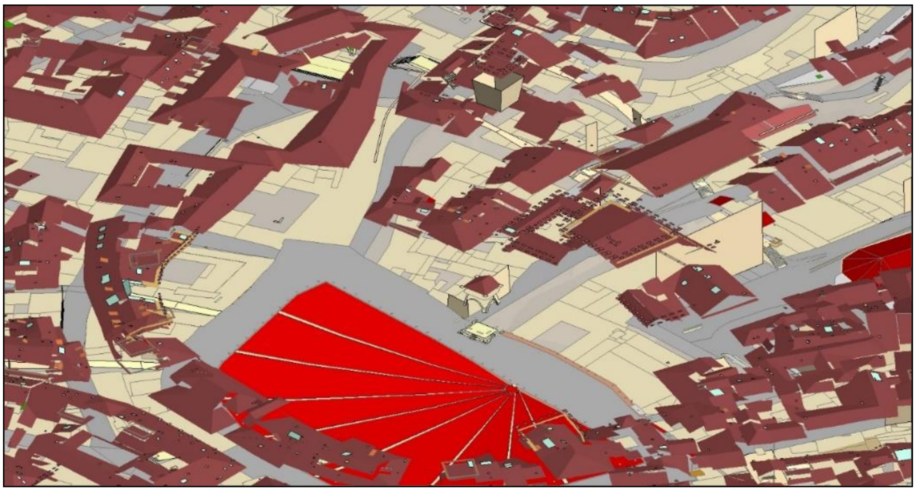


Figure 9 - data view before extrusion

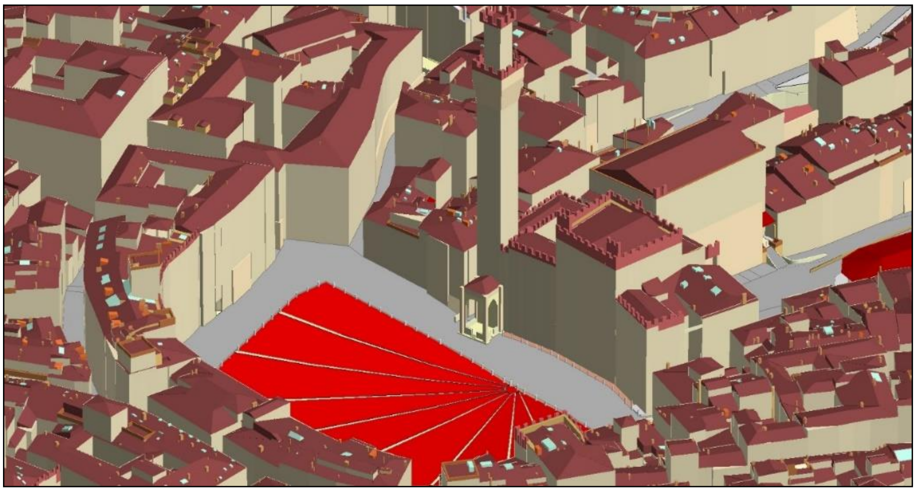


Figure 10 - data view after extrusion

- To allow a proper representation of certain architectural parts, such as roofs, walls, porches, medieval battlements (Fig. 11), projecting towers (Fig. 12), etc.

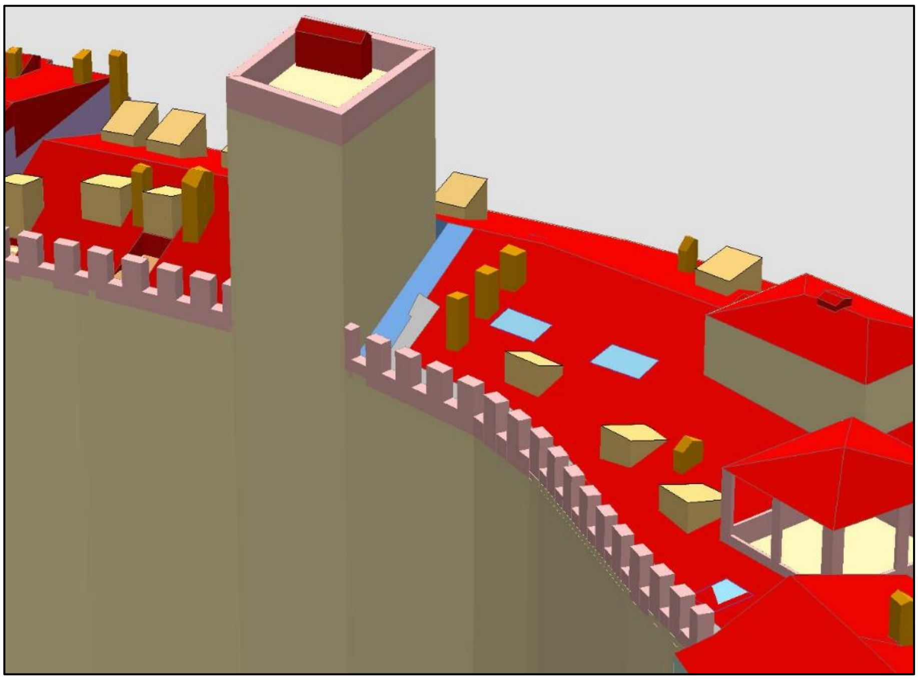


Figure 11 - architectonical parts on the roof

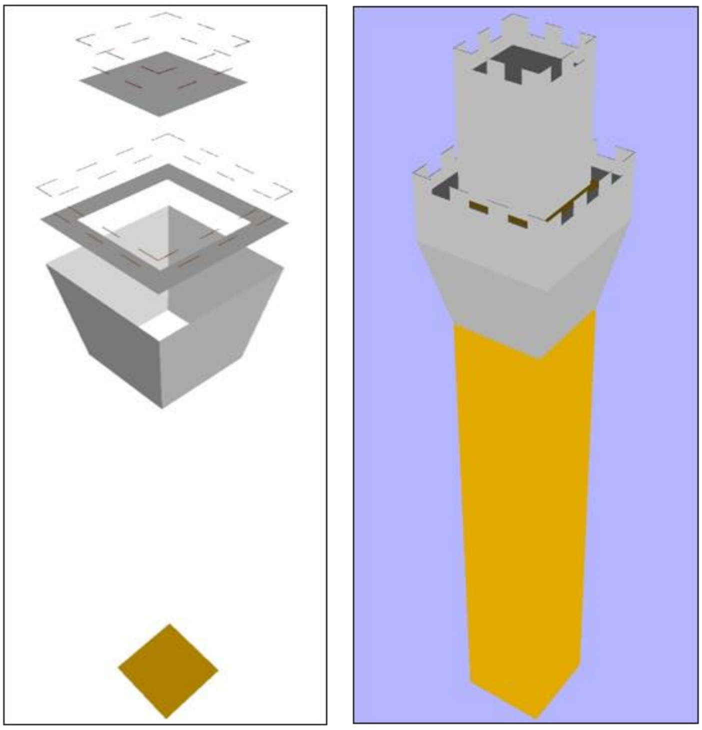


Figure 12 - particular of the Torre del Mangia: before extrusion (left), after extrusion (right)

- enable special object model representation: the dome of the Siena cathedral (Fig. 13);

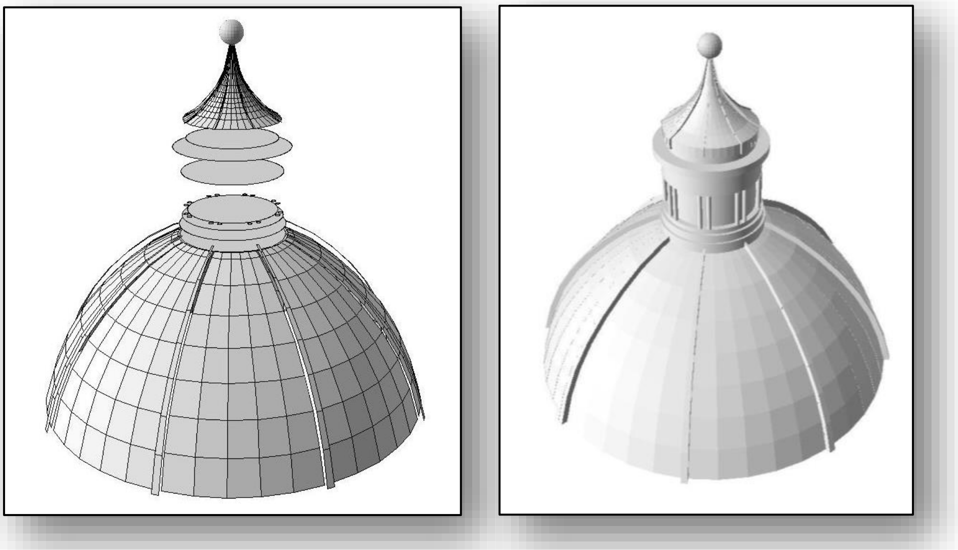


Figure 13 - dome modelling special object: before extrusion (left), after extrusion (right)

- deliver the 3D GTDB through well-known physical formats. For instance, the SienaGTDB has an implementation model based on SHP+DBF14 files.

The overall approach was implemented in the Siena medieval old Town case study, where some 2000 buildings, 15000 roofs and 18000 architectural objects have been captured with a density of 500 features per hectare adopting an input process made of paired and engineered “views” of aerial photos (photogrammetric method for roof level) and field surveys (for foot level) combined with not-metric imageries for attributive details.

3.4 Some conclusions on SienaGTDB and CityGML comparison

One of the main advantages in the SienaGTDB model adoption refers to its potential evolution toward a B-Rep of a 3D city model (i.e. CityGML model). A research topic, currently underway, refers to the automatic derivation of that B-Rep, starting from a model such as SienaGTDB. Preliminary results suggest that it could have a quite insignificant economic impact in comparison with the expensive costs implicated by the generation of a brand new 3D city model through an ad hoc data acquisition.

One of the most critical aspects of CityGML is related to the acquisition of information about objects that often cannot be directly obtained from traditional cartographic processes (e.g. airborne photogrammetric or topographic survey). Instead, SienaGTDB does not require such a level of information: it exploits an intermediate approach, which makes it ready for typical GIS applications. In this case, 3D visualization can be obtained thanks to the extrusion methodology starting from footprint and roof cartographies.

Moreover, in terms of interoperability, the use of a 3D GTDB addresses the integration of the 3D spatial features into an object-relational database (Tupper, C. D., 2011). In Italy, this approach is consistent with national/regional SDIs because based on the same reference specifications (MD, 2012), according to a multi-source DB approach (see Chapter 8).

The SienaGTDB model and the CityGML model have different perspectives: the first one aims at a detailed description (1:500 scale) of the 3D topographic point of view organised in a database, whereas the second one points towards a 3D semantic description of the urban context.

Basic differences could be summarised as follow:

- CityGML adopts both solids and the B-Rep geometrical model while SienaGTDB uses only reference surfaces at footprint/roof level together with the extrusion

---

[14] https://en.wikipedia.org/wiki/.dbf

methodology. 3D visualization is quite similar in the approaches, except for windows and doors (and, more in general, vertical surfaces) that are not included in the SienaGTDB model.

- CityGML adopts a semantic model relating elements belonging to the same object (i.e. wall, floor, roof of a building) as a part of it at a single LOD per time, while the SienaGTDB model implements a spatial database where objects could be related with classes and spatial component at different LODs (i.e. for architectonical elements, roofs, elementary volumes, etc.) (Fig. 14).

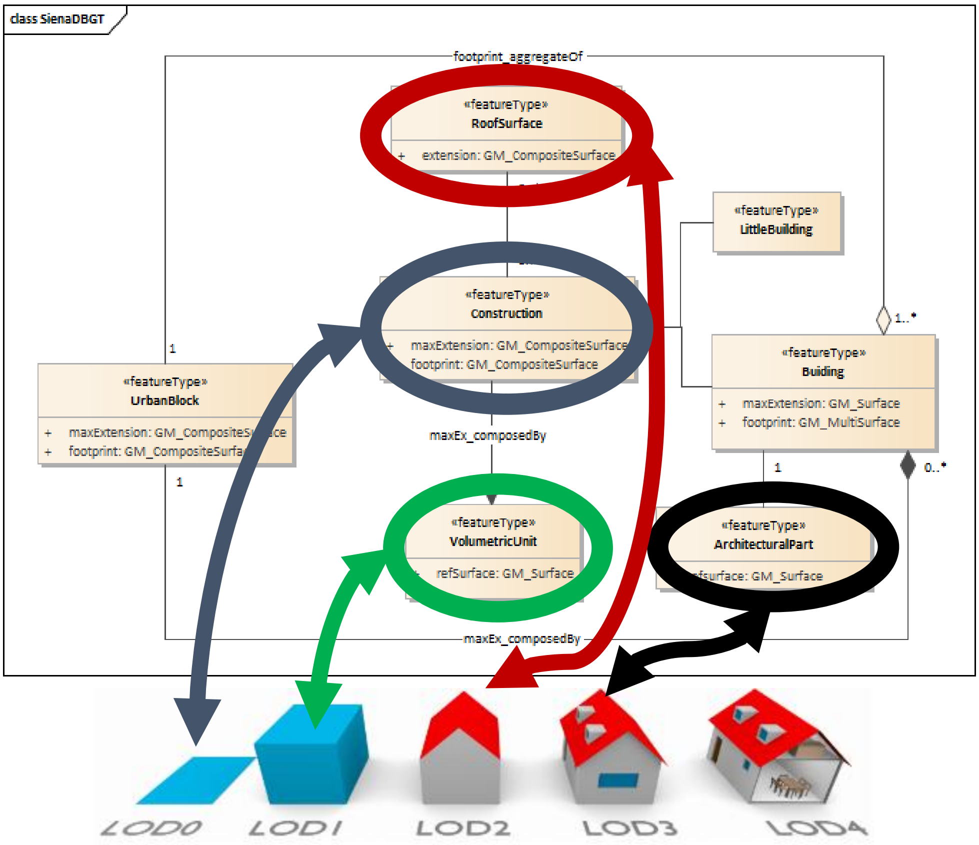


Figure 14 - SienaGTDB - cityGML model semantic correspondences

- In the SienaGTDB model, it's possible to define complex objects (Fig. 15) as abstract classes that aggregate different classes of objects (Fig. 16), while in the CityGML model this opportunity is not allowed in the core (Fig. 17) but, if needed, it can be defined as ADE (Application Domain Extension) of the model (Biljecki et al., 2018).

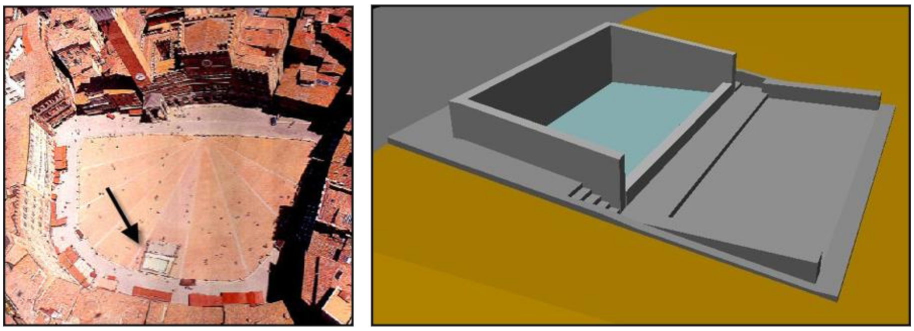


Figure 15 - Fonte Gaia as complex object: located in Piazza del Campo (left), 3d view (right)

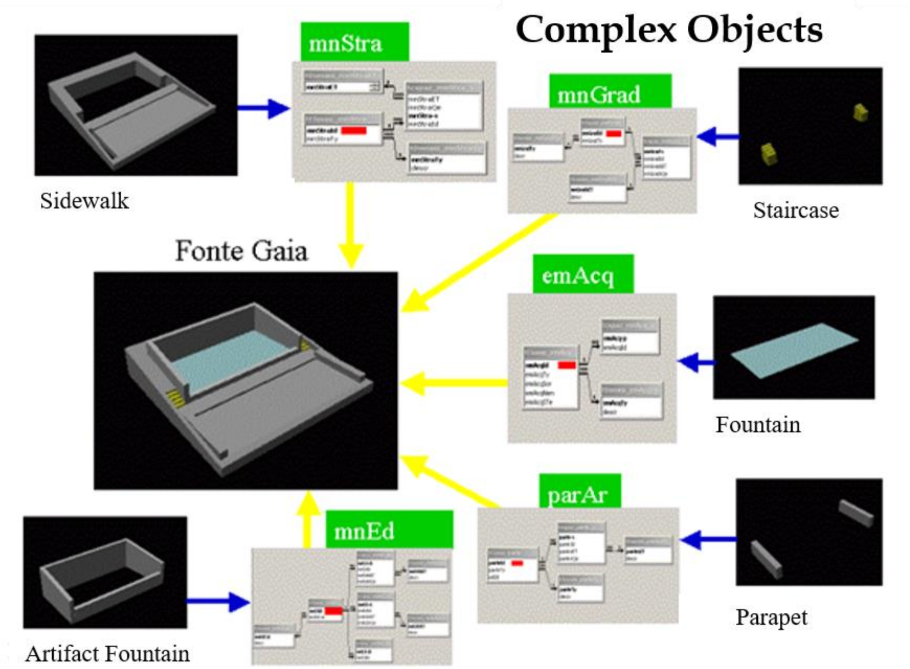


Figure 16 - Fonte Gaia: aggregation of different objects

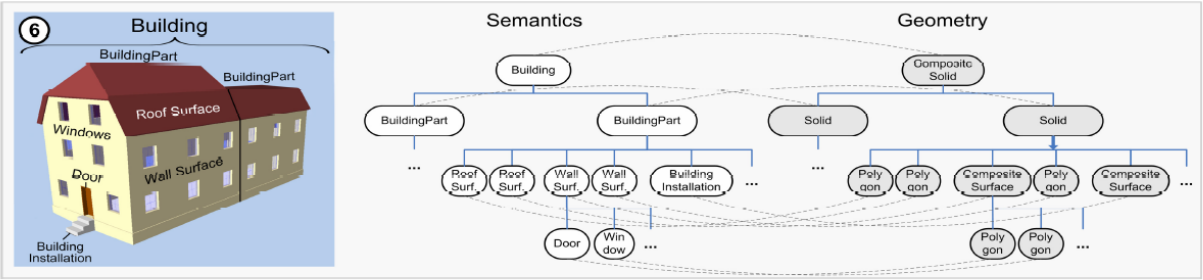


Figure 17 – CityGML: semantics - geometry relationship (from Stadler and Kolbe, 2007)

- acquisition costs to obtain a CityGML model are more expensive than those required by the production of standard cartography (in the latter case direct acquisition of lateral object surfaces is not required).
- From the visual and volumetric points of view, such two approaches are quite similar.
- The SienaGTDB model refers to all geographical objects with the same approach, while CityGML is more oriented to a detailed definition of buildings and their parts.
- From the semantic point of view, the SienaGTDB has a closer relation to the GML3 (Geographic Markup Language) standard (OGC GML, 2016) than CityGML. In fact, SienaGTDB is oriented to a territorial description in an abstract way like GML3.
- The SienaGTDB model is ready to be integrated with different Spatial Data Infrastructures as required by the INSPIRE Directive (Directive 2007/2/EC, 2007), recalled as Open Geospatial Web Services (OGC-OWS)[15] together with other territorial datasets, or implemented in a specific SDBMS. Instead, several CityGML examples have been implemented to covering the gap between SDBMS and geoportals (Yao et al., 2018), enabling the integration of core concepts independently of specific application schemas (Jetlund et al., 2020), as detailed in the Future development chapter (Chapter 9).

---

[15] https://www.ogc.org/standards/owc

4. CHAPTER 4 – THE BIM-GIS CONNECTION

4.1 BIM-GIS connection: the state of the art

Nowadays, the information management of spatial built environment using BIM or GIS systems is one of the research subjects of main interest; therefore, there has been quite intensive research production on this topic during the last decade. BIM is a new paradigm of digital design and management and its use is already mandatory in certain Countries (Tucci et al., 2019). Thanks to 3D data acquisition geomatics technics, in recent years, BIM has been increasingly applied to manage documentation and information of historical (Lopez et al., 2018; Pocobelli et al., 2018).

Nevertheless, the integration of BIM and GIS, namely GeoBIM, could be a helpful way in 3D city modelling, but several incompatibilities have to be solved. GIS started modelling the environment performing mainly 2D spatial operators in large areas. However, recently automated GIS workflows have generated detailed 3D data, thanks to increased performance in surveying and modelling 3D data, even if related to an individual building, that is a traditional domain of a BIM application (Arroyo Ohori et al., 2018).

From the building information point of view, BIM models are, for each object, much more detailed and semantically richer than GIS models that otherwise cover large territories.

BIM overpass the concept of data modelling; in fact, such digital representation could make available additional information about building lifecycle, construction processes, operation, and maintenance (Kumar et al., 2019).

The leading open standard to deliver a BIM is the IFC (Industry Foundation Classes) (ISO/TC 59/SC 13 16739, 2018), while the 3D GIS one is the CityGML (OGC CityGML, 2012). These two formats focus on different information domains, so converting data raises some interoperability issues between them (Matrone et al., 2019).

From a practical point of view, back and forth actions between BIM and GIS software are often supported only by few proprietary native formats.

Moreover, often the conversion cannot avoid loss of information, thus requiring recreating, sometimes manually, portions of datasets (Kavisha, 2020).

In the past decades, common requirements for bridging the data models and workflows from the AEC[16] (Architecture, Engineering, Construction) and the geospatial community have been formalized to overcome the single weaknesses of BIM and GIS, making available the BIM objects in both the IFC and CityGML standards. Therefore, GIS and BIM datasets differ fundamentally concerning their semantics, geometries, and level of details (Ohori et al., 2017).

However, some interesting remapping experiences have been conducted. For instance, the differences in describing building systems between the two data models refer to the adopted LOD, since it refers to a significant difference in ontologies. For instance, considering the two file formats and comparing main entities in a hierarchical way (Fig. 18), it appears that IFC decomposes the building more in detail than CityGML does. As a result, doesn't exist a one-to-one mapping relationship between classes (Cecchini, 2019).

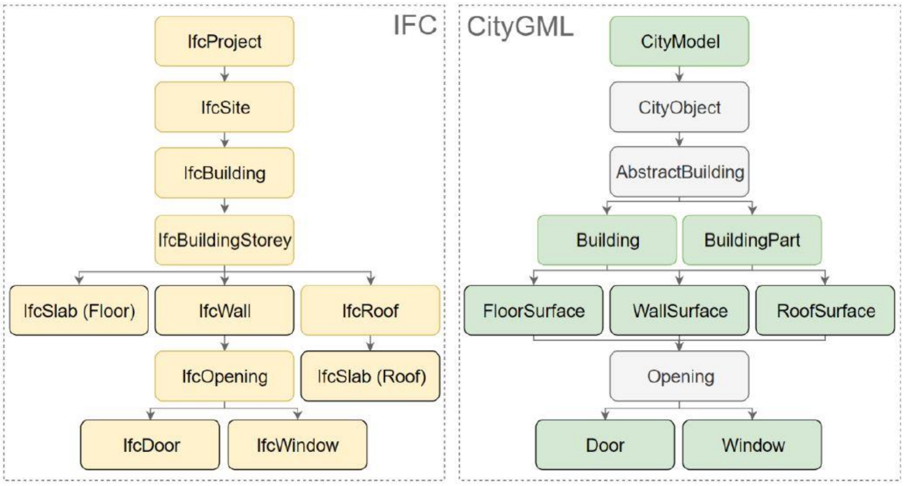


Figure 18 – "Comparison between the hierarchical scheme of IFC and CityGML. The blocks with the black contour carry geometric information. The blocks filled in grey shall forward the attributes without proper identification" (Checchini, 2019).

A specific geomatic ISO standard (ISO/TC 211 19166, 2018) defines a BIM to GIS criteria to mapping element information from BIM to GIS.

. In this standard three mapping mechanisms are defined:

- BIM to GIS Perspective Definition (B2G PD);
- BIM to GIS Element Mapping (B2G EM);
- BIM to GIS LOD (Level of Detail) Mapping (B2G LM).

The conceptual mapping mechanism uses existing Standards such as GML (OGC GML, 2016), CityGML (OGC CityGML, 2012), and IFC (ISO/TC 59/SC 13 16739, 2018), but a unified information model developed between BIM and GIS is out of the scope of this standard.

---

[16] https://www.igi-global.com/dictionary/building-information-modeling-australian-architecture/1428

The lack of a shared model approach concerning geometry, semantics, and schema is the main reason why the two domains remain disconnected even though at the physical level the integration of BIM and GIS could advantage both domains. For example, if on one side more detailed 3D city models can be built by reusing the BIM data, on the other side BIM designers give detailed information georeferencing their model and considering the territorial context information. (Kumar et al., 2019).

Nowadays, in practice, a great part of the development efforts has been focused on the unidirectional transformation: from BIM to GIS. Probably because the translation concerns a one-off necessity to transform physical datasets, indeed users avoid considering a metamodel that could link the two domains at a theoretical level, because it doesn't resolve implementation issues.

Another aspect related to BIM-GIS connection, particularly in the Italian context, since the main part of urban areas consisting of historical buildings, the BIM concept is integrated with the characteristic of Heritage, namely HBIM.

In general, the binomial HBIM-GIS is solved by managing an object-oriented model representing HBIM into a GIS environment, thus obtaining support for planning but simplifying as much as possible the analysis at parameter levels or virtual trends that are commonly used in an exclusive HBIM environment (Tucci et al., 2019b).

It's worth noticing that, considering each existing building to been modelled by a related HBIM, the geo-referencing accuracy is one of the crucial aspects that need to be taken into account in a transaction that building in a 3D GIS environment.

Moreover, an increasing amount of CH (Cultural Heritage) data needs to describe complex systems like cities or vast territories combining several high LODs and different scales (Colucci et al., 2020).

It means that to integrate HBIM into GIS is necessary to follow a multi-accuracy approach, this issue is not been fully developed yet. A practical solution for multi-accuracy DB management has been detailed in Chapter 8.

4.2 The LandInfra/InfraGML standards as a BIM-GIS "connecting bridge" for railways.

A new standard, namely LandInfra (OGC LandInfra, 2016), aims to cover the lack of conceptual modelling inBIM-GIS process integration. LandInfra defines a Conceptual Standard Model about land and civil engineering infrastructure facilities throughout some implementation-independent concepts. Conceptual contents refer to facilities, projects, alignment, road, rail, survey, land features, land division, just to name a few.

The main interoperability characteristics of the landInfra are detailed as follows:

- It is an OGC standard, so compatibility with other Geographic Information Standards, such as OCG and ISO/TC 211, is guaranteed.
- The implementation level is represented by the InfraGML standard (OGC InfraGML, 2017), where geometric entities refer to the GML standard (OGC GML, 2016). Hence GML functionalities are inherited, such as geospatial feature concepts, coordinate reference systems, linear reference system. In particular, the relationship with the LRS standard (ISO/TC 211 19148, 2012) makes it possible to integrate many of the transportations requirement to locate information.
- It is based on a UML conceptual model, developed prior to GML encoding.
- It is synchronised with the concurrent efforts carried out by buildingSMART International (bSI)[17] in the development of infrastructure-based IFCs, although this one uses the EXPRESS language (ISO 10303, 2004) instead of GML language of the OGC ones. bSI is an open, not-for-profit, and neutral organization representing the worldwide industry body, driving the digital transformation of the built asset industry.
- It is more easily integrated with CityGML. There are some potential overlaps between LandInfra and CityGML, but the level of detail is significantly different: CityGML focuses on building whereas LandInfra focuses more on land and infrastructures facilities other than buildings.

For all this reason LandInfra could be defined as a ‘connecting bridge’ between the BIM and GIS domains (Kavisa, 2020) (Fig.19), above all in the GIS infrastructure context.

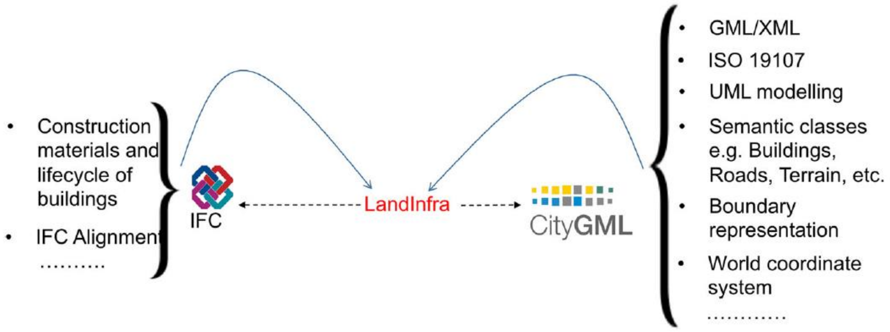


Figure 19 – “LandInfra a connecting bridge between IFC and CityGML, but is conceptually, semantically, and geometrically closer to CityGML” (from Kavisa, 2020)

---

[17] https://www.buildingsmart.org/

In addition, the Alignment package of LandInfra has been developed jointly with the bSI IfcAlignement of IFC. The two standards are quite similar. Some differences could be observed according to different domains of application. This consistency should enable linking of geospatial and BIM database upon linear location.

In the specific Railway case, which is the case study developed in detail in the next chapters, there is compatibility between LandInfra and bSI IFC (bSI IFC railway, 2015) models, even though the bSI IFC one is more detailed than LandIfra's correspondent (OCG LandInfra, 2016).

The spatial structure of railway engineering adopted in bSI Railway standard is shown in Fig. 20: "The railway project (IfcProject) may contain one or more railways (IfcRailway) and one or more railway terminals (IfcRailwayTerminal). IfcRailway may consist of one or more alignments (IfcAlignment), one or more tracks (IfcTrack), and one or more sites of tunnels (IfcTunnel), subgrades (IfcSubgrade), bridges (IfcBridgeIfcBridge), stations (IfcRailwayStation) and buildings (IfcBuildingIfcBuilding)" (bSI IFC railway, 2015). Moreover, a railway terminal (IfcRaifcRailwayTerminal) may also consists of a series of railways (IfcRailway) and stations (IfcRailwayStation).

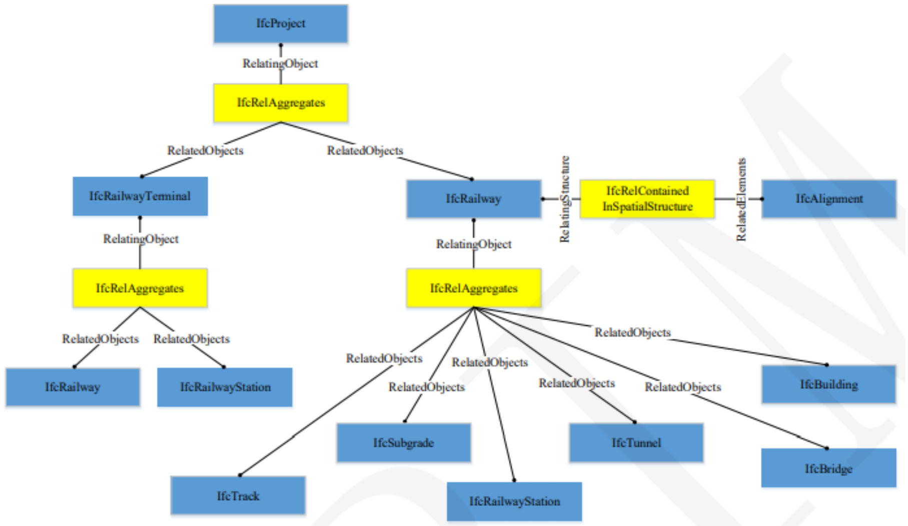


Figure 20 - Railway engineering spatial composition (from bSI IFC Railway, 2015)

In turn, each part of railway engineering is defined by a specific spatial composition. A building part (IfcBridgePart) refers to the various parts of the Bridge (IfcBridge), depending on the type of this one. For instance, considering a Bridge as shown in Fig. 21, it is possible to decompose the bridge in different BrigePart such as gird, abutment, pylons, cables, arch, suspenders, foundation, suspended tendons and bridge floor system.

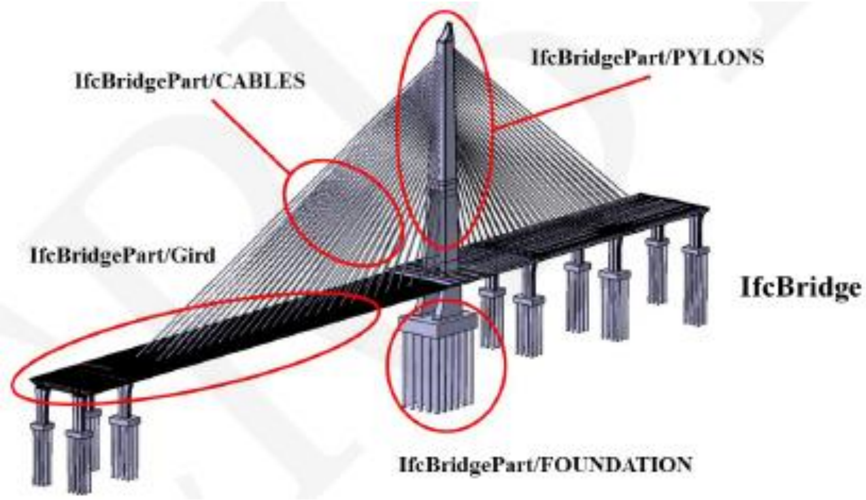


Figure 21 - bridge spatial structure decomposition (IfcBridge(ELEMENT)) from (bSI IFC, 20015)

InfraGML (OGC InfraGML, 2017) Encoding Standard presents the implementation-dependent of concepts supporting LandInfra. For this thesis, the “Part 5 – Railways” of this standard is considered (Fig. 22).

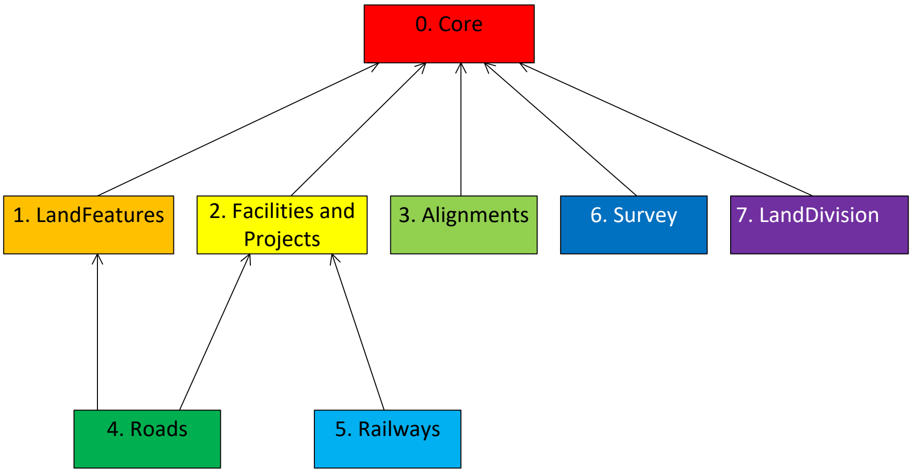


Figure 22 - InfraGML part Dependencies from (OGC InfraGML, 2017)

In this standard the Railway requirement class supports the exchange of information related to the design of a railway, avoiding supplying the entire IFC format. Contents refer to Railway Elements such as track geometry, chainage discontinuities and superelevation (cant).

An example of a railway object (railway switch), defined according to InfraGML standard (OGC InfraGML, 2017) could be found on the B.2.

4.3 The contribution of point clouds in railway BIM-GIS integration

Traditionally, point clouds have been considered source data. Generally, to use them in a GIS environment, they are converted into grids or vector objects as a base to support further processing. In recent years, point clouds have been increasingly applied for visualization purposes, especially in web applications.

In the meantime, the suitability of DBMS for managing point cloud data continues to be debated (Oosterom et al., 2017).

to increase the realistic representation of virtual 3D city models, texturing became a valuable property associated with each object, such as building façades or roofs, often combining aerial and ground-based laser scanning sources with aerial or satellite imagery (Peters et al., 2017).

Meantime laser scanning has become the predominant data source in BIM development, highlighting the need to integrating BIM and GIS domains in a homogeneous framework. BIM and GIS can model the built environment in 3D (representing both indoor and outdoor features). Moreover, the usability can be improved by implementing a specific SDBMS and taking advantage of the more accessible query capability of a DB, not only focusing on repository aspects.

BIM and GIS both can represent the world “as is,” providing an efficient method for managing, documenting, and visualizing spatial and non-spatial information. BIM has been characterized as “a modelling technology that combines the design and visualization capabilities of CAD with the rich parametric object and attributes modelling of GIS” (Ellul et al., 2017).

To recap, some main differences between BIM and GIS approach are:

- in BIM, the semantic information is prominent; different engineering project aspects can be adequately visualized, e.g. construction materials can be represented by patterns and hatching. The BIM objects such as walls, ceilings, stairs, windows, stairs are modelled according to a parametrical approach. Hence, some detailed information (wall thicknesses, structural components, etc.) can be visualized. Each BIM model refers to a single project. No georeferenced information is generally provided. About pre-existing buildings or infrastructure, often geometries came from the restitution of a laser-scanning survey.
- In GIS: semantic information is organized in spatial and no spatial attributes. Geometries represent geospatial features. Data are often collected in object-relational

DBs, mainly accessible through the SQL (Select Query Language) language. In any GIS, the Spatial Reference System management and topological operators have been defaulting implemented. Generally, volumes could be described according to a B-Rep model (Foley et al., 1995).

BIM and GIS integration, thanks to recent technologies, allow gaining considerable benefits in the lifecycle management of infrastructures. Moreover, “BIM has been used to facilitate the integration, interoperability, collaboration and automation of processes in the construction industry” (Isikdag and Zlatanova, 2009).

Particularly about Railway infrastructure, there is an implicit need to monitor the railway geometry and surrounding environments, and a significant roadmap has been designed on that assumption (i.e. http://big.yonsei.ac.kr/railbim/). Therefore, three-dimensional modelling is necessary to maintain safety levels despite the inevitable wear and tear of the railroad over time (Neves et al., 2019). Moreover, one of the primary sources came from a Mobile Mapping Systems (MMS) mounted directly on the train where point clouds are the output of the acquisition phase and the railway solutions. In any case, point clouds are not natively segmented or classified, i.e., they represent all of the elements surrounding the railway track at a given moment. . The implantation of a method that allows an efficient classification of the points representing a specific object among the millions of cloud points, must be defined (Gézero & Antunes, 2019).

Recently, several studies have been focused on extracting elements from point clouds datasets acquired by MMS, such as in (Pastucha, 2016), (Rodríguez-Cuenca et al., 2015), (Yu et al. 2014), (Che et al., 2019) just to name a few. However, particularly for railways, the most recurrent subject is represented by the track lines segmentation (Zhu & Hyyppa., 2014). Moreover, considering the derivation of point clouds to obtain a BIM for railway modelling considerable experiences have been explained in different studies, among them (Bensalah, 2018) and (Nuttens et al., 2018).

Hence, it seems that BIM-GIS connection in common platforms can bridge the gap between the world scale and detailed data (Kurwi et al., 2017). Collaboration may play a crucial role to solve existing problems, but from the research point of view, the lack of common and general solutions on an effective integrated framework is already an open issue.

4.4 Some conclusions on BIM GIS integration

Effectively BIM-GIS conversion, translation and extension require the use of existing standards. This process can be manual or semi-automatic, but in general, involves semantic adaptations that may lead to unavoidable information loss. In semantic web technologies and services-based methods, a common approach is oriented to enabling and sharing data rather than converting existing standards or developing new ones; both are addressed. Otherwise, information loss might happen during integration or the data filtering process.

Moreover, as most integration works at the process level, it requires extensive human intervention so in the transition phase of integration low productivity has to be taken into account. In terms of re-use of processes, seems unavoidable a certain level of revision in order to convert, translate and extend standards(Liu et al., 2017).

Following a BIM evolution point of view, some scenarios have been assumed as the future development of integration: 1) technology integration, where BIM and GIS are partially utilized together to address specific problems, as mostly happened so far, 2) science integration, where location-based theories and technologies could be integrated into BIM. 3) source integration, where the BIM could be considered like remote sensing (RS) in monitoring natural resources by LIDAR (Laser Imaging Detection and Ranging) and photogrammetry (Song et al. 2017) (Fig. 23).

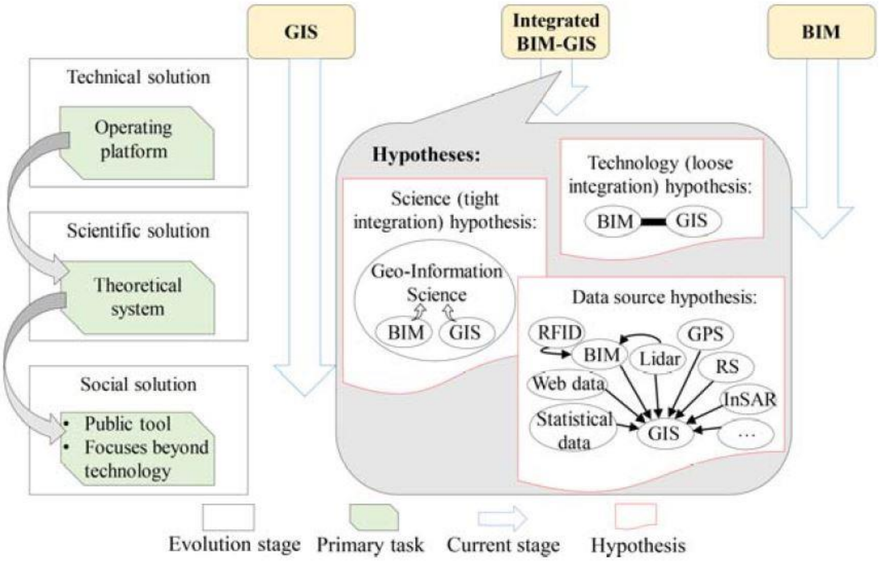


Figure 23 – "Hypotheses of future development of BIM-GIS integration" (from Song et al. 2017)

Despite the unbalanced point of view on BIM found in literature, the predominantly GIS orientation of this thesis, on the contrary, focus on a definition of a metamodel from a GIS/SDI context that could be used as a linked model to connect different domains as GIS and BIM are (Fig. 24).

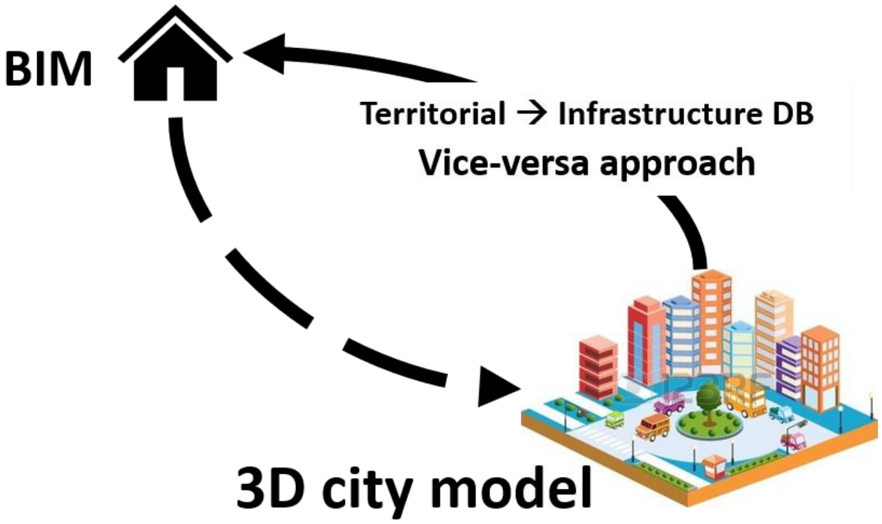


Figure 24 - From GIS to BIM link metamodel

In this sense, the join model could be represented by CityGML if the connection refers to applications strictly related to buildings, while landInfra/InfraGML are more performant for general datasets both in the urban and extra-urban environments. Moreover, for LandInfra/InfraGML the repeatability is assured in an interoperability way.

5. CHAPTER 5 – THE CASE STUDY

This chapter is mainly based on the (Corongiu et al., 2018) scientific publication.

Despite the existence of several public transportation methods, railway transportation still represents one of the most used ones for both short and quite long travels. The maintenance of railway systems typically requires monitoring thousands of kilometres of lines, which is quite challenging and time-consuming. The Italian railway system includes more than 16 thousand kilometres of active lines, with approximately 900 million passengers per year. The Italian Railway Network Enterprise (RFI), which manages the Italian railway system, launched the MUIF (Unique Model of the Physical Infrastructure) a few years ago, to manage spatial information in a single model related to the infrastructure of Italy’s railway system (Corongiu et al., 2020).

The inspiration for this thesis starts with datasets and the spatial model of the MUIF project. According to thesis objectives, MUIF requirements imply to design in single GIS:

- Reference geospatial information delivered in a 3D GeoTopographic Database (GTDB);
- Railway assets integrated into the 3D GTDB congruently;
- Multi-source/multi-accuracy data and spatial data model design

Indeed, up to now, the georeferenced railway network information has been managed aimed at two basic aspects: the first one addressed to asset monitoring, the second one for commercial tasks. Before MUIF, the common nature of geospatial data has been managed as independent information. Any stage of the update on each side of the two systems did not impact the other one because the two systems worked in separate compartments. So, the building of a single integrated model focuses on the evolution of the existing network model to:

1. guarantee the connection among railway specific objects, network models and different operational applications within the RFI;
2. access data at different levels of detail;
3. deal with different temporal scenarios.

Based on this context, the MUIF project represents both the first step of a big industrial project and, at the same time, an excellent chance to research new integrated technology and methodology to manage geographic information over time and different scenarios. Therefore, some thesis objectives refer to the design of the integrated GTDB, such as general interoperability, 3D aspects, reference Vs. thematic GIS connections. Some other objectives aim analysis and tests to support future MUIF update phases such as BIM-GIS connection and, as developed in the next chapters, multi-source 3D SDMBS management.

5.1 Geographical Extent

As the MUIF project refers to the Italian Railway Network, the extension of the case study could be identified with all the Italian territory. Territorial branch structures are in charge of carrying out production activities related to network maintenance/management processes for safe train circulation and station quality. These administrative sectors are called Dipartimenti Territoriali di Produzione (DTP), alias Territorial Production Departments, and are coordinated by the Central Production Department (Fig. 25).

Figure 25 – DTP extent in the Italian territory

The delivery of the MUIF project is organized in DTP lots, each of them validated and matched with the previous ones already provided. Currently (as of December 2020), data provision progress covers about three-quarters of the entire national extension. For the different experimental phases of this thesis some DTP datasets have been used:

- DTP Bari;
- DTP Reggio Calabria;
- DTP Venezia;
- DTP Firenze.

5.2 Source datasets

MUIF survey campaign is carried out using different sources and sensors. Hence all data, with their temporal and spatial dimensions must be organised into a single model of spatial railway information.

About positional requirements, data refer to a comparable level of detail and sometimes to many source accuracies. The accuracy of the surveyed data depends on different parameters:

- the cartographic error of survey at a specified scale;
- the quality of data sources;
- the Spatial Reference System (SRS);
- the transformation algorithm among different Coordinate Reference Systems (CRS).

Source data refers to:

1. an airborne photogrammetric survey (medium Ground Sample Distance – GSD – accuracy = 0.08 m);
2. a LIDAR survey by plane (point density> = 4 pt./m^2) (Fig. 26);

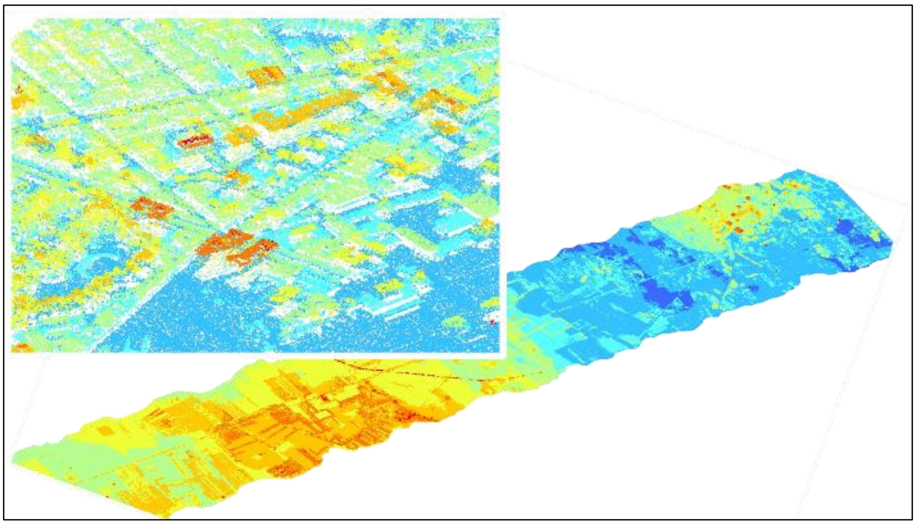


Figure 26 - airborne LIDAR survey

3. a Mobile Mapping System (MMS) by train, side-by-side images (Fig. 27), panoramic images (Fig. 28) composed of six synchronous shots, acquired in six different directions, and telemetry survey point clouds(Fig. 29).

   The laser scanner is located at the end of the train. It surveys a buffer zone of40 m from the track centreline. Stereoscopic images also supply synthetic clouds for an amplitude track band of 25 m and approximately every 2 m.

   .

Figure 27 - side-by-side images

Figure 28 - panoramic image

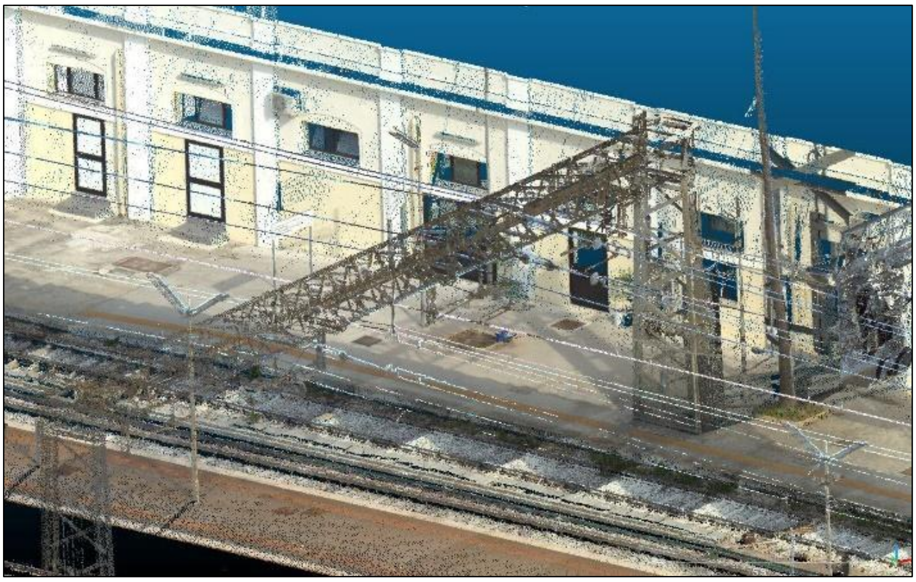


Figure 29 - point clouds from train telemetry laser scanner

4. point clouds data coming from a scanner located in the upper part of the train. It is addressed to railway infrastructure global monitoring). This set consists of 3D points of the railway ballast and objects very close to the tracks.

Processed datasets refer to:

1. a railway track centreline;
2. orthophotos (pixel size =0.10 m) (Fig. 30);

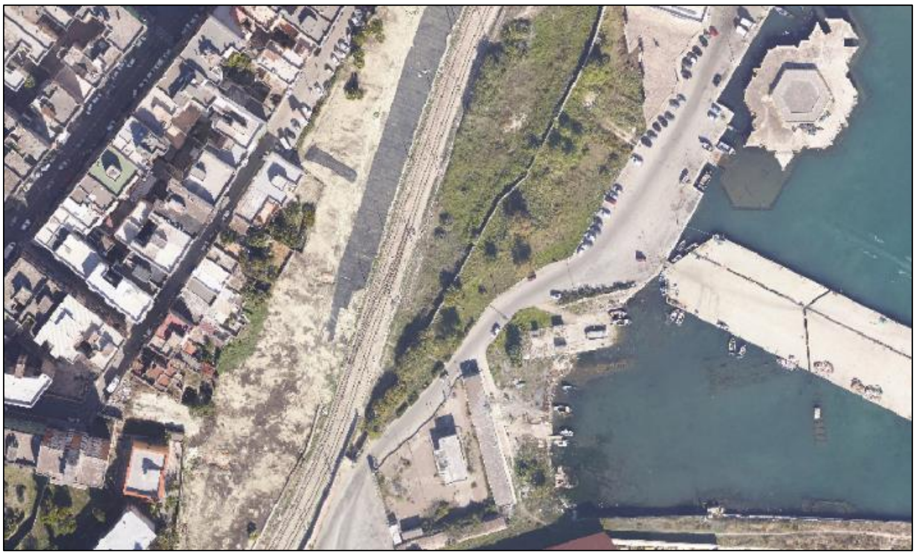


Figure 30 - orthophoto

3. Dense Terrain and Dense Surface Digital Models (DDTM-DDSM) (resolution =1 m) and a Digital Elevation Model (DEM) (resolution = 5 m) as derived from LiDAR point clouds;
4. Point clouds of telemetry survey with RGB attribute from panoramic images;
5. Railway asset and topographic contents integrated into a Multiaccuracy GTDB.

5.3 The point cloud datasets

Datasets coming from the train laser scanner refer to telemetry point clouds together with a sequence of simultaneously panoramic images used to associate their RGB value to the point cloud data.

The RGB colours were associated with point clouds extracting information from georeferenced panoramic images. The telemetry survey is equipped with a GPS IMU system which, at every moment, estimates coordinates and position of both the laser scanner and the camera.

The terrestrial LIDAR data comprises all the railway and its surroundings, including the ground, buildings, streets, etc. An additional survey is carried out in the railway station surroundings to integrate objects outside the radar range. To be more precise, a laser scanner backpack device is adopted to acquire additional railway assets. As a result, other point clouds have been developed with similar characteristics as those obtained by the mobile train system and merged in a unique, delivered dataset (Fig. 31).

Figure 31 - compound train and backpack LIDAR data in the station zone

The survey consists of a sequence of train runs, temporally ordered, each run giving rise to one or more point clouds. For the sake of delivering and managing convenience, the point clouds are split into sub-clouds: each of such sub-cloud represents approximately a 1 km long area along the railway track, and it is stored in a LAS file. During the execution of the point cloud splitting procedure, an overlap is kept between adjacent trunks, in such a way as to ease their match (Fig. 32).

Figure 32 - The extent of delivery overlapped point clouds

It is also worth mentioning that several (overlapping) point clouds might cover the same area. Because the case study considers just the first implementation of the MUIF project, it is due to the closeness of different railway lines, leading to overlapping point clouds collected during data collections on such other lines. Therefore the same situation may occur in future periodical monitoring of such area.

The LIDAR data are in LAS 1.2 classification format with seven classification types: “Unassigned”, “Ground”, “Low Vegetation”, “Medium Vegetation”, “High Vegetation”, “Building” and “Noise”.

5.4 Multiaccuracy GTDB

In the GTDB different zones have been defined to include territorial objects at different spatial/thematic accuracies. For each zone, a specific 3D accuracy has been identified according to cartographic scale representation:

- 0.40 m for planimetry and altimetry dimension related to a 1:1.000 scale cartography;
- 0.60 m for planimetry and altimetry dimension related to a 1:2.000 scale cartography.

Three different accuracy zones have been defined (Fig. 33):

1. Railway Station zone (Fig. 33 blue zone): it is the area where both positional (1:1.000 scale) and thematic (completeness of objects in the GTDB) accuracy is the highest considered;
2. 120 m buffer zone of the track centreline (Fig. 33 red zone): it represents the surroundings of the main railway with the positional accuracy related to a 1:2.000 scale cartography, and the thematic accuracy related to the completeness of objects in the GTDB;
3. 120 m – 500 m buffer zone of the track centreline (Fig. 33 yellow zone): it represents the more distant surroundings of the main railway with the positional accuracy related to a 1:2.000 scale cartography, and the thematic accuracy related to only Building layer of objects in the GTDB.

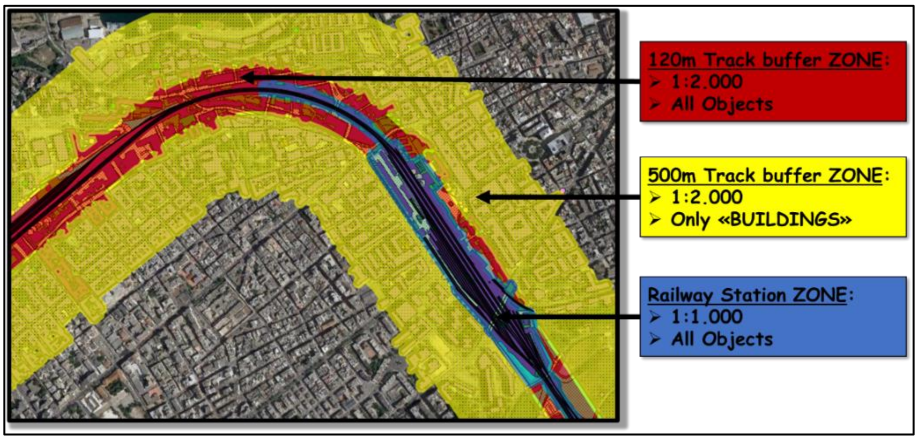


Figure 33 - different zones of the multiscale GTDB

5.5 The vector semantic model

Data on the vector model are organized in a GTDB following Italian Specifications (MD, 2012) and the GeoUML methodology as explained in Chapter 2. Therefore, the GeoUML Catalogue has been used to support the definition of the conceptual schema specifications, the Implementation Model, called “FLAT SHP”, has been physically delivered in shapefile (SHP) and DBase File (DBF) tables, being directly generated thanks to the GeoUML tools, then the GeoUML Validator has been used to perform the conformity check of a dataset or database with respect to the specification on the GeoUML Catalogue (Tucci et al., 2020).

The third dimension has been modelled in a relation-free structure, independently from the specific use case application. So this is not specifically oriented for visualisation, project analysis scopes, or a specific field of study. The GTDB has been designed in a single DB, integrating heterogeneous information coming from different data sources and combining and harmonising data according to the higher level of interoperability (conceptual interoperability of Fig. 1). The challenge of this approach has not been limited to solve the 3D data modelling but has been focused on the integration of 3D data coming from different data sources, with different structures and accuracies, into one harmonised data model aimed at several tasks such as management, updates, etc. as detailed in Chapter 8.

Regarding 3D, the model is compliant with the SienaGTDB detailed in Chapter 3, where “volumetric units” are intended as a natural extension of technical cartography contents, and obtained extruding each reference surface. This technique allows the volume disposition of a building or any other topographic object to be represented. To obtain a visual 3D B-Rep, a specific field “extrusion” has been populated (Fig. 34).

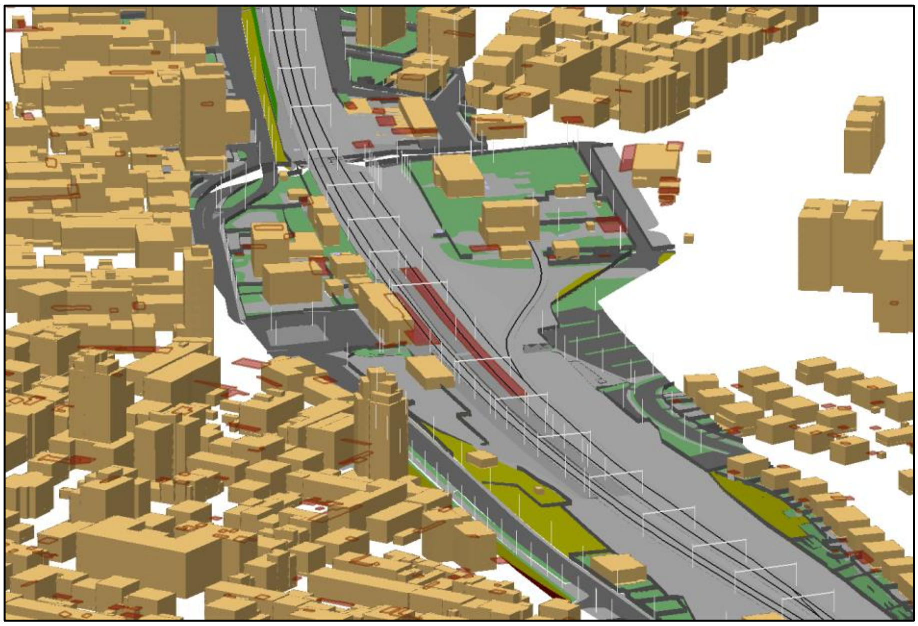


Figure 34 - 3D realistic visualisation of GTDB by extrusion

Furthermore, some examples of 3D experimental modelling have been explained in Chapter 8.

5.5.1 Asset integration

Based on existing standards, topographic and railway assets have been organised in a unique integrated database(Fig. 35). The assets describe different functional objects of railways such as railway switches, intersections, or artefacts (compliant with specific railway standards as detailed in Chapter 6). They have been modelled in the integrated GTDB, following the same approach for topographic objects that is location and footprint geometry-based. In turn, the GTDB complies with the Italian national specifications (MD, 2012).

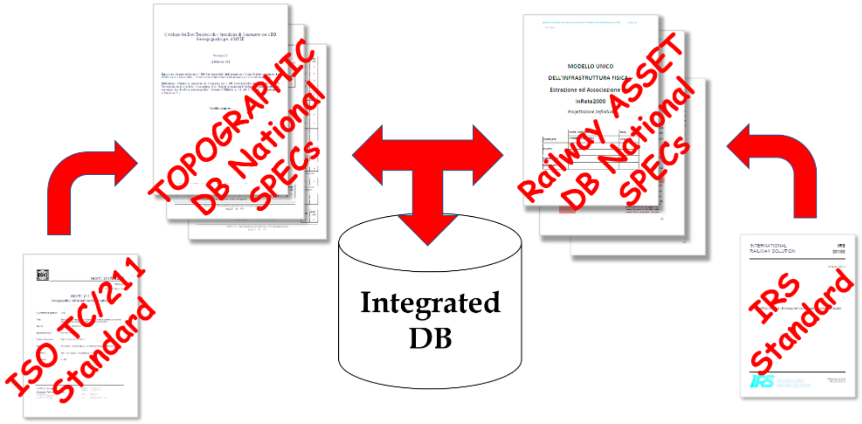


Figure 35 - content profiling based on standards

Practically, contents have been organised in a single database where specific railway objects, the assets, have been integrated with all the other topographical objects (Fig. 36).

From the geometrical point of view, to be compliant with cartographic requirements, the traditional topographic objects are generally described by spatial models focused on the representation of their footprint position. Otherwise, the assets describe railway objects that not necessarily could be represented in terms of location or their footprint. However, the geometries of assets have been forced to follow the same approach of all the geographical data of traditional cartography, for instance describing footprint spatial extents or locations.

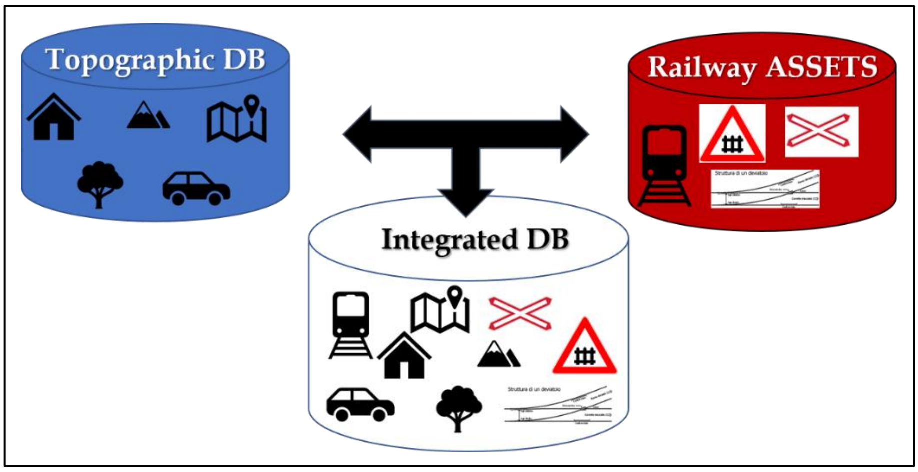


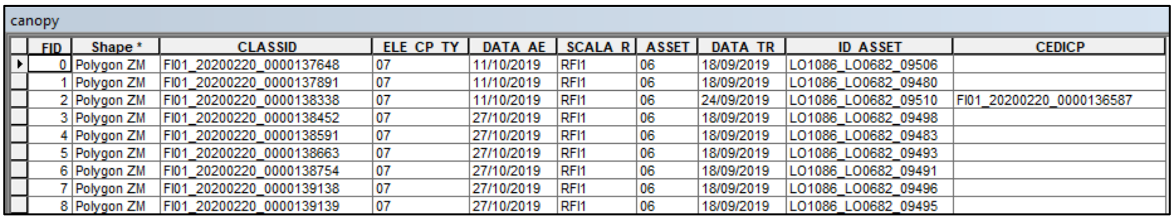
Figure 36 - integration of topographic and railway objects

From the instance point of view, the integration of assets within the GTDB has been carried out by designing new object classes, if never described in the GTDB, or new attributes of a class to add some new object properties, or, finally, by integrating code lists to insert new types (Tab. 2).

Table 2 - in green data catalogue asset integration; new classes or attributes or domains

Asset	Class	Geometry	Attribute	Domain	
Railway access point	GM_DER	Point			
SSE (power substation)	PE_UINS	Polygon	PE_UINS_TY	0602	Railway power plant
BTS (Base Transreceiver Station) – Antenna	TRALIC	Point	TRAL_IMP	09	Railway antenna
Balise	MN_INT	Point	MN_INT_TY	0901	Balise
Railway Signal	MN_INT	Point	MN_INT_TY	0902	Railway signal
Watchtower	EDI_MIN	Polygon	EDI_MIN_TY	21	Watchtower
Wall	MU_SOS	Lineare	MSOS_TY	04	Retaining wall
Noise-reduction barrier	EL_DIV	Polyline	EL_DIV_TY	20	Noise-reduction barrier
Hydraulic systems	OP_REG	Polygon/Polyline	OP_REG_TY	13	Cliff (river)
Cliff (marine)	F_NTER	Polygon/Polyline	F_NTER_TY	02	Cliff
Well (drainage)	MN_RTC	Point	MN_RTC_TY	0503	Well (drainage)
Railway Mast portal	PT_FER	Polyline			
Mast	PALO	Point	PALO_IMP	01	Railway mast
Transponder SSC	MN_INT	Puntuale	MN_INT_TY	0903	Transponder SSC
Railway bumper	MN_INT	Puntuale	MN_INT_TY	0904	bumper

Each asset is further related to its respective railway office that is in charge to properly manage it on site. For this reason, a specific external object-ID (ID_ASSET field) has been implemented into the topographic database (Fig. 37). Then, an RFI data lake will manage GTDB and integrated asset, several primary datasets (orthophotos, point clouds, DEM, etc.), and different secondary geospatial datasets (cadastral data, geological information etc.) data Lake System.



canopy

FID	Shape *	CLASSID	ELE CP TY	DATA AE	SCALA R	ASSET	DATA TR	ID ASSET	CEDICP
0	Polygon ZM	FI01_20200220_0000137648	07	11/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09506	
1	Polygon ZM	FI01_20200220_0000137891	07	11/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09480	
2	Polygon ZM	FI01_20200220_0000138338	07	11/10/2019	RFI1	06	24/09/2019	LO1086_LO0682_09510	FI01_20200220_0000136587
3	Polygon ZM	FI01_20200220_0000138452	07	27/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09498	
4	Polygon ZM	FI01_20200220_0000138591	07	27/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09483	
5	Polygon ZM	FI01_20200220_0000138663	07	27/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09493	
6	Polygon ZM	FI01_20200220_0000138754	07	27/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09491	
7	Polygon ZM	FI01_20200220_0000139138	07	27/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09496	
8	Polygon ZM	FI01_20200220_0000139139	07	27/10/2019	RFI1	06	18/09/2019	LO1086_LO0682_09495	

Figure 37 - asset external ID (ID_ASSET field)

5.6 The point clouds model

The MMS is principally focused on providing an accurate high-resolution dataset collected along the railway system. In fact, the 3D spatial information is gathered employing a laser scanner mounted on the top of the train. Since the laser scanner is moving along the railway centreline, the acquired point cloud is highly dense close to the rail line, and hence it is well suited to properly (geometrically) describe most of the assets of interest (Fig. 38). Consequently, this dataset is of paramount interest when dealing with the problem of extracting a 3D description of each asset.

Figure 38 -MMS laser scanner dataset

The mobile laser scanning dataset, acquired by the MMS mounted on the train, is integrated with telemetry data and 360° panoramic images, collected along with the MMS survey. In particular, a proper system calibration allows to associate RGB colours, provided by the panoramic images, to the laser scanning points.

5.7 Some conclusions about MUIF datasets and future updates

The MUIF project is already ongoing, half of the Italian territory has been already supplied. However, the data sources that refer to photogrammetry and laser scanners are aimed at the restitution of the MUIF geodatabase in terms of the vectorial model. For the moment, the integration efforts between geotopographic objects and railway assets concerned the database contents. So that the point clouds models obtained from train laser scanners are stored as source data but not further related to the SDBMS. Starting from the datasets thus structured for the MUIF, some analysis and elaboration of the data have been conducted towards the integration of different topological structures, towards the connection of the point clouds in the DBMS, towards the automatic classification of the assets starting from the point clouds, towards the 3D modelling. The deepening of these aspects, treated in the following chapters of the thesis, are further objectives of the thesis, to be considered in support of the future developments of the MUIF project.

6. CHAPTER 6 - GEO-REFERENCE VS GEO-THEMATIC INFORMATION SYSTEMS

6.1 The concept of a reference base and the ontology for topographic mapping.

The common point to share geographic information among different applications is related to a unique topographic base map. The critical issue in the primary integration process issue refers to harmonise corresponding instances. This semantic matching process is only possible if an object's meaning is clear (Uitermark et al., 2005). To obtain interoperability, it is necessary to make objects definition clear and make datasets semantically transparent to each other. "Any successful communication requires a language that builds on a core of shared concepts in this sense; ontology can play a crucial role" (Kuhn, 2001).

The reference model could associate concepts from a domain ontology and information from surveying rules. Then some Additional structure related to the dataset's combination could be integrated (Fig. 39). Relationships between reference model concepts and application ontology concepts define datasets semantics (Uitermark et al., 2005). In general, "in computer science, an ontology has to do with the explication of knowledge to overcome the semantic diversity of different information sources" (Visser et al., 2001). More precisely, in a domain ontology for topographic mapping, definitions for topographic concepts have to be supplied, such as "road", "railway", or "building".

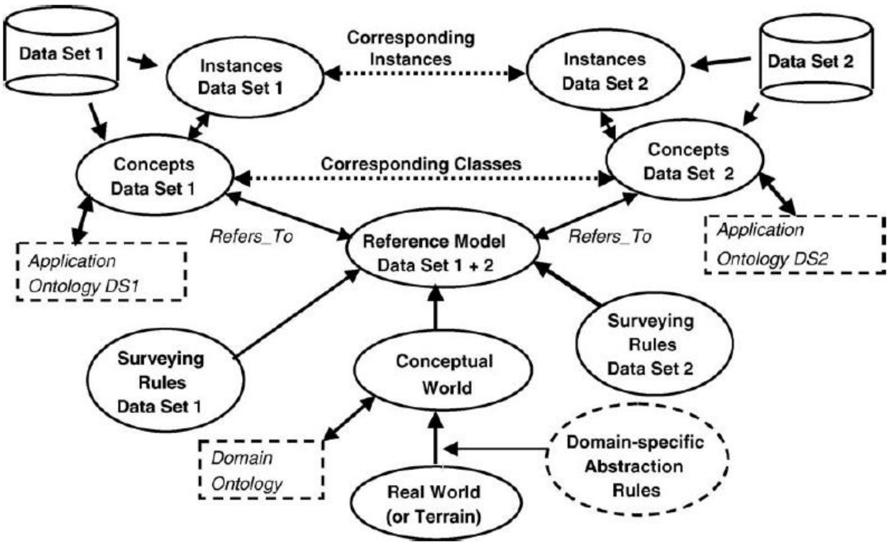


Figure 39 – "An ontology-based framework for the integration of topographic data sets" (from Uitermark et al., 2005). Arrows and text box boundaries are continuous when representing active integration steps, dashed when representing domains and dependencies of actions.

The approach is addressed to grain enough information and design a semantically rich structure so that every dataset semantic similarity became evident. For this reason, a best practice is about role defining. Depending on the role definition, a dataset class could be associated with another one with a rank of equivalent class or subclass or superclass, composite class etc. (Uitermark et al., 2005).

6.2 Reference - Thematic content connection

Thematic information is intended to communicate a single theme or a narrow set of themes. From the cartographic point of view, thematic maps are different from reference maps. The primary purpose of a reference map is to deliver geolocations and general orienteering information to the users. Geographic features tend to be represented as same as map elements. In other words, geospatial features and portrayed map elements are based on equal geometries.

The reference maps represent the territory as is, avoiding any interpretation in the representation of spatial elements. Examples of standard reference maps include topographic maps such as those created by the United States Geological Survey and Canada National

Topographic Series (NTS) where image maps can be obtained from satellites or aircraft and available through online mapping services[18].

In this context, the GTDB intended as a reference geospatial content shared among different applications is one of the main topics of this thesis. From a semantic point of view, such a model comprises (besides the spatial and graphical aspects) the ontological meaning expressed in terms of thematic classes, attributes, and their interrelationships. Objects are classified in homogeneous classes according to logical criteria and consider graphical aspects relating to modelling features given or observed in the real world.

For this reason, for example, within the GTDB the street model is based on the surface and the graph representation. Therefore, an abstract class named “Street” is related to two spatial components: street areas for cartographic scopes and element roads for network transportation analysis (Fig. 40).

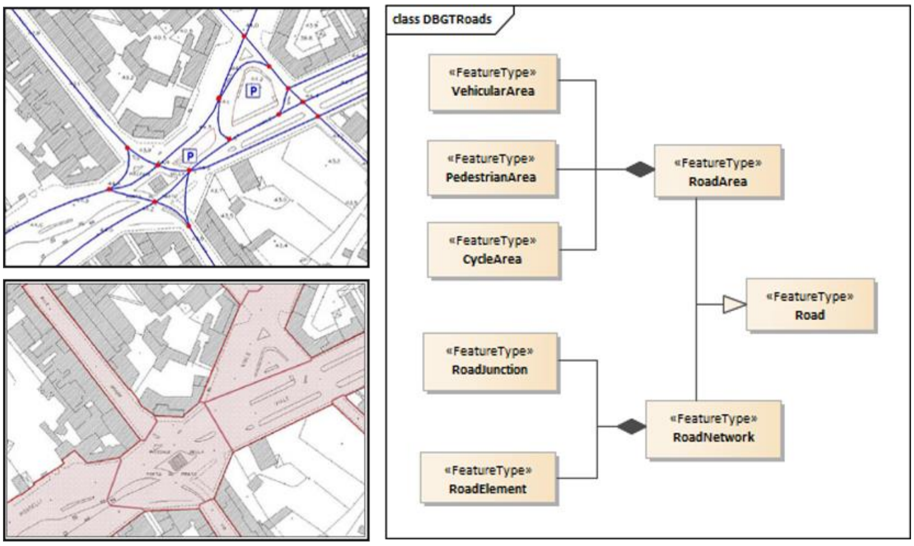


Figure 40 – DBGT road representations and UML schema

The definition of semantics and data models in the GTDB has been aimed at managing the complex geospatial database as a reference for spatial infrastructure in a relation-free approach, independently from a specific thematic application.

As a result, the multi-spatial components associated with each object class allow using the GTDB as a reference base of geographic information in thematic contexts such as railway topics. Similarly, this approach could be addressed to manage the multi-source multi-representation (Stoter J. et al., 2008) databases as detailed in Chapter 8.

---

[18] https://openpress.usask.ca/introgeomatics/chapter/thematic-maps/

6.3 Railway Infrastructures Geospatial Model

In the last decades, railway management systems aimed at cost reduction by simplifying the exchange of technical information between stakeholders. According to standards and following a machine-readable approach, sharing information systems requires avoiding non-standard data conversion and removing inefficiency due to misinterpretation.

There is an urgent need to reduce the data exchange process's complexity by describing and collecting interlocking data in a machine-readable format, standard, and correct (Bosschaart et al, 2015). Moreover, there is a need to describe railways at different level of details, to be suitable for different specific purposes (Hlubuček, 2017).

As a transportation network, railway infrastructure should be described considering assets located along the track centreline and considering the network according to the graph theory.

This family of models considers the railway network's topological structure as a graph where nodes are logical objects corresponding to track elements such as signals, track detection sections, switches, etc.

So, elements are modelled as the nodes, whereas the edges in the node-edge diagram represent the physical connections between these interconnected net elements (Hlubuček, 2017) connected in a route. Therefore, a route is a sequence of infrastructure elements, track detection sections, movable track elements (e.g., switches), and signals set and locked once a train travels from one signal to the next. Then routes are dynamically determined by algorithms. Moreover, it is possible to define a signal aspect with an edge between successive signal nodes. The edge can have attributes to reflect the signalled Automatic Train Protection (ATP) speeds. This kind of representation is used by the graphical specification language EURIS (European Railway Interlocking Specification). Interpret the plans and or the files, manually extract the data, and convert them to a specific format readable by their engineering tools (Bosschaart et al, 2015).

6.4 Standards on Railway network model

6.4.1 The geographic standards on Railways

The reference for modelling thematic information is represented by the INSPIRE Directive (Directive 2007/2/EC, 2007). The reference railway data model is detailed on "Data

Specification on Transport Networks (INSPIRE D2.8.I.7, 2014), where integrated transport network within each national border is detailed.

In this standard, transportation data includes topographic features related to road, rail, water, and air transport. “Data specification provides a coherent approach to the forms of the representation (physical topographic area objects or centreline representations) and consistency between data sets, the latter as different types of coherence (between spatial objects of the same theme at different levels of detail, between different spatial objects within the same area or coherence at state boundaries)” (INSPIRE D2.8.I.7, 2014).

To ensure a consistent approach across all network themes, Inspire Standards refers to the INSPIRE Generic Network Model (GNM) Directive (INSPIRE D2.10.1, 2013). All the spatial data sub-themes have to follow the indications of that standard to be shared by any spatial data theme network (e.g., rail).

The GNM model describes different aspects: first, the network connection mechanisms, establishing the cross-border connectivity or the intermodal connectivity, then object external referencing to support the reuse of information; finally, define the linear referencing to support and link the different transport properties to the transport elements.

In turn, the GNM model is compliant with ISO geomatic series standards about specific aspects of interoperability to be provided explicitly.

Therefore, the reference standard based on LRS (ISO/TC 211 19148, 2012) requires fulfilling intelligent transport systems such as Location-Based Services (LBS).

About the geometrical and topological model of Transport networks are concerned (INSPIRE D2.8.I.7, 2014), the conceptual schema defines locations related to a one-dimensional object as the measurement along (and optionally offset from) that object. . It describes the data and operations that need to use and support linear referencing. It is suitable for transportation, utilities, location-based services and other applications that define linear objects’ locations.

The data specification includes three types of geometries: the first one is about topographic areas, the second one refers to the centreline and the third define objects represented as points. The first two may be alternative representations of the same real-world phenomena dependent on the level of detail you need to model. The third type only included in the marker posts’ specification in addition to the network nodes. Generally, the representation type is based on 2D vectors.

Moreover, other specific definitions refer to topological characteristics: “Topology is handled in the data specification implicitly rather than explicitly, mainly to keep the model as simple as possible since it is expected that most applications will use the network data within a topological environment. Consequently, there is the prerequisite for “implicit topology”, where the data provided must be sufficiently clean and capable of automated topological construction within a user’s application” (INSPIRE D2.8.I.7, 2014). This concept is framed by specific requirements, including data quality information.

Generally, network models only represent the network's 2D topography focusing on the logical network connectivity information. A 3D geometry representation could be associated with multi-patch features; however, these would not be coupled to the network modelling visualized as pure 3D.

Each domain data model represents a commodity-specific 2D GIS-based abstraction of the real world's respective rail network. Hence, the GNM cannot be used efficiently for a different type of rail network, and thus no common model/database integrating different models is provided. Nevertheless, this standard allows network tracing and some specific spatial analysis related to the network (Adolphi et al. 2013).

6.4.2 The thematic standards on Railways

Railway infrastructure information needs to share their data internally across an organization and externally between organizations on the strictly thematic point of view. Several data formats exist focused on both rail and non-rail aspects. Each design is primarily performant to requirements specifically related to a given discipline. Since the iron network and its elements are similar in every country, there is no need to reinvent it.

Standards in the railway sector are compiled for more than 100 years and collected by the International Union of Railways (in French "Union Internationale the Chemins de fer", in short UIC). The UIC is the largest railway organisation worldwide, delivering specific railway regulations called International Railway Standard (IRS). The most relevant one is the RailTopoModel (RTM), which become an IRS standard in spring 2016 under the name IRS 30100 (IRS 30100, 2016). This standard is intended to be used in all business processes dealing with the design, construction, operation, and maintenance of a railway network. Moreover, this standard focuses on quickly exchanging and storing data unambiguously, and avoiding errors among different business processes.

About contents, the RTM specs describe the railways business objects universally, thus independently of usages (usage-agnostic). Structural modules refer to topology, external references, infrastructure (i.e. bridges), signalling, life cycle properties, and others.

The RTM abstracts the underlying necessary concepts in the form of a UML 2.0 class diagram.

The railway topology represents the essential module where concepts have been described a generic model, in such a way that it applies to any aggregation level in which a railway network may be represented.

Consequently, the topology is directly or indirectly related to each object class according to their appropriate aggregation level.

Railway objects also are featured based on several physical characteristics and located according to different kinds of positioning methods.

The overview below (not following UML conventions) introduces RTM's main objects and dimensions (Fig. 41).

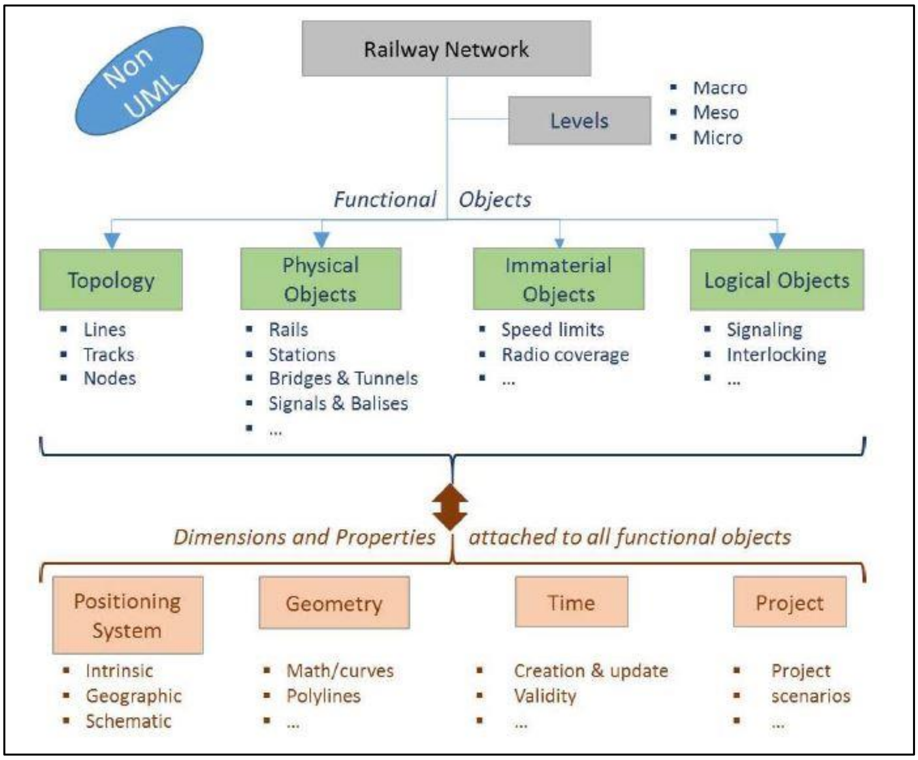


Figure 41 - Functional coverage of the RailTopoModel (from IRS30100, 2016)

Compared with ISO Standards (TC 211 191xx series) described in the previous paragraph (see paragraph 6.4.1.), it is worth noting that the RTM standard complements that ISO one by specifying semantics and providing relevant functionalities in the railway context. So, the RTM standard facilitates the implementation of infrastructure management information systems; it includes the geographic dimension natively, and therefore fulfils, inter alia, the INSPIRE Directive requirements (Directive 2007/2/EC, 2007) when these requirements apply to railway infrastructure. No current ISO 191xx series standard deals with the independent-scale requirements needed to model the railway infrastructure since these ISO standards stand at a higher level of abstraction.

Moreover, some other content tasks of the RTM standard are:

- the concept framework supports the railway infrastructure description about the iron-type and business elements such as infrastructure assets, referencing and positioning, behaviours and relationships among objects.
- The structure and topology network are described at various levels of detail, depending on usability and data availability. Various stakeholders mainly use it for purposes not precisely known in advance, e.g. network design and maintenance, traffic

scheduling, and traffic management. Different details can be described; each object can be detailed at line level, track level by its physical component description such as switches, signals, or balizes. Several properties can be associated with each class: properly conformity assessing, technical characteristics explaining, life cycle data defining, economic aspects describing, and others.

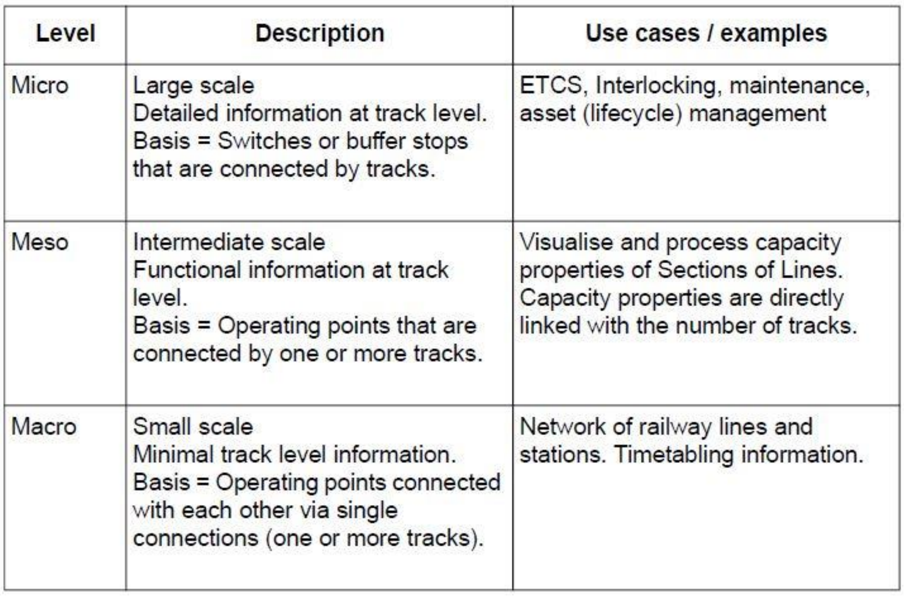
The railway network can also support associated events since the model is based on a graph theory as far as topology is concerned. Hence, the iron network is provided by a topological representation although schematically visualized. Any detailed level of the track locations, from corridors down to tracks, is supported and displayed.

Moreover, permitted routes can be selected from the model, thanks to network topology and related events querying, such as track possessions, power supply characteristics, signalling assets, etc.

.

Finally, the model supports multiple referencing systems, thus ensuring consistency during the transformation from one referencing system to another.

The standard defines three LODs: Micro, Meso, Macro (Fig. 42)



Level	Description	Use cases / examples
Micro	Large scale Detailed information at track level. Basis = Switches or buffer stops that are connected by tracks.	ETCS, Interlocking, maintenance, asset (lifecycle) management
Meso	Intermediate scale Functional information at track level. Basis = Operating points that are connected by one or more tracks.	Visualise and process capacity properties of Sections of Lines. Capacity properties are directly linked with the number of tracks.
Macro	Small scale Minimal track level information. Basis = Operating points connected with each other via single connections (one or more tracks).	Network of railway lines and stations. Timetabling information.

Figure 42 - RTM LODs (from IRS 30100, 2016)

For the connection with the topographic environment of the railway, only the Micro-level has been considered.

The RTM representation in UML consists of four packages: Base, Topology, Positioning Systems, and Net Entities as shown in Fig. 43.

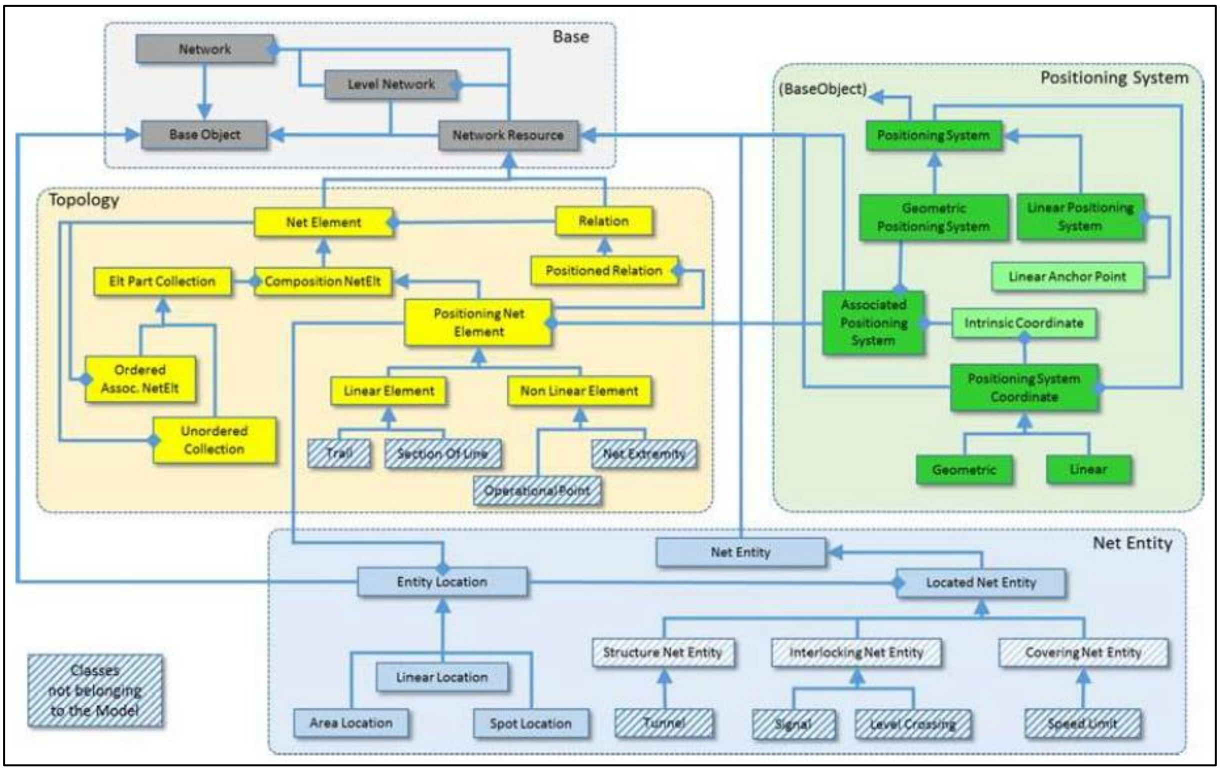


Figure 43 - RTM class diagram (from IRS 30100, 2016). It is organized into four packages: Base package contains base object classes related to network concepts and their levels; Topology package contains classes that define topological relationships; Position System define the reference coordinates based on LRS; Net Entity define objects according to graph theory.

There exists the RTM implementation level, developed by the organisation RailML.org[19], exactly called RailML. RailML is based on the XML language (eXtensible Markup Language) to simplify data transfer through a common data structure. Indeed data files include data and descriptions of the data they contain (Nash et al., 2004). Therefore RailML is limited to an interchange format always within the rail sector, so not yet implemented in data translation software such as FME (Feature Manipulation Engine) [20]. For this reason, RailML has not been considered as a step of interoperability of this thesis.

6.5 Railway Geospatial Model: Graph topology vs Spatial topology

The case study datasets include semantic modelling of railway assets based on the RTM model in the context of geospatial data where both assets and topographic contents have been integrated into the GTDB. So the connection between reference and thematic GIS has been solved thanks to the integration of contents directly. Therefore, from a technical point of

---

[19] https://www.railml.org
[20] https://www.safe.com/

view, GTDB and RTM's topology structure are profoundly different. In the first case, its topology refers to a CRS based on X, Y, Z coordinates of geometry vertices, while in the second one topology refers to the LRS along with the network. As it is declared in standards such as the GDF for network transportation (ISO 14825, 2011): "it is not allowed the use of topologies, i.e. a GDF Layer can only contain data from one type of topology".

It means that Spatial and Linear systems are not compatible from the topological point of view. So data have to be modelled in different datasets even in the same DB. It means that the direct spatial constraints between the railway network's topologies and the spatial DB could not be defined. The chosen solution has been addressed to combine assets both belonging to the Topographic and Railway DBs considering its higher granularity, i.e. "railway element" and "railway junction" classes into the integrated GTDB where topological constraints could be validated, as a basic link between models (Fig. 44).

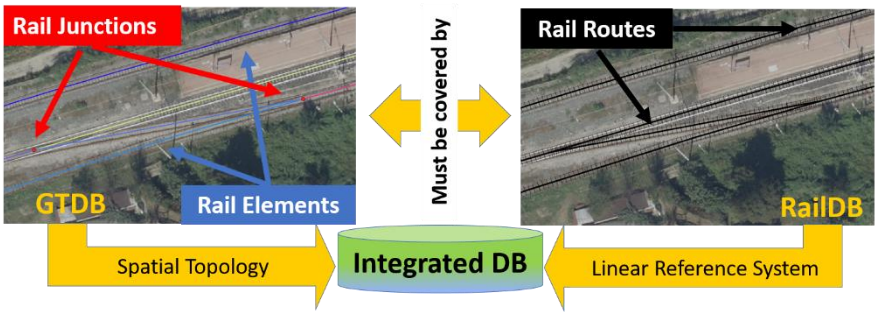


Figure 44 - rail route network base on source element of the GTDB

Hence, the approach to obtain integration in the GTDB has been addressed on the modelling of elementary objects (rail junction and rail element) as source objects that will be used for the implementation of routes based on LRS. A topological constraint "must be covered by" between source objects and rail route network has been imposed.

7. CHAPTER 7 - POINT CLOUDS IN SDBMS.

7.1 Overview and objectives

In addition to the use of point clouds as a source for vector information extraction, recently the direct management of point clouds in DBMS environments became a quite hot research topic. Some experiences suggest the translation of point clouds in point 3D spatial features (Zlatanova, 2006). Some others define a methodology for patching and indexing point clouds in an SDBMS to relate them with polygon areas (Ramsey, 2014) or simply use a collection of files and specific tools to select, analyse, manipulate, and visualize the point cloud data. The latter approach aims at managing point clouds as-a-whole. The spatially discrete point cloud representation plays in 3D a similar role to that of rasters (and orthophotos) in a 2D context. Both raster data and point cloud data are often quite massive, but relatively static. (Oosterom et al., 2015).

The leitmotiv of this thesis is that of ensuring interoperability, avoiding translation, and loss of information. In agreement with this idea, some efforts have been made to properly relate basic vector features in the SDBMS representing an asset with the corresponding points in the cloud. Hence, this chapter describes a methodology to effectively use point cloud objects in an SDBMS. Two main steps have been considered, summarised as follow:

1. the first step focuses on determining the connection between the acquired 3D points of the specific objects and the corresponding geographic features (ISO/TC 211 19107, 2002) in the SDBMS;
2. the second one implements, just in a preliminary testing phase, the automatic extraction of candidate objects of interests (i.e. assets) from the point clouds. This part must be considered as a starting point of the following developments (Chapter 9), out of this thesis.

7.2 The point clouds-SDBMS connection: methodology and procedure

For what concerns the first step, an Object-Relational database approach (Tupper, C. D., 2011) has been considered to properly store and represent topographic objects in an SDBMS. Each object is described by one or more reference 3D geometries. A specific geographic feature, which for instance includes the object footprint, is described by its 3D geometries, namely the “spatial component”. One or more segmented point clouds can be associated with each instance of a spatial component and linked to such component through hyperlinks.

Since the object spatial component-point cloud is a many-to-many relationship, when new geometric information on an object is available, for instance after a new survey, the object spatial component can be updated by adding the proper segment of the newly acquired point cloud. This approach enables also the management of spatial-temporal dimensions as different LoDs.

More details on vector models and hyperlinks are provided later, but it is worth noticing that at each update there is one reference spatial component that participates in relationships and topological constraints with all other database objects. Nevertheless, the relationship between geographic features and point clouds could be recalled on-demand, using the hyperlink. It happens if you need to give more detailed information about an object of interest, as well as if you need to model according to the full 3D Boundary Representation (B-Rep) approach (Shapiro V., 2002). Moreover, it is possible to associate new point cloud spatial information from cyclical monitoring systems, during the time. Finally, this process allows using segmented point clouds in an SDBMS as classified spatial components of objects.

To make the point cloud segmentation process as feasible as possible for each railway asset type, a specific procedure has been set up, implemented in a Python 3.7 environment, Conda ecosystem and Numpy, Whitebox, LasPy modules.

The asset geographic features stored in the database can be extracted from the point cloud by segmenting the portion that describes each asset: a topological overlay of the point cloud with the asset footprint, derived from its features, is performed to such aim.

To be more specific, the simplified geographic features of each asset, along with its geolocation, are stored in the GTDB. Depending on the considered asset, its associated geographic features can be either points, lines, or polygons (Fig. 45).

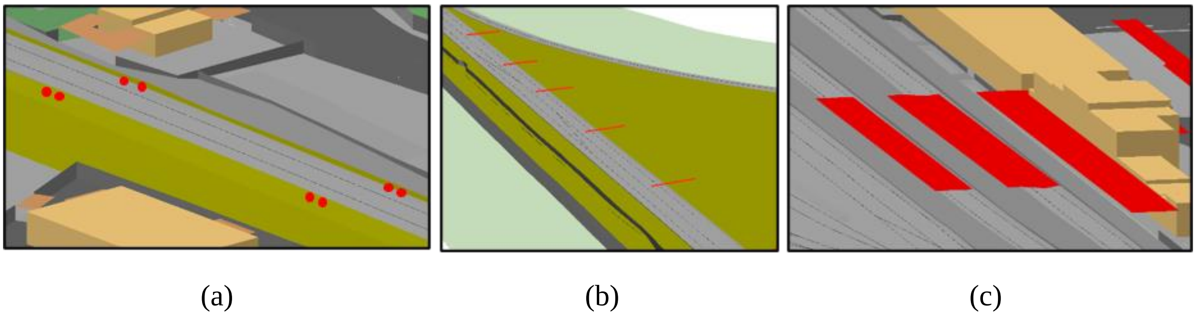

(a) (b) (c)

Figure 45 - (a) point geographic features (red points), (b) line geographic features (red lines), (c) polygon geographic features (red polygons)

In the first two cases (Figure 9(a), Figure 9(b)), a 2-meter buffer on the horizontal coordinates of the asset features is used to select from the cloud the points describing such asset (Fig. 46).

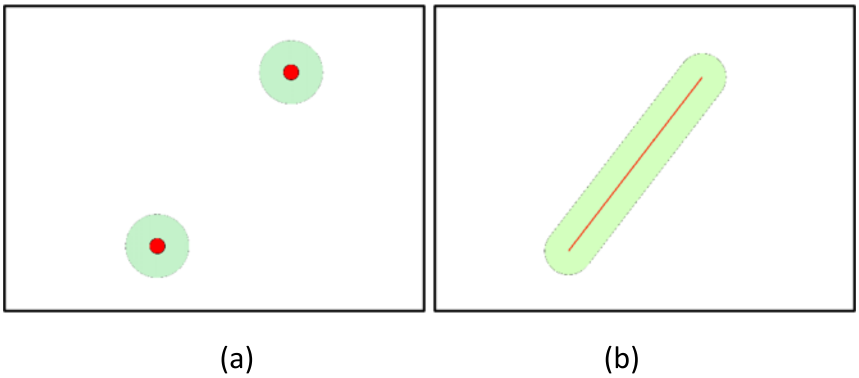

(a) (b)

Figure 46 - (a) 2 m point buffer (green polygons), (b) 2m line buffer (green polygon)

Instead, since polygons already represent the footprint extents of topographic objects, they can be directly used for the asset point selection.

A SHP file is created for each asset, summarizing its geometric characteristics as described above (i.e. its footprint, either expressed as its polygonal boundary or its 2-meter buffer).

Then, a dictionary representing the relationship table from the LAS file of each train sub-clouds (approximately 1 km long) to the SDBMS assets is defined through the following operations:

- A unique ID is associated with each asset SHP file, named hereafter ID_ASSET.
- For each LAS file, all the assets within the spatial extent of the area described by such file are properly determined and reported in a table, as shown in Fig. 47. To be more precise, the search for each asset to be reported in such a list is done by checking the spatial overlay between the area described by the LAS file and each asset boundary region, which is summarized in the associated SHP file. Since there typically is an overlap between the spatial extents of the LAS files, an asset might be reported in the table of more than just one LAS file.

	Type	Size	Value
2019Carro06TLM_206_113.las	list	50	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_112.las	list	48	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_110.las	list	46	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_109.las	list	52	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_107.las	list	56	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_106.las	list	48	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_103.las	list	50	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_104.las	list	49	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_100.las	list	50	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_116.las	list	47	[[illegible]DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_114.las	list	72	[illegible]ATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_117.las	list	20	[illegible]C:\[illegible]TI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_115.las	list	66	[illegible]I\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_111.las	list	4[illegible]	['C:\DA[illegible]\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_105.las	list	50	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_101.las	list	50	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_108.las	list	51	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…
2019Carro06TLM_206_102.las	list	49	['C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682…

Index	Type	Size	Value
0	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_00781.shp
1	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_00782.shp
2	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_01013.shp
3	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_01014.shp
4	str	1	C:\DATI\CASE STUD[illegible]mast\LO1086_LO0682_01015.shp
5	str	1	C:\DATI\CASE STUD[illegible]FI[illegible]mast\LO1086_LO0682_01153.shp
6	str	1	C:\DATI\CASE STUD[illegible]FI0[illegible]ast\LO1086_LO0682_01745.shp
7	str	1	C:\DATI\CASE STUD[illegible]FI[illegible]ast\LO1086_LO0682_02153.shp
8	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_02337.shp
9	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_02338.shp
10	str	1	C:\DATI\CASE STUDY\FI01\mast\LO1086_LO0682_02339.shp

Figure 47 - Dictionary set up: table A represent the list of LAS files, table B represent the list of SHP assets connected with each raw of table A. Between table A and table B a one-to-many (1:m) relationship there exists

The resulting dictionary is used to select and extract, for each asset instance, the 3D points within the footprint (which can be easily obtained by the previously created SHP file) of that geographic feature. The selected points are saved as a new LAS file, with the file name encoded as follows:

XXXX__YYYY.las

where XXXX is the asset ID ID_asset, and YYYY is the name of the source LAS file.

For example, LO1086_LO0682_10411__2019Carro06TLM_206_108.las is the segmented LAS file for an asset mast portal where XXXX="LO1086_LO0682_10411" is the ID_ASSET, whereas YYYY="2019Carro06TLM_206_108" is the name of the original LAS file containing the 3D data collected during such specific train run.

Hence, considering both the updates over time and the overlap between sub-run sequences, the connection between geographic features and segmented point clouds is a many-to-many (n:m) relationship, as formalised in UML language (Fig. 48). To explain the n:m relationship a joint table has been defined (JOIN_ASSET_LAS) splitting the n:m relationship in two: the first one a is 1:n relationship between ASSET features and JOIN_ASSET_LAS table, the second one is m:1 relationship between the JOIN_ASSET_LAS table and the PC_LAS code list.

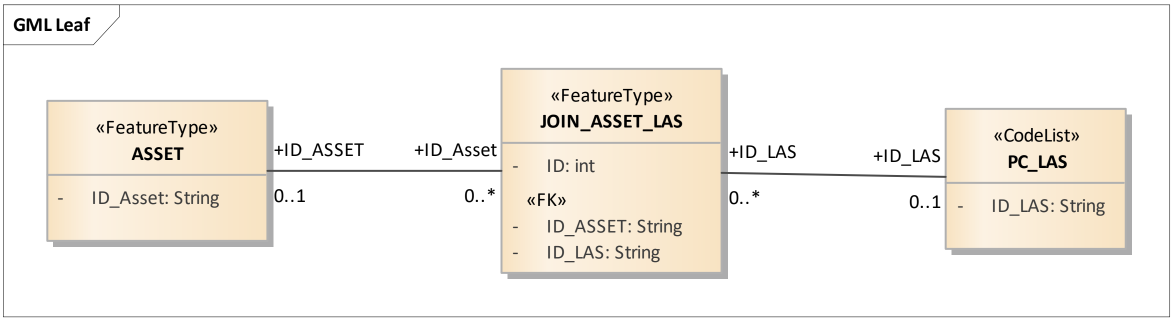


Figure 48 - many-to-many (n:m) relationship between geographic features (ASSET) and LAS point clouds (PC_LAS) though a join table (JOIN_ASSET_LAS)

This approach keeps the connection updated between geographic features in an SDBMS and a more detailed 3D modelisation. As a result, monitoring of railway assets can be carried out over time. Moreover, a full 3D modelisation of each asset can be quickly recalled on-demand by means of the connection through simplified vector geographic features stored in the DB, hence, hence avoiding an onerous extraction via B-Rep of each asset, which is quite unsuitable due to the high LoD of the source point clouds.

Some examples of topographic objects as point, linear, and polygonal features are reported below.

1. Single mast represented as a point on the ground (red points) and the corresponding extrusion results (light blue line) (Fig. 49);

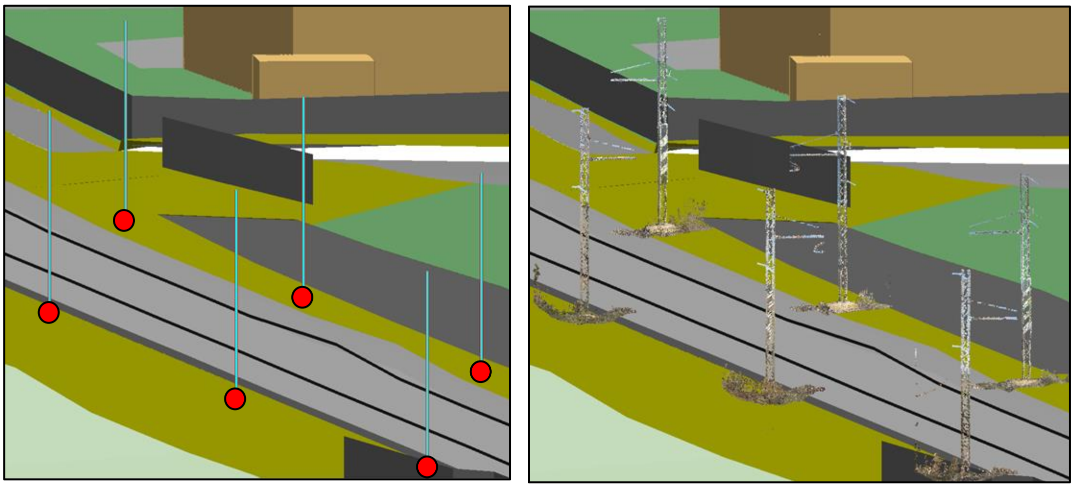


Figure 49 - asset represented as point (in red) geographic features in the SDBMS on the left image, as segmented point clouds on the right image

2. Mast portal represented as a line (red line) (Fig. 50);

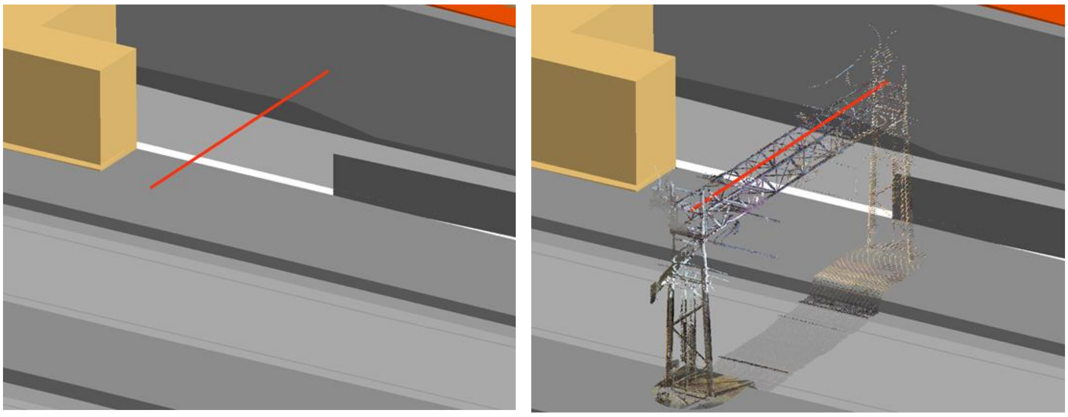


Figure 50 - asset represented as line (in red) geographic features in the SDBMS in the left image, together with the segmented point clouds on the right image

3. Canopy station represented as a polygon (light blue polygons) (Fig. 51)

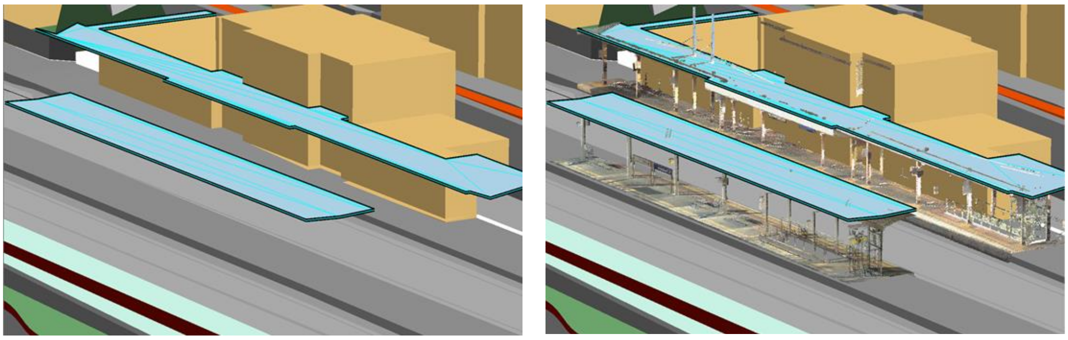


Figure 51 - asset represented as line (in red) geographic features in the SDBMS in the left image, together with the segmented point clouds on the right image

7.3 Some conclusions on the point clouds - feature connection

This research work has been motivated by the need for properly managing in a common way survey data coming from different sources: airborne photogrammetry, laser scanners, drones etc. Since each data source typically uses a specific spatial data model, often there are not enough connections between them: conformity of processes involving heterogeneous data is typically not guaranteed, and the interoperability can be ensured at the cost of some information loss. Instead, the proposed approach maintains the structure of the source datasets (point clouds), whereas the connection with vector features has been obtained by properly segmenting and linking the point cloud data to the objects in an SDBMS.

As a result, the following advantages can be obtained:

- sustainability of territory monitoring over time through continuous updates: the relationship between a feature class and the segmented point clouds allows to associate a single spatial component to different updates over time;
- directly relating source to feature data, avoiding any translation of the first one, allows associating a basic reference spatial component to each asset. Thanks to the hyperlink, it is possible to quickly retrieve a higher level of detail in an extemporaneous query. Moreover, the connection allows to model asset according to a 3D B-Rep requirements, even after the acquisition;
- the adopted spatial model pertains both to a repository SDBMS and a management SDBMS (that could change over time according to process analysis) requirements;
- The considered approach extracts and stores in the database only information of interest, avoiding the use of point clouds “as-a-whole”. Direct consequences are:
    - drastic reduction of the storage requirements;
    - focusing only on the object of interest;
- the hyperlinks between features and LAS files allow effective management of point cloud data in an SDBMS.
- “near real-time” procedures for recognising new assets in newly acquired point cloud datasets can be particularly interesting in monitoring applications. More details on this subject will be provided in the next paragraphs, along with a short description of the rudimental procedure currently implemented to this aim.

7.4 Point clouds classification and procedure overview

This part of the thesis is based on a test conducted during my PhD course and presented in the paper (Corongiu et al., 2020).

Regarding the second objective, segmented point clouds are classified, as ground truth, to train the algorithm to automatically extract object of interest from new acquisition by laser scanning in the same or different areas (temporal and spatial update). Hence, the advantages refer to monitor existing assets and relationships between a simplified feature and point clouds over time. Training is carried out of a Convolutional Neural Network (CNN - Krizhevsky et al., 2012) to automatically detect new assets to be updated in the SDBMS in near real-time.

The implemented procedure aims to automatically detect objects of interest from point clouds provided by the train and backpack mobile mapping systems, extracting the corresponding position and semantic information to be inserted in the above-mentioned

SDBMS. Fig. 52 shows a case study example of point clouds acquired with the mobile mapping system mounted on a train.

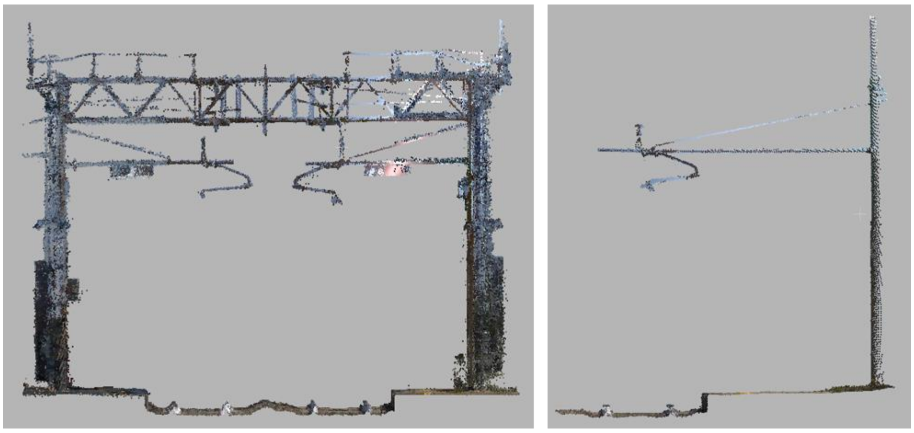
Figure 52 – "Example of train mobile mapping point cloud acquired in the proximity of a station" (Corongiu et al., 2020)

This preliminary test is addressed to demonstrate how the SDBMS railway assets could be updated in near real-time by automatic classification of point clouds.

The adopted procedure uses the ground truth to train a CNN to automatically detect railway assets from point clouds.. This aspect is focused on update new assets in the SDBMS.

The extraction of railway objects from point clouds data has already been considered in the literature by several authors: (Neubert et al., 2008) considered the detection of rail tracks, catenary, and contact cables from orthophotos, whereas (Elberink & Khoshelham, 2015) extracted such information from points clouds by using template matching techniques (Arastounia & Oude Elberink, 2016). Similarly to this case study, some other authors exploited mobile laser scanning systems (Pastucha, 2016) or airborne LIDAR (Arastounia, 2017) for the data acquisition.

The implemented solution, first of all, considers a pre-processing step, aimed to properly select a set of candidate objects then fed as inputs of a deep learning classifier. In particular, each candidate is selected considering its point cloud local property. Then, eigenvalue-based segmentation (Maalek et al., 2018) is used to discard ground and vegetation points not related to the object of interest. CNN uses the modified Fisher vectors to classify objects (e.g. masts) from point clouds according to the methodology proposed by (Ben-Shabat et al., 2018). Fig. 53 shows two asset types such as (a) portal and (b) cantilever, to be detected from the point clouds. The extracted semantic information is compared with that one already stored and validated in the MUIF database (Tucci et al., 2020).

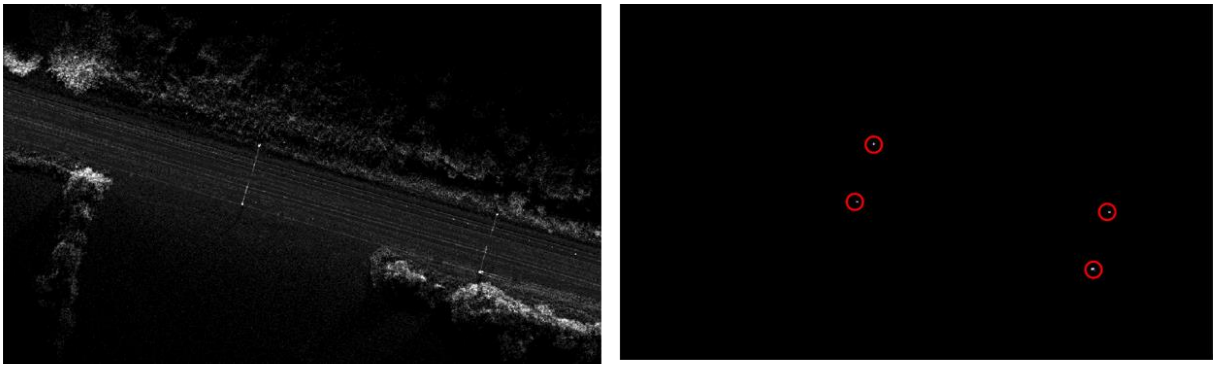


(a) (b)

Figure 53 - Examples of objects to be classified (rails are included to ease the readability of the figures on the top).

7.5 Candidate extraction

The test phase considers, in particular, the problem of detecting cantilever masts and mast portals. A simple procedure extracts a set of candidates to be tested. this step is mainly motivated by the need of reducing the computational burden related to the execution of the classification step via neural networks. In the case of cantilever masts and mast portals, it worth noticing that are typically object quite high, so the candidate extraction procedure is mostly based on the identification of such areas with a high planar point density (Fig. 54(a)), and with quite high differences in altitudes between points in such areas. Then, connected components are computed and the centroid of each connected region is considered as a potential object of interest candidate (Fig. 54(b)).



(a) (b)

Figure 54 - Example of candidate extraction: (a) planar point density (b) extracted candidates.

7.6 Pre-processing

To prepare objects to feed into the neural network classifier, a pre-processing phase is performed. For each candidate, the pre-processing workflow is as follows:

- a subset of the point cloud extraction in the neighbours of s candidate (e.g. based on a buffer cylindrical zone of 2 m radius, centred in such location).
- subset local reference system changing to align the coordinates with the railway track direction (Fig. 55).
- Detection and discard of points not related to the objects of interest, such as vegetation ones (Fig. 56, 57, 58).

The detection method is based on the eigenvalues of the covariance matrix of the neighbourhood of a 3D point (Maalek et al., 2018, Weinmann et al., 2015). Then, the three eigenvalues of such covariance matrix, called $\lambda_{1,}$ $\lambda_{2,}$ λ_3 be, and e_i, *i* = 1, 2, 3, as the normalized version of the eigenvalues, i.e. $e_i = \sum_{j=1}^{3} \lambda_j$. Ere considered. Next, similarly to (Weinmann et al., 2015), the following 3D features are used to summarise the point geometrical characteristics: linearity L_λ, planarity P_λ, scattering S_λ, omnivariance O_λ, anisotropy A_λ, change of curvature C_λ:

$$L_\lambda = \frac{e_1 - e_2}{e_1} \quad (1)$$

$$P_\lambda = \frac{e_2 - e_3}{e_1} \quad (2)$$

$$S_\lambda = \frac{e_3}{e_1} \quad (3)$$

$$O_\lambda = (e_1 e_2 e_3)^{1/3} \quad (4)$$

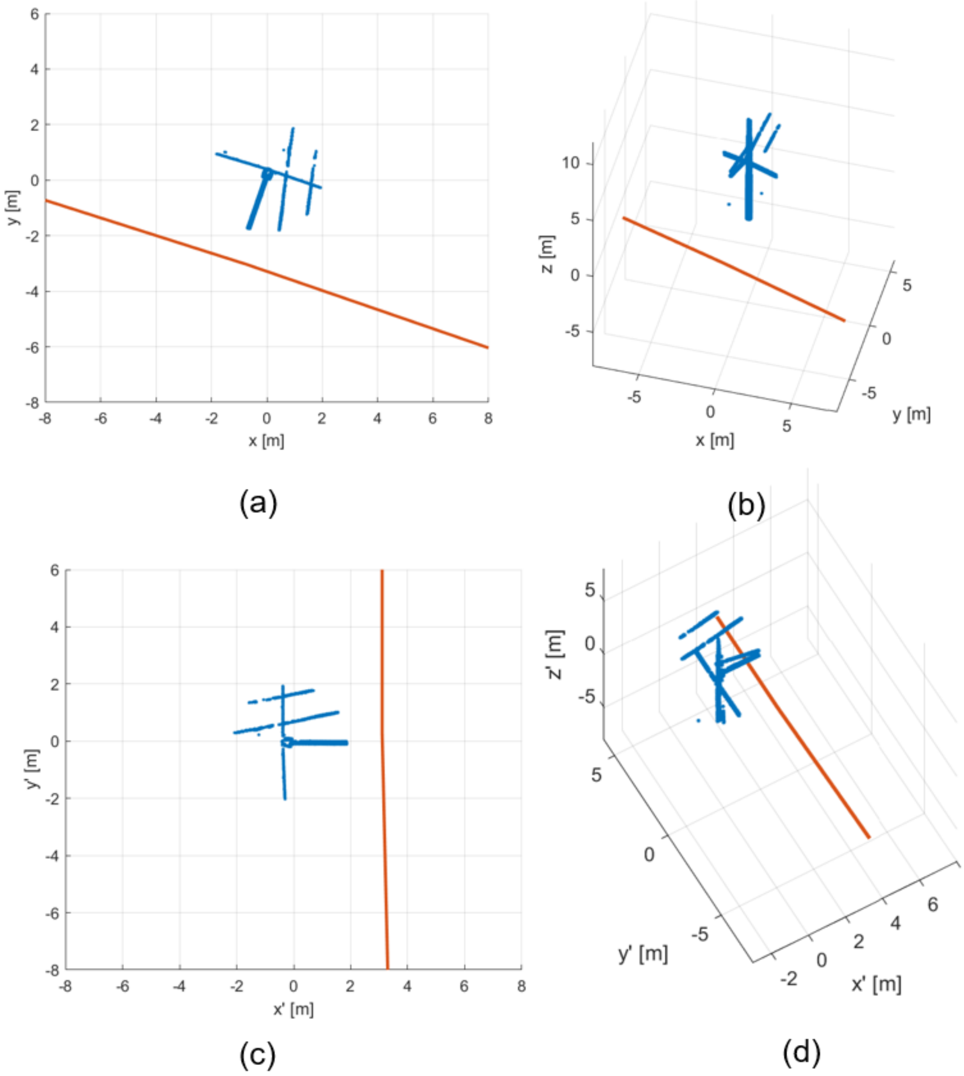
$$A_\lambda = \frac{e_1 - e_3}{e_1} \quad (5)$$

$$C_\lambda = \frac{e_3}{e_1 + e_2 + e_3} \quad (6)$$

Such 3D features are used to discard from the objects of interest point related to vegetation, walls, and other not interesting objects. Just to give some examples, ires and other metallic objects in parts of masts and portals could easily be identified as linear features, whereas the upper part of the walls could be represented by planar features.

The overall computed 3D features are used as input of a Support Vector Machine (SVM) classifier, which aims at separating not interested object points. SVM classifier is trained on

9458 randomly sampled pre-classified points. About accuracy, (true positives + true negatives) in relation with the number of samples, on the training set was 93.5%, whereas, on a validation set of about 100 k samples, the accuracy was 93.2%.



(a) (b)

(c) (d)

Figure 55 - Example of point cloud pre-processing: mast (blue)and railway track (red line). (a) and (b) original data, (c) and (d)local reference system with y′ coordinate aligned with the railway track.

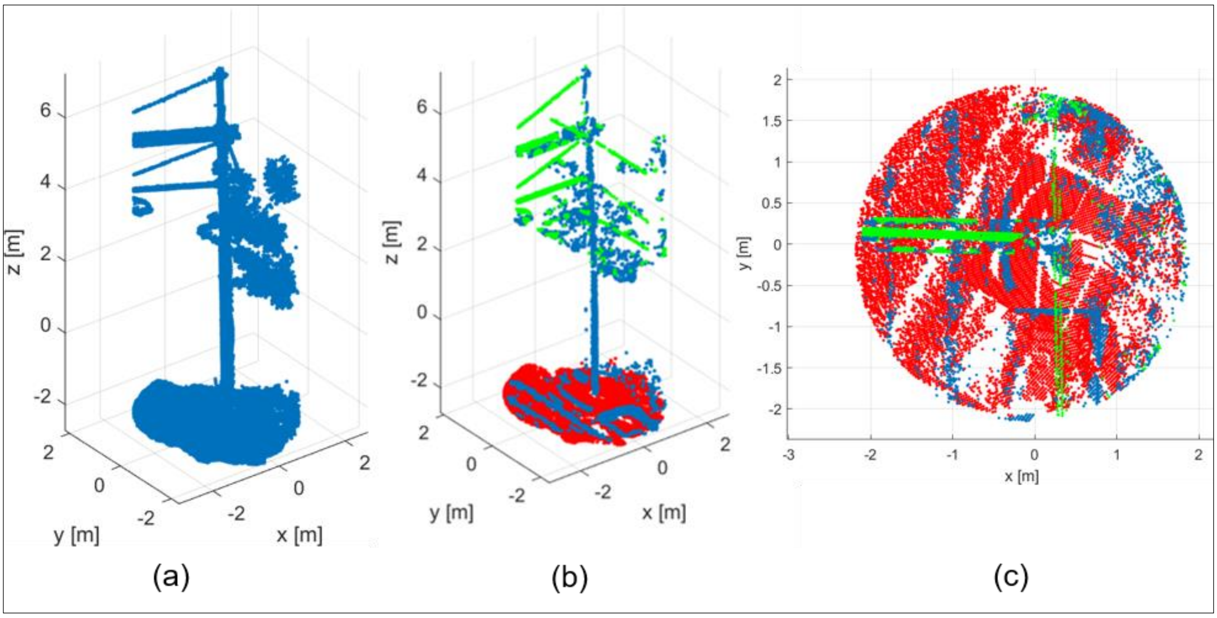


Figure 56 – "Example of point cloud pre-processing: (a) original data, (b) segmentation side view, mast (blue), cantilever and wires (green), ground (red), (c) segmentation re-oriented top view." (by Corongiu et al., 2020)

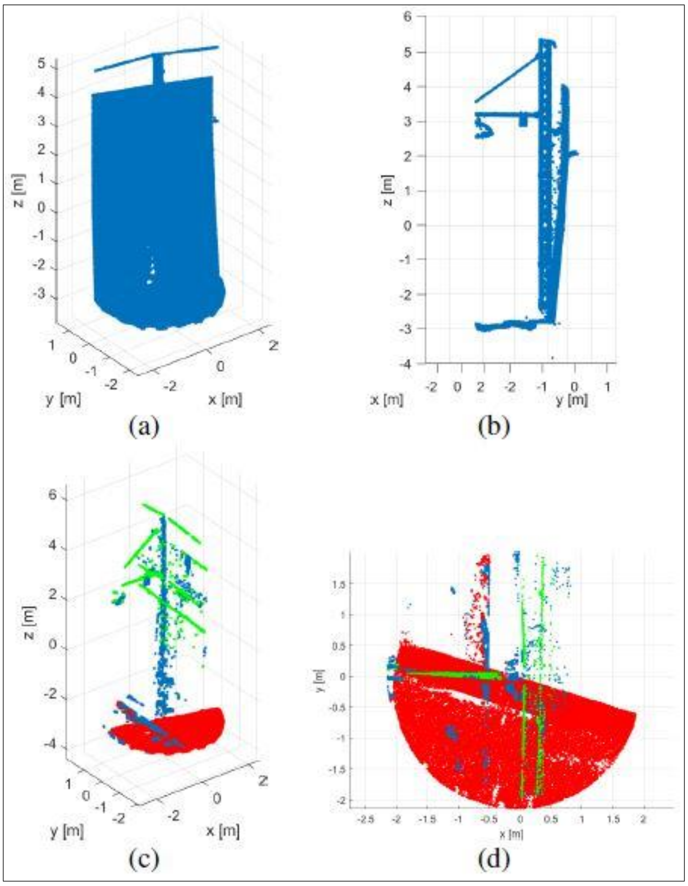


Figure 57 - Example of point cloud pre-processing (discarding wall): "(a) and (b) original data, (c) segmentation side view, mast (blue), cantilever and wires (green), ground (red), (d) segmentation re-oriented top view" (by Corongiu et al., 2020)

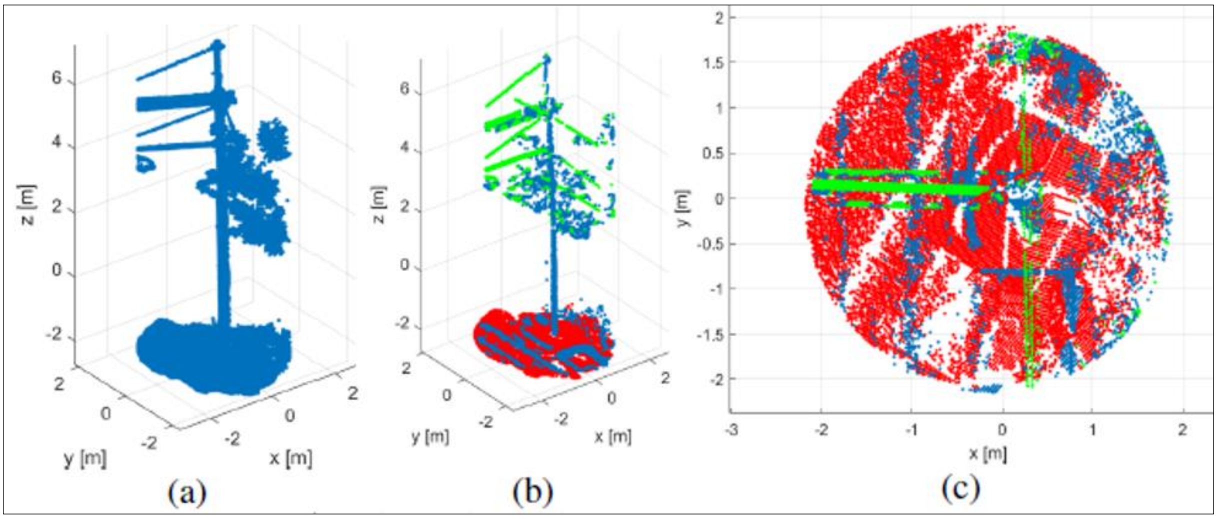


Figure 58 - Example of point cloud pre-processing (vegetation elimination): "(a) original data, (b) segmentation side view, mast (blue), cantilever and wires (green), ground (red), (c) segmentation re-oriented top view." (by Corongiu et al., 2020)

7.7 Classification

The classification step is based on a deep learning approach, as an adaptation of the 3D modified Fisher vectors (3DmFV) approach proposed in (Ben-Shabat et al., 2018). The used notation to describe the mathematical foundations is similar to those of (Sánchez et al., 2013) and (Ben-Shabat et al., 2018).

7.7.1 Fisher Vectors

Let $X = \{x_1, \dots\dots, x_T\}$ be a set of T observations of a certain process, whose statistical behaviour is assumed to be statistically described by a probability density u_λ, which depends on a set of parameters λ. To be more specific, u_λ describes the generative process of the observations.

The generative process has considered the value of the partial derivative of $\log u_\lambda$ related to each parameter. Generalizing to the whole parameter set, we can compute the gradient values $G_\lambda(X)$:

$$G_{\lambda}(X) = \nabla_{\lambda} \log u_{\lambda}(X) \qquad (7)$$

Then, considering the Fisher kernel, similarity measures have been taken between two samples, i.e. the inner product between gradient vectors, weighted by the Fisher information matrix.

So, since the observations are independent, $u_{\lambda}(X)$ can be factorized in the product of the density function, and consequently:

$$G_{\lambda}(X) = \sum_{t=1}^{T} \nabla_{\lambda} \log u_{\lambda}(x_t) \qquad (8)$$

7.7.2 Association of Fisher vectors with Gaussian Mixture Models

The Gaussian Mixture Model (GMM) allows approximating arbitrarily well any continuous distribution. So a GMM appears to be well suited as generative density:

$$u_{\lambda}(x_t) = \sum_{k} \omega_{\mathrm{k}} u_{\mathrm{k}}(x_t) \qquad (9)$$

Where $u_{\lambda}(\cdot)$ is the *k*-th Gaussian, whereas ω_{λ} is its weight. The mean and the covariance matrix of the *k*-th Gaussian are u_{λ} and Σ_k, respectively.

Using the soft-max formalism (Sanchez et al., 2013), ω_{k} can be substituted with α_k:

$$\omega_k = \frac{e_k^{\alpha}}{\sum_j e_j^{\alpha}} \qquad (10)$$

Let $\gamma_t(k)$ be defined as

$$\gamma_t(k) = \frac{\omega_k \, u_k(x_t)}{\sum_j \omega_j \, u_j(x_t)} \qquad (11)$$

Then, the gradients concerning the parameters can be computed as follows

$$\nabla_{\alpha_k} \log u_\lambda (x_t) = \gamma_t(k) - \omega_k \tag{12}$$

$$\nabla_{\mu_k} \log u_\lambda (x_t) = \gamma_t(k) \frac{x_t - \mu_k}{\sigma_k^2} \tag{13}$$

$$\nabla_{\sigma_k} \log u_\lambda (x_t) = \gamma_t(k) \left[\frac{(x_t - \mu_k)^2}{\sigma_k^3} - \frac{1}{\sigma_k} \right] \tag{14}$$

Where Σ_k was assumed to be diagonal, with the values on its diagonal equal to σ_k^2.

It is worth notice that the soft assignment $\gamma_t(k)$ is usually sharply peaked, e.g. the t-th observation can be quite safely assigned to its closest Gaussian $u_\mathrm{k}(\cdot)$.

If the above observation holds, then the Fisher information matrix becomes approximately diagonal (Sanchez et al., 2013), and its effect can be summarized by a normalization of the Fisher vector as follows:

$$G_k(X) = \frac{1}{\sqrt{\omega_k}} \begin{bmatrix} \sum_{t=1}^{T} \nabla_{\alpha_k} \log u_\lambda (x_t) \\ \sum_{t=1}^{T} \nabla_{\mu_k} \log u_\lambda (x_t) \\ \sum_{t=1}^{T} \nabla_{\sigma_k} \log u_\lambda (x_t) \end{bmatrix} \tag{15}$$

Where $G_k(X)$ is the part of the Fisher vector related to the k-th Gaussian. The overall Fisher vector can be obtained concatenating all the $\{G_k(X)\}$.

7.8 Results and discussion on classification

The training was performed considering several thousands of cantilever masts, portals and other assets taken from the MUIF telemetry point clouds of the province of Venice lot. Only railway assets identified by at least 700 points were considered.

The above-explained procedure was applied on a railway test area, approximately 1 km long, in the province of Venice (Italy), corresponding to about 260 million points, not included in the training dataset. A total of 470 objects were included in such area (but only 4 portals).

The obtained classification results are reported in Table 3.

Real Class	**Classification results**		
	Portal	**Cantilever mast**	**Other**
Portal	100%	0%	0%
Cantilever mast	3%	86.2%	10.8%
Other	0%	2.5%	97.5%

Table 3 - Classification results

Fig. 59 shows two examples of classification errors, related to cantilever masts misclassified as a portal and as another object.

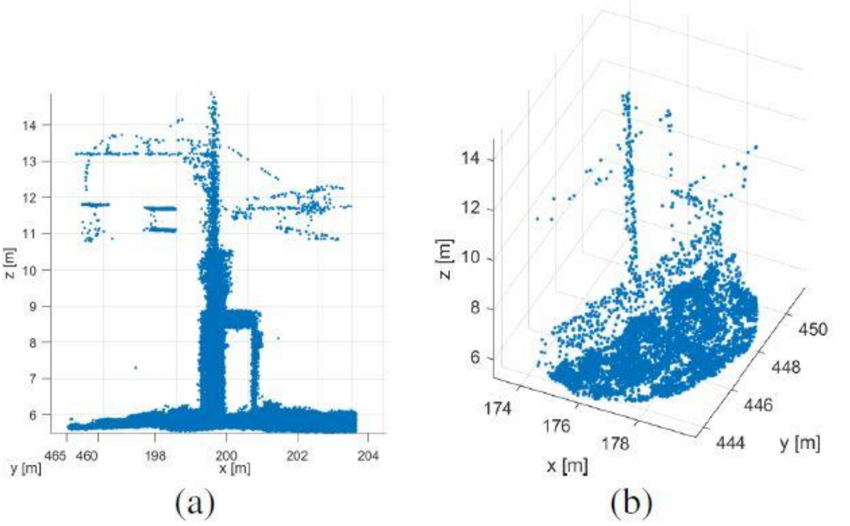


Figure 59 – "Examples of classification errors: objects classified as (a) portal, (b) "other", instead of cantilever mast." (by Corongiu et al., 2020)

Fig. 60 shows an example of false-positive classification as cantilever mast.

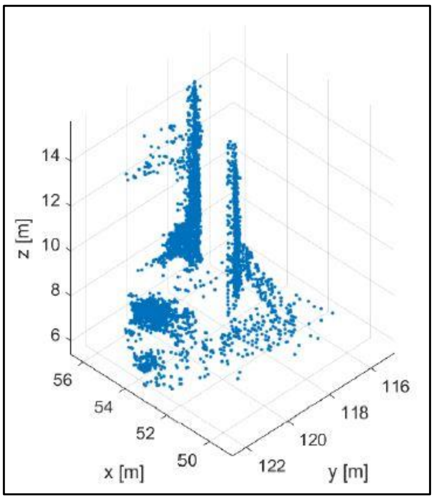


Figure 60 - Example of classification error: object classified in the cantilever mast class instead of "other".

Finally, Fig. 61 shows the distribution of the position errors of the classified cantilever masts in comparison to the related asset geospatial feature of the MUIF DB.

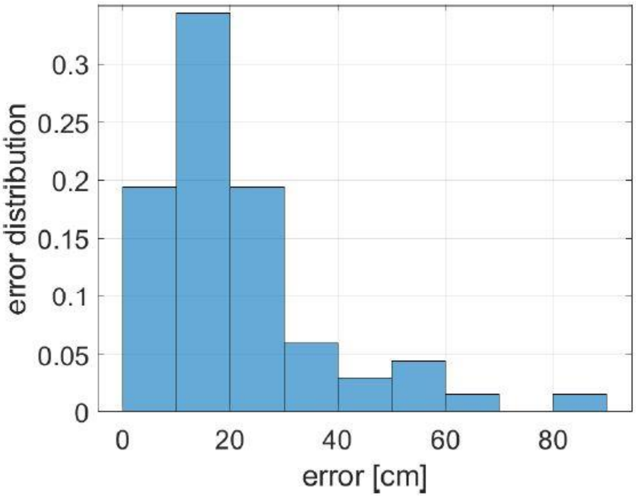


Figure 61 - Distribution of the position error of the classified cantilever masts.

The main motivation of such kind of approach is aimed to reduce the classification sensitivity concerning updates on the input for instance in relation with different point densities or obstructions.

The use the Fisher feature vector can allow obtaining a linear classifier with a similar performance to those obtained with a nonlinear one (Sanchez et al., 2013).

The candidate extraction steps proved an effective procedure to properly extract assets That despite the presence of extra candidates, could be considered relatively fast, reducing the overall computational burden, i.e. notably reducing the number of points to be examined by the neural net.

The pre-processing step allowed to obtain quite good results in terms of reducing the number of outlier points, not related to the object of interest, i.e. vegetation, walls.

The classification performance could be considered quite acceptable in particular about portals, they a quite easily identified even it is necessary, for feature work, to consider a much larger number of samples for a statistically more reliable result.

Several cantilever masts were classified as “other” objects. This was probably mostly due to the presence of certain cantilevers that are thin enough to be described by few points in the mobile mapping 3D reconstruction (Fig. 59(b)). Furthermore, some types of cantilever masts are similar to portals support part (Fig. 59(a) and fig. 60).

Despite the obtained classification errors might be acceptable taking into account only the information provided by a local subset of the overall clouds, the addition of information provided by the context should allow reducing the rate of such errors.. Furthermore, the extension to the classification of other railway assets will be considered as well, along with more depth analysis of the influence of the point density on the classification results. Finally, an extension to the analysis of buildings and structures close to the railway will also be considered (Park et al., 2007, Chen, 2012, Bitelli et al., 2004, Guarnieri et al., 2015, Masiero et al., 2015, Boreggio et al., 2018).

To conclude, the position of the detected objects was quite well estimated to the reference one, with an error usually smaller than 30 cm, as shown in Fig. 61.

7.9 Some conclusions on automatic classification

This part of the thesis presented the current state of development of an automatic approach for the extraction of information about railway assets from large point clouds collected by a mobile mapping system mounted on a train.

Such approach aims at reducing the need for human interaction needed during geospatial information extraction, potentially also speeding up the overall process. The workflow is implemented in the considered approach: determine a set of candidate asset areas, pre-process the subset of points in such areas and finally feed them as input to a 3DmFV neural net classifier.

The obtained results show acceptable performance in the classification of certain railway assets, however, the proposed procedure shall be extended to other objects in our future work.

8. CHAPTER 8 – AN INTEGRATED SPATIAL MODEL FOR HETEROGENEOUS DATA: MANAGEMENT, UPDATE AND VALIDATION

This chapter is mainly based on (Corongiu et al., 2018) and (Tucci et al., 2019) scientific papers.

8.1 The SDBMS requirements

In a given SDBMS, data may be modelled in classes as parts of an object-oriented geo-model. Defining implementations for geo-models that efficiently store and retrieve the models in spatial databases received enormous attention from researchers. To properly visualize conceptual contents into a physical geo-model stored in a spatial database a specific remapping is needed. Moreover, the heterogeneity both of geometric and topological aspects can be efficiently solved by a geospatial DBMS implementation (Li et al., 2020).

The motivation to design the SDBMS as a better solution to manage heterogeneous datasets is related to the usability in spatial analysis and management capacity over time. It means that one of the main tasks in modelling has been addressed not only on contents but also toward process management. Hence if all the suggestions to obtain interoperability have been carried out according to the steps detailed in the previous chapters, an integrated multi-LOD multi-source general-purpose 3D SDBMS has been set up, able to be continuously updated.

The idea should be to realise a unique database in a logical sense, but physically distributed and shared throughout the territory, that could permit different users to manage data in an independent and integrated model, but still maintaining the ownership of data in different contexts, according to INSPIRE Directive principles (Directive 2007/2/EC., 2007). By defining a topographical reference base, it will be possible to guarantee the sharing of topologically consistent geographical information in the first version and the updated versions. At the same time, a global information system will be set up which can grow gradually and incorporate integrated information at different levels.

Therefore, to maintain and allow dynamic evolution over time, such SDBMS has to be founded on some basic characteristics:

- 3D, multi-LOD, multi-source management and evolution;
- The validation process of different phases;

- Continuous updates of heterogeneous spatial and temporal dimensions.

8.2 Multi-source, Multi-accuracy, Multi-LOD Data

Over recent years, harmonisation of different datasets into homogeneous DBs become a common practice since the acquisition phase can be based on different systems and sensors: aerial surveys, terrestrial surveys, drones, LIDAR, etc. each of which with specific products. As a result, the evaluate the quality of every single dataset and the congruency level of all information in an integrated geospatial DB have to take into account both multi-resolution and multi-accuracy aspects. Another issue to be considered refers to the design of continuous updates that such geospatial DB need to support over time. These characteristics need to be managed according to a BIG data approach and with a robust solution since the beginning of the implementation phase.

The considerable amount of data produced by the systems mentioned above and sensors has been generated by continually growing and implementing data acquisition techniques. The collected data vary by mainly considering the 5 Vs: Volume, Velocity, Variety, Value and Veracity. Furthermore, data are characterized by different properties and be structured, semi-structured or unstructured. BIG data is a prevalent keyword in a wide range of fields such as data handling and storage and more efficient computing algorithms, although related to technological progress in sensors development. BIG data coming from different sources has also opened a debate on technological improvements in surveying data that could broader toward new geospatial application areas (Thaduri et al. 2015).

Concerning integration, the data fusion approach represents an effective process to obtain that multiple data of the same real-world object could be consistently, accurately and useful accessible (Dong et al., 2000; Bleiholder & Naumann, 2009). For example, there are several construction datasets for building domain generated by different data providers. Data fusion aims to harmonise different datasets into a database with a consistent data schema removing duplicate and integrating them through the automatic ingestion processes (Li et al., 2020). Fig. 62 presents the paradigm of general data fusion.

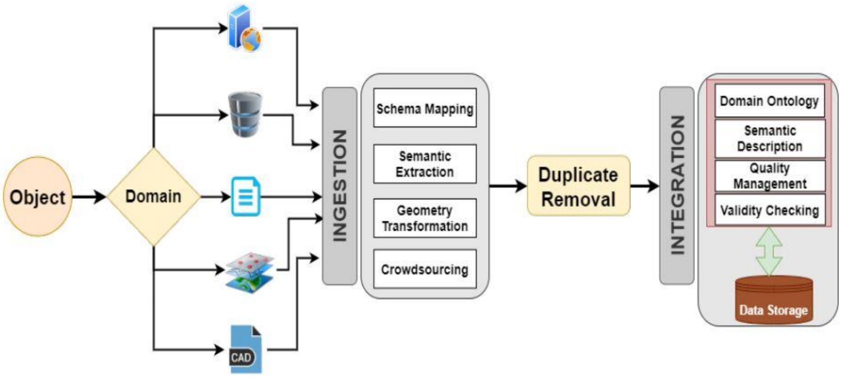


Figure 62 – The paradigm of data fusion (from Li et al., 2020)

About multi LODs, a depth analysis is described in (Biljecky, 2016b): "Techniques for producing 3D data can derive data in multiple LODs (e.g. an airborne laser scanning survey can result in both block models, and models with detailed rooftops). However, multi-LOD data of the same real-world object are still seldom available. This deficiency hinders applications that require them as input, such as visualisation. There are a few possible reasons why multi-LOD datasets are rare, among others: (1) GIS software packages are generally not programmed to utilise multi-scale representations of 3D models; (2) 3D city models are usually acquired for one purpose in mind; hence they are acquired in the optimal (single) LOD; and (3) there are limitations in the acquisition and storage process, e.g. a 3D modelling software is not capable of simultaneously producing two or more representations, and to store them consistently and efficiently" (Biljecki, 2016b).

Regarding the GTDB used as a case study of this thesis, territorial objects have been described through different spatial components that in turn have been referred to different LODs (i.e. for buildings, streets, artefacts, bridges, etc.). This characteristic came from the compliancy with both the GTDB National Specs (MD, 2012) (as detailed in Chapter 2) and the SienaGTDB (as detailed in Chapter 3). Thanks to this approach the GTDB could be defined as a MultiLOD SDBMS.

About BIG data concerns, two additional aspects must be taken into account: different phases to supply geospatial information from several data sources and the evolution of the technical cartography in a DB structure as detailed in Chapter 2.

8.2.1 Multi-dimensional spatial data supply among different phases

In the context of a cartographic map production process, a significant increase in the availability of different tools and sensors leads to the greater complexity of the data structures. Analysing the state of art, it is worth notice that the past's classical acquisition process was mainly supplied by aero-photogrammetric surveys, aimed at delivering a map at a specific time and about a specific territory. Methodologies for validating processes were, in that context, almost established and standardised. Nowadays, the growing availability of multi-source devices requires considering multi-resolution characteristics to validate such new data supply. In this case, the validation process should be carried out in compliance with the adopted technology and the integration of a complex geographical infrastructure.

The resulting cartographic product validation is carried out by a comprehensive elaboration of interesting data sources. The resulting cartographic product validation is carried out by a comprehensive elaboration of data interested sources. The overall process is almost always made up of steps that start when the previous one is concluded. The need to validate each intermediate phase depends on the process management to be carried out.

For instance in Italy, a Regional Authority cartographic service can promote a survey campaign made by aero-photogrammetric survey and related aerial triangulation in the first phase then, in the next phase, will entrust a related spatial database creation. In this case, photos and aerial triangulation could be considered as final products. Moreover, those managing the infrastructures need to survey vast territories to implement a topographic database by supplying related orthophotos and DTMs (Digital Terrain Model). Preparatory products for carrying out final compositions are only significant concerning their supply.

Another side effect of adopting an SDBMS instead of traditional cartography is the cartographic representation and symbolisation of geographical objects, not so developed in research literature because not a topic related to SDBMS. It seems straightforward to distinguish between the database and cartographic representation from a theoretical perspective, since inaccuracies because of symbolisation are avoided in the database. Otherwise, the database product is a vector representation of the map. A multi-scale topographical database requires model data models to be compliant at all scales consistently. The separation between model geospatial feature and cartographic portrayal requires at the same time that geometric represented entities and databases spatial features keep consistent. (Stoter et al, 2008).

The symbolisation (OGC SE, 2006), the styled layer description (OGC SLD, 2007) and generalisation processes to represent different scales of geographical objects are out of this thesis's scope but considered crucial for future developments of the SDBMS published in Geoportals on the WEB.

8.3 3D aspects and evolution modelling

It appears that the advances in geo DBMS are progressively striking: nowadays, one of the mainstream related to SDBMS is focused on supporting data types to handle 3D. (Verbree & Zlatanova, 2007).

The case study considers datasets integrated into a single GTDB. Some advanced processing analyses have been performed to evaluate which railway assets should be modelled in 3D starting from vectorial and point cloud source data (so integrating both aero photogrammetry and train MMS). Railway assets are objects that the Railway Enterprise needs to manage and update for its institutional purposes continuously, and the GTDB has explicitly been aimed to model them in a GIS. In traditional technical cartography, railway assets are mainly represented by points or lines geometries just to locate their position.

For example, railway portals and pylons having a small footprint, are represented as curves and points, but because of they pass over the railway track, they maintain a specific 3D shape. For this kind of spatial objects, where volumes are not obtainable from the original geometric features of the implemented DB, the acquisition of some additional geospatial components has been carried out just to tests a comprehensive and detailed 3D view by the extrusion approach (Fig. 63).

The points belonging to these assets have been isolated from the LAS file’s initial point clouds and the footprint has been digitalised in 3D polygon shapes. Then, each of these polygons has been extruded according to its correct height. Finally, these spatial components have been included in the GTDB as specific classes related to their proper assets.

So assets have been described through some additional information about their volumetric shape (heights below portals, thicknesses of pylons, heights below bridges) implemented as attributes of the related spatial components.

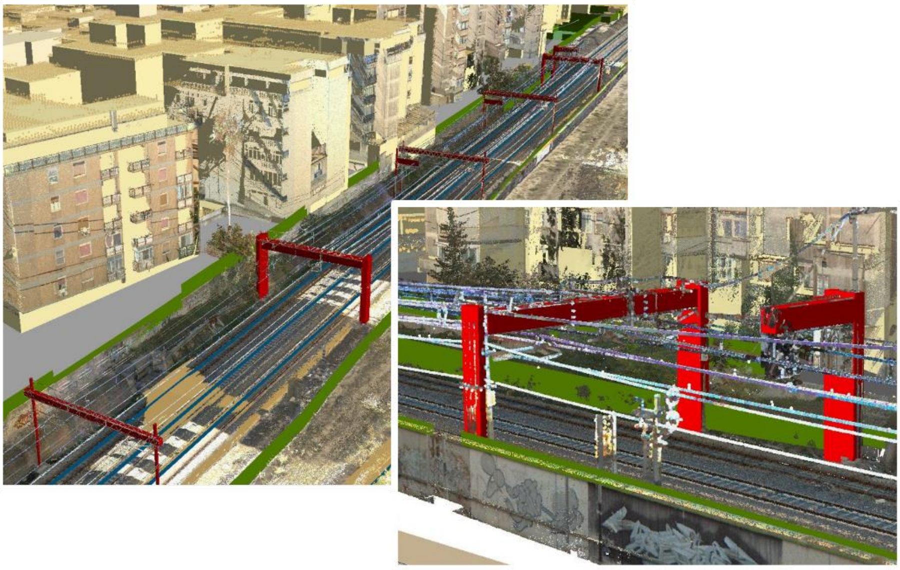


Figure 63 - 3D integration of different sources of railway assets: examples of mast portal 3D modelling (represented in red) by geospatial parameters defined as attributes into the GTDB.

8.4 The harmonisation processes

This part of the research has been aimed to harmonise different types of data in the same GIS environment. Hence, a deep analysis has been carried out using spatial operators applied to several aspects of geographic information. As an example, some investigations have been conducted about shared infrastructures between different transportation network such as rail and routes or considering which part of the hydrography network could impact the railway network.

The first step involved re-projecting all datasets toward a single Spatial Reference System (SRS). Then datasets have been analysed and overlaid together (and not only visualised by a re-projection on the fly). In this way, spatial topological constraints have been combined among data coming from different sources.

The main critical task of this step has been related to the multi-resolution management among several objects into a sole DBMS to achieve integrated contents. So, information coming from many data sources have been integrated into each class of objects into the database.

For instance, the railway network supplied by the airborne photogrammetric source has been integrated with the point clouds by train laser in the underground path (e.g. in a gallery) to obtain a composite track-centreline with an attribute that specifies their position relating with the DTM (overground, on the ground, underground, etc.).

Even almost ordinary for topographic objects, this harmonisation process has been more complicated if related to railway assets. Assets, before integration, were implemented in external classes not integrated with the territorial context and not yet semantically described in a topographic 3D model. Hence, assets have been implemented as 3D spatial components so that a B-Rep could be then derived.

This homogenising process has been only an intermediate step to obtain fully B-Rep modelling since the research is already ongoing. The idea has been focused on finding the best way to acquire spatial components of object classes that generally belong to a 3D city model. Such geospatial components must be compliant with the railway construction information needed to manage and maintain the existing asset DB. From a semantic point of view, the main critical encountered task was to define the geometrical attribute of railway assets following the same approach of cartography where reference surfaces are identified to extrude them in order to obtain a B-Rep visualisation. In fact, this happens because not always railway assets have a significant footprint. The connection between complete B-Rep modelling of data and a BIM structure for management will become the objective for profiling data.

the unique IDs of assets have guaranteed correspondence between spatial feature and the point clouds of the telemetry survey and the connection procedure explained in Chapter 7. The GTDB consists of a number of classes that define geospatial objects as a content of a 3D geotopographic cartography. Also, properties defining the type of extrusion and the corresponding elevation value have been implemented to prepare the 3D representation necessary to show their real volumetry (Fig. 64).

Figure 64 - 3D views of building extrusions

Moreover, even if all the above datasets have been memorised with the same planimetric Spatial Reference System (ETRS2000 RDN2008, EPSG:7794[21]), the Vertical System referred to altitude to ellipsoid instead of orthometric elevation, a geodetic re-projection was carried out.

For instance, that re-projection has been necessary to make orthophotos homogeneous with all other datasets, according to their high-level resolution (0.10 m pixel size). For this reason, the use of the DSM by TIN only for ground areas has been carried out. For the Triangulation

---

[21] https://epsg.io/

phase, only LIDAR point clouds with a “ground” classification in the LAS file have been selected. Next, the orthophotos were moved to the correct 3D elevation level using the TIN as a reference surface. Afterwards, LIDAR datasets and the GTDB buildings were spatially overlaid to classify roof and balcony point clouds (Fig. 65).

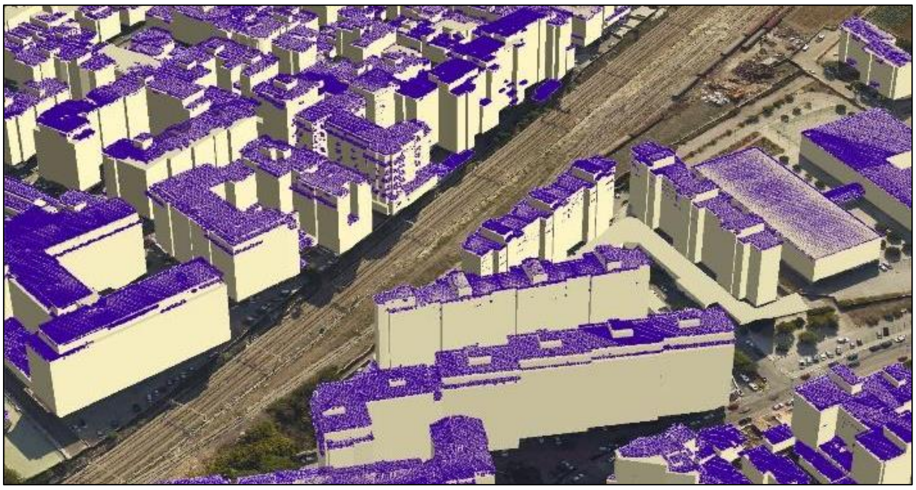


Figure 65 - Roof and balcony point clouds coloured in purple.

Facades were not easily acquirable from the aerial photogrammetric survey, thus the architectural parts such as balconies, terraces, bowindows (have been integrated using point clouds coming from the train telemetry survey (case study data source explained in Chapter 5) (Fig. 66).

Figure 66 - Point clouds of façades from telemetry survey of the case study (chapter 5).

Concerning their density, it has been essential to visualize point clouds with as many points as possible so that holes could be avoided, and full detail has been obtained. Thus, the original point clouds were partitioned to manage them better and using the maximum resolution. Some architectonical elements of façades have been integrated as a tested solution only for buildings along the railway track, attained from telemetry point clouds.

8.5 The validation processes.

This part of the thesis focuses on critical aspects encountered during the validation phases of the spatial data in the MUIF project. In particular, new techniques to validate a multi-source DB have been taken into account together with traditional ones related to cartography quality certification. The first ones haven’t yet reached a consolidated application level due to the rapidly advancing technological evolution, but profiling of quality standards has been followed. For this reason, a comprehensive introduction on multi-sources DB state of the art and related technologies have been detailed. Therefore, considering the use case, the validation process has been carried out by a step-by-step approach. A consolidated validation methodology has been adopted for traditional products (by airborne survey) such as photograms orthophotos and lidar, while a comparison with ISO standard quality specification (ISO TC/211 19157, 2013) has been followed for the main innovative survey (by MMS - Mobile Mapping Systems mounted on the train). Finally, for the GTDB both massive (informatic procedures) and traditional thematic evaluation accuracy have been combined. Therefore, the adherence with standards has been referred to consider both the quality of data and the conformity to data product specifications.

Moreover, to guarantee compliance among different phases, the evaluation of the quality has been carried out at the end of each step. A rethinking of the evaluation process in the context of spatial databases has been thus necessary. About the GTDB, different tools for validating it has been performed, not just considering a significant sample but the DB as a whole. Procedures should be aimed at validating the data content's compliance and structure according to the logical model defined into technical specifications, including the geometrical and topological constraints (Carrion et al. 2008). The main difference between a traditional product of cartography validation and that one applied to a more complex SDI (Spatial Data Infrastructures), is about the needed validation of dependency, relationships and constraints of the latter. In this context, the basic effectiveness of a data product specification is related to make understandable the type of procedure to be performed to a dataset to evaluate the quality both at single and overall product.

A data product specification aims to a detailed description of a dataset or dataset series, that in addition to information on how to supply data (format file, extension, lots, etc.) will allow to replicate and re-create the same dataset, with the same quality aspects, following the same rules and process. It is a detailed technical description of the data product regarding the requirements that it will or may fulfil. The purpose of this International Standard is to provide requirements on the content of data product specifications, in compliance with other existing standards for geographic information (ISO/TC 211 19131, 2018).

8.5.1 Data Product Specification and Quality Standards

The description of geographic data quality facilitates the comparison and the selection of tailored data set for specific application needs and requirements. Complete descriptions of a data set's quality will encourage the sharing, interchange and use of appropriate data sets. Information on the quality of geographic data could be found in the specific standard (ISO/TC 211 19157, 2013). It allows a data producer to evaluate the compliance of a dataset to the criteria outlined in its product specification. The ISO19157 standard also supports users in evaluating particular application requirements. For this evaluation, clearly defined procedures are used consistently Moreover, according to this standard, different phases have been taken into account:

1. Deliverables
2. Validation
3. Accessibility
4. Updating

Particularly about the “Deliverables”, in the case study some critical aspects, have been concerned in establishing which the different steps and which lot subdivision were necessary to distribute to each phase expert their specific part. Then the different parts have been led up to the next ones considering passages from deliverables and validators and vice-versa, harmonisation of different multi-source datasets into a single database validation.

Another reference standard during the case study validation process has been the Data Product Specifications (ISO/TC 211 19131, 2018). This standard impacts on structures chosen to evaluate the quality both for each lot and at a general level certification of the results.

Regarding quality evaluation, for the different phases, a specific standard (ISO/TC 211 19157, 2013) has been considered (Fig. 67).

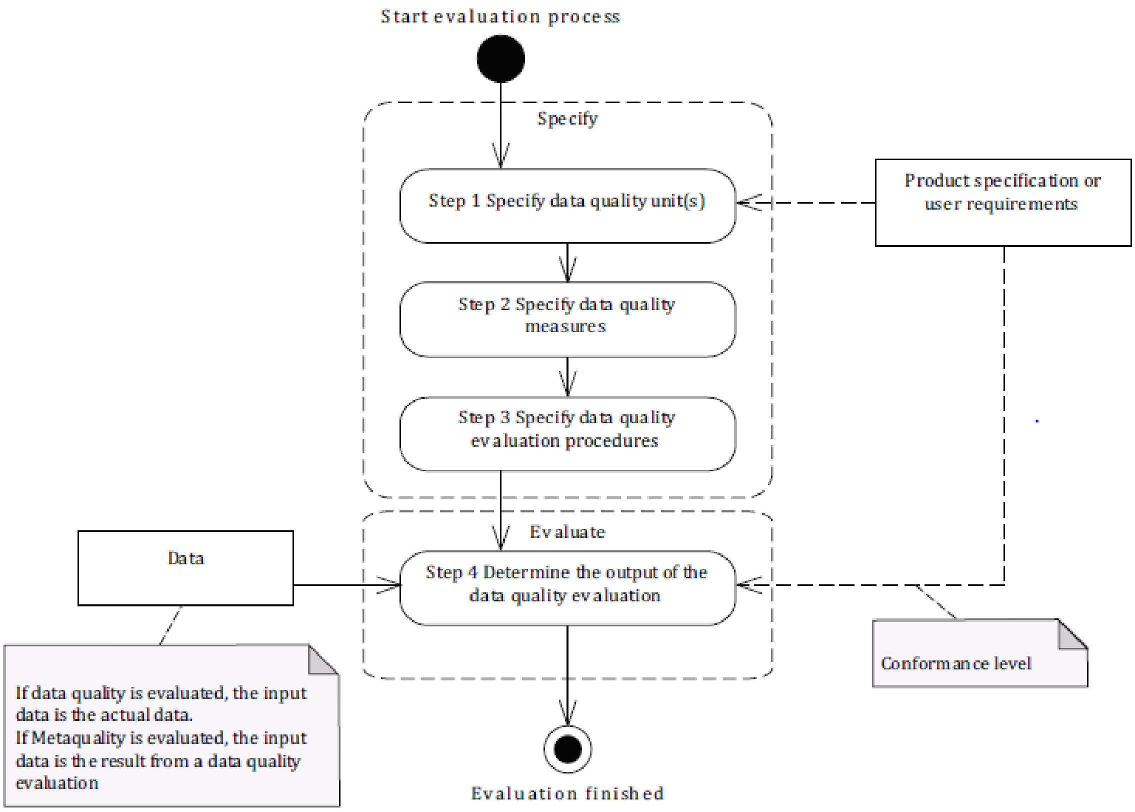


Figure 67 - Relationship between data product specifications and quality evaluation steps (from) expressed in terms of activity diagram: continuous text boxes represent actions, dashed regions represent the activity background, dashed arrows represent dependencies, continuous arrows show the sequence of activities

In the *Quality* standard (ISO/TC 211 19157, 2013), the basic elements and their quality characteristics are described through *data quality units*. The data quality unit combines *scope* and *data quality elements*. D*ata quality elements* are components describing certain aspects of geographic data quality, and these have been organised into different categories (Fig. 68).

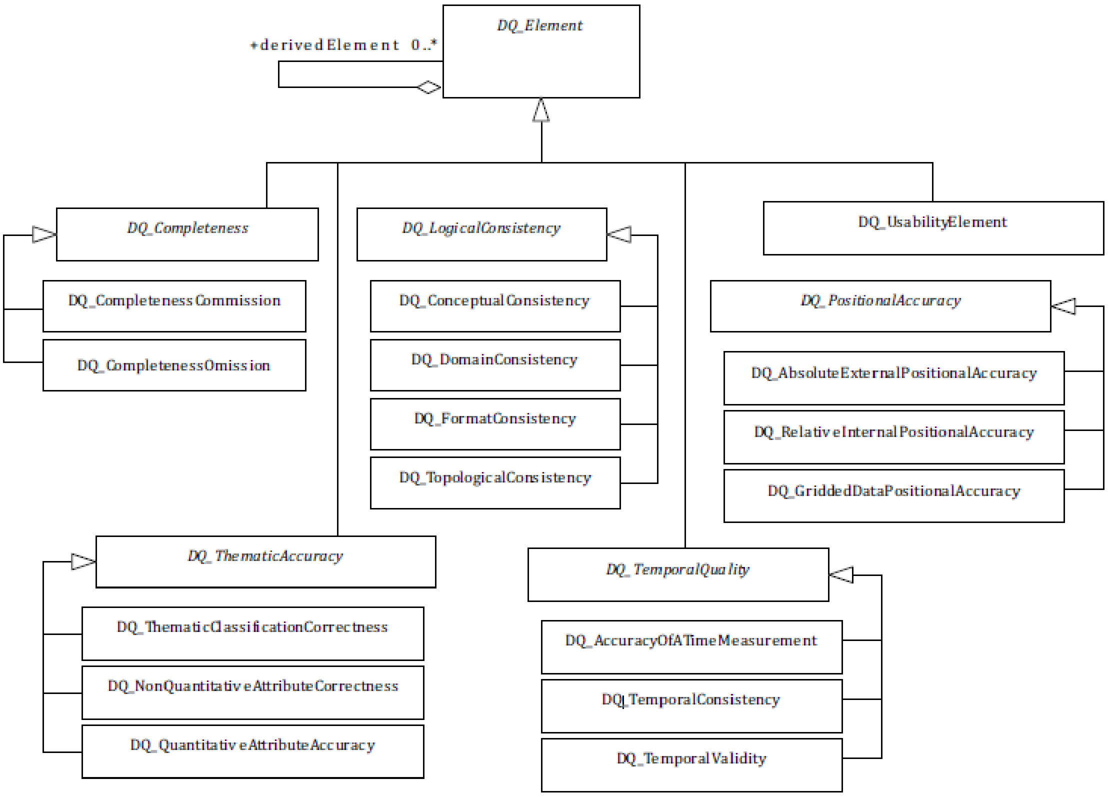


Figure 68 - Data Quality Elements as described in the ISO 19157 standard (from ISO/TC 211 19157, 2013)

The categories that have been considered are:

- Completeness (C1): This category defines the presence or absence of features, attributes and relationships. It consists of two data quality elements:
  - commission (C1_1): excess data present in a data set
  - omission (C1_2): data absent from a data set
- Logical consistency (C2): defined as the degree of adherence to logical rules of data structure, attribution, and relationships (data structure can be conceptual, logical or physical). Just to give an example: where logical rules are documented in a data product specification, then the source should be referenced in the data quality evaluation. It consists of four data quality elements:
  - conceptual consistency (C2_1): adherence to rules of the conceptual schema
  - domain consistency (C2_3): adherence of values to the value domains
  - format consistency (C2_4): the degree to which data are stored under the physical structure of the dataset

- — topological consistency (C2_5): correctness of the explicitly encoded topological characteristics of a data set

- Positional accuracy (C3): This category defines the accuracy of the position of features according to a spatial reference system. It consists of three data quality elements:
  - — absolute or external accuracy (C3_1): closeness of reported coordinate values to values accepted as or being true
  - — relative or internal accuracy (C3_2): closeness of the relative positions of features in a data set to their respective relative positions accepted as or being true
  - — gridded data positional accuracy (C3_4): closeness of gridded data spatial position values to values accepted as or being true
- Thematic accuracy (C4): is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and the classifications of features and their relationships. It consists of three data quality elements:
  - — classification correctness (C4_1): it compares classes of features or their attributes to a universe of discourse (e.g. ground truth or reference data)
  - — non-quantitative attribute correctness (C4_2): it measures if a non-quantitative attribute is correct or incorrect
  - — quantitative attribute accuracy (C4_3): it verifies if the value of a quantitative attribute is closeness to a value accepted as or known to be true
- Temporal quality (C5): This category defines the quality of the temporal attributes or the quality of feature temporal relationships. It consists of three data quality elements:
  - — accuracy of time measurement (C5_1): it defines if reported time measurements is closeness to values accepted as or known to be true
  - — temporal consistency (C5_2): it defines if the order of events is corrected
  - — temporal validity (C5_3): it defines if data concerning time is valid.

Finally, the component called *Data Quality Measure* has been considered.

- Data quality measures (D1): a data quality element should refer to one measure only, using a measure reference, providing an identifier of a measure that is fully described providing the name and a short description of the measure.

8.5.2 The validation methodologies

In the railway case study, datasets are characterised by a significant level of details both about the spatial and the temporal dimensions, combining railway assets along with the railway network with all the topographic objects in their surroundings. About spatial dimension, it is worth noticing that generally mobility infrastructures were developed along a predominant longitudinal direction and cross vast territories. Moreover, railway objects to be managed need to be detailed in each part and with the defined constraints with all the geospatial objects that cover the territory. Due to the longitudinal development across different reference topographic data, the validation process must consider the multi-dimensional aspects of the case study DB.

The relative European policies state that railway infrastructure managers should focus on reducing operational costs while simultaneously increasing financial assets and safety performances. In the railway context, the CMMS (Computerized Maintenance Management System) is one of the system's principal component. Such system implements the LAM (Linear Asset Management), uses dynamic segmentation allowing to project assets as *events* along-track centreline and characterising them as attributes of the rail network. The CMMS Linear referencing also enables the description of structural linear assets of the network such as intersections or switches (Thaduri et al. 2015).

In fact, railway assets have been modelled considering their semantic meaning into the data model, as it happens for topographic ones, offering good data integration possibilities, adaptability, and compatibility compared to traditional approaches. without forgetting that they are created for a specific purpose, it is necessary to model assets in the most abstract way possible. The primary usability requirements for a shared semantic data model are strictly related to the level of abstraction in defining railway assets. Indeed, each investor builds bespoke railway lines according to different standards, technologies, thus obtaining independent routes.

In the past years, the lack of interoperability has incentivised several has uncoordinated scenarios where different companies often have operated on the same destinations with competitive routes. Across borders, each European rail operator has freely invested in run services avoiding stop to swapping locomotives, or passenger changes. Another issue affecting trains that serve multiple countries is the need for proper control systems to interface with each country's signalling infrastructure, thus increasing complexity and reducing reliability (Tutcher, 2014). The context's impact of this in terms of interoperability between different data models is the central critical aspect to focus on finding a standard solution also in the management of the railway spatial data.

To solve the interoperability issue and taking into account the railway network's extension, the datasets have been organised in different deliverables (according to data sources) and several lots. Thus, the validation processes have also considered the congruence between lots and contents of different sources. For the first congruency issue, adjacent lots of the same type of dataset have been considered. For the second one congruence, different kinds of datasets (orthophotos vs LIDAR, etc.) have been evaluated.

As a result, the overall certification has considered all the dependencies of the geospatial datasets: congruencies in a multi-source/multi-dimensional DB have been implemented trough to validate both time and spatial dimensions. Then ad hoc procedures on rules and constraints have also been applied to every single type of dataset. This approach has required a definition of a shared dataset model, defined in a data product specification.

8.5.3 The validation of aerial photogrammetry surveys

Compared to the past, nowadays the entire process could be based on faster and less critical solutions , and the quality certification could be well-structured with massive actions. Therefore the availability of several sensors often with different characteristics implies that quality certification must pay attention to choosing performant indicators.

If once upon a time a scale of a photo and their photographic grip distance implies the cartography scale, nowadays to evaluate if the required level of detail is compliant with a chosen sensors, it is necessary to consider precisely its characteristics and related optical tools.

The availability of GNSS (Global Navigation Satellite System) receivers and IMU (Inertial Measurement Unit platform) onboard sensors () simplifies the role of aerial triangulation, to a homogenising and quality process. The matching between photograms became completely autonomous, robust and reliable algorithms can guarantee the correctness of coverage. Then, each block could be considered extremely rigid, and together with the onboard receivers, these aspects dramatically reduce the number of support points on the ground. As a result, the number of checkpoints necessary for quality evaluation was also considerably reduced. The numerical images are then easily processable, avoiding the previous parameter optimisation and only making some adjustments to transform them.

To summarising, only two basic indicators have been considered to certify the suitability of a photogrammetric flight:

1. The validation of the GSD (Ground Sample Distance) has been calculated according to information of the camera parameters, the onboard sensors, and a general unrefined terrain model;
2. A measurement of limited conveniently arranged points to validate the aerial triangulation phase.

8.5.4 The validation of (DTM/DSM datasets) by airborne LIDAR survey

Regarding mobility infrastructures, the knowledge of orography represents one of the most important framework data for understanding connections and critical aspects, to facilitate spatial analysis and risk evaluation managing the related Spatial Data Infrastructure. Datasets from aircraft LIDAR sensor have been supplied in raw strip delivers, and organised in sheets of territory. A tiling process is often an implicit tool within the software where it is difficult to store information about the algorithms used. The unavailability of such process lineage metadata makes the quality certification a very problematic aspect. Because of supplied dataset are organised in final (row strips) and intermediate (point clouds tiled into sheets) products a) the evaluation has been carried out certifying the final supply and the processes followed to obtain this one.

For instance, the interpolation process is delivered in grid files. Interpolation could be considered effective if the source data is dense enough, to make sure that no under-sample areas exist only raw strips and point clouds can be evaluated. The altitude correctness, considering the checkpoints and all the products, also verifies that no anomalies exist in the source data, but if they exist in the grid files, it means that an error has occurred during the interpolation process.

8.5.5 The validation of terrestrial survey

As mentioned in Chapter 5, all the case study datasets have been supplied in delivery lots, as homogeneous elementary informative units, for ensuring good BIG data management thus allowing a sustainable quality evaluation of the deliverables. In fact, point clouds take up about 3-4 TeraBytes (TB) of storage for each delivered lot. Since at the end of the MUIF project there will be about 90 lots, a comprehensive archive size of about 200 TB will be provided. According to the *Quality* standard (ISO/TC 211 19157, 2013), the evaluation phases have been carried out following a rigorous approach and subsequent step definition.

According to the above standard, the *Data Quality Scope* describes the overall elements and characteristics that identify data on which the quality controls are conducted (see Fig. 69):

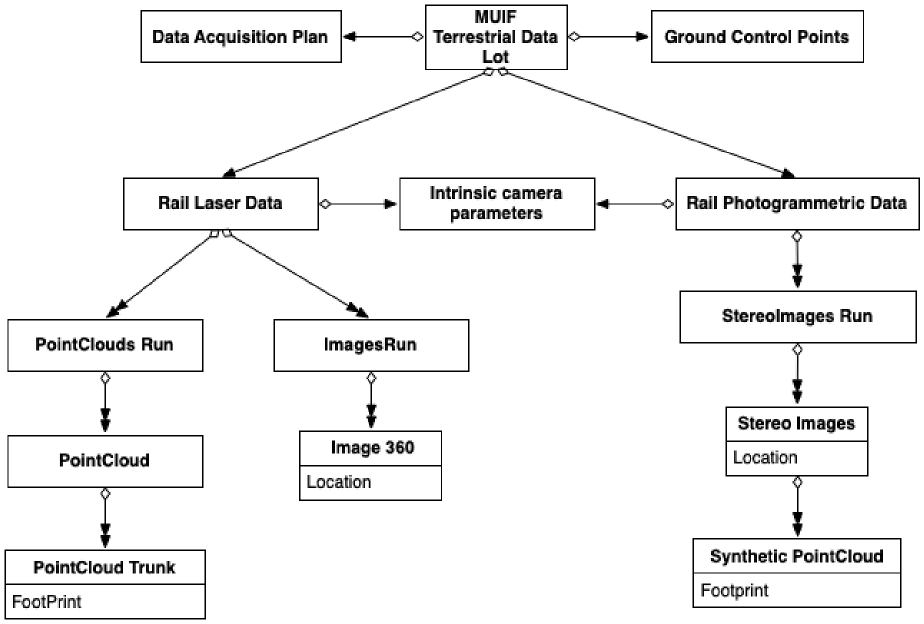


Figure 69 - Basic elements of the ISO/TC211 19157 standard considered for the case study data quality scope definition

Referring to the above figure, the case study basic elements have been:

- Data Acquisition Plan – this element defines data acquisition targets related to geospatial features;
- Ground Control Points (GCP) – this element includes the measured ground reference points that are used as a comparison parameter for point cloud precision assessment;
- Rail Laser Data – this element includes data carried out by the laser scanner mounted on the train. For each lot, only one kind of dataset exists. Rail Laser Data consists of one or more Point Cloud Runs and one or more Images Run;
- Point Cloud Run – this element consisting of a sequence of laser shots, temporally ordered. Each shot gives rise to one or more point clouds (Point Cloud). The point clouds have in turn been split into sub-clouds (Point Cloud Trunk), along about 1 km of the track centreline, to lightening the size of the related files. Between adjacent trunks, an overlap area is maintained to allow their matching;
- Images Run – this element consists of a sequence of panoramic images (Images 360). Such images are simultaneously acquired from the laser scanner data source. They have also been used to associate their RGB value to the point cloud data. As the Point Cloud Run element, the sequence is temporally ordered but grouped differently. Moreover, each panoramic image is geolocated as a point geometry in a specific geospatial feature;

- Rail Photogrammetric Data – this element refers to stereoscopic data and associated ballast synthetic point clouds. They consist of one or more StereoImages Runs;
- StereoImages Run – this element consists of a sequence of temporally ordered stereo photograms (SteroImages). Thanks to SfM (Structure for Motion) techniques a synthetic cloud has been carried out(Synthetic PointCloud). Each stereo image is geolocated as a point geometry in a specific geospatial feature.

In line with the ISO quality standard (ISO/TC 211 19157, 2013), the quality components used were:

- C1 - to verify whether all the files previewed in the data product specification were supplied with their appropriate structure;
- C2 - to verify congruence between Point Cloud Run indices and panoramic images, and between stereoscopic images and synthetic Point clouds;
- C2_4 - to verify both panoramic and stereoscopic images about format and luminance;
- C2_5 - to verify if panoramic images are contiguousness particularly verifying if the acquisition frequency is compliant for every run and if the point clouds (both laser and synthetic) have been correctly coloured;
- C2_5 - to verify if panoramic and stereoscopic images are compliant with the acquisition plan;
- C2_5 - to verify if the laser point cloud footprint is compliant with the acquisition plan coverage;
- C2_5 - to verify if Point Cloud Trunks are contiguously overlapped points according to different Runs;
- C3_1 - to verify the preciseness of the laser point clouds with the Ground Control Points;
- C3_2 - to verify the stereoscopic point clouds' preciseness according to the neighbouring Ground Control Point.

Another quality considered component has been the *Data Quality Measure.* It implied some critical evaluation aspects because it can be associated with one or more evaluation Data Quality Measures (DQMs) related to each Data Quality Element (DQE). As an example being implemented, about the case study, two DQMs have been detailed (C2_4 and C2_5 described in section 8.5.1).

Regarding C2_4, an objective method for highlighting anomalies has been focused on the luminance aspect. The percentage of images where exposure entails alignment difficulties and colouration of clouds was relatively high. Since each lot includes about 250,000 images, establishing a-priori quality confidence in order to avoid processing all the datasets has been a crucial aspect. However, the adopted procedure aims to calculate the grey value histogram and considers their median. Images, where the value is k times outside the standard deviation, were selected as outliers (being over or under exposed).

The definition of the ***k*** amplitude and the ***p*** percentage of such anomalous images has been one of the main debatable aspects in designing certification procedure since this threshold can cause the rejection of the dataset. In the data product specifications, this value wasn't defined, so a specific extemporaneous procedure established ***k*** = 2.5 and ***p*** = 0.03. Moreover, to reduce time processing, a procedure to select a significant sample instead of the whole dataset has been set up. According to the Quality standard (ISO/FDIS 19157, 2013), about the *SampleBasedInspection* evaluation, a statistical procedure was carried out to reduce time by selecting a sample but still maintaining a reliable evaluation.

For this case of study, the percentage of error in a reduced sample with an amplitude ***N***, is assumed to estimate the unknown real error. The confidential range was calculated with Significance Level (***alfa***) = 0.01. In practice, a positive validation was attributed if the estimation value was lower than ***p*** and the upper limit in the confidential range was lower than 0.06, which is double real acceptable error ***p.*** As mentioned above, the acceptable error was established at a value of 0.03. Moreover, to was established a sample amplitude of ***N*** = 300. This value is intended as a trade-off between the time needed to processing data and the acceptable amplitude range.

For what concern C2_5, the main issue was related to the data's size to be validated. It is worth noting that each Point Cloud Trunk can hold over 20 GB and trunk numbers can exceed 200 units. Then the contiguity constraint between different point cloud files has been automatically validated. In the case study, the boundary trunks have been calculated for each run. Moreover, about FootPrint overlaps, a random internal point was chosen for comparing its existence in the few point cloud dataset in which it was contained. This evaluation was carried out using a cylindrical buffer with a radius **R** and height **h** where the sub-point clouds were compared with the Iterative Closest Point (ICP) algorithm. Validation is acceptable if the RMS is lower than the threshold **e**. The parameter values selected were **R** = 0.5 m, **h** = 2.0 m, and **e** = 0.10 m. As an option, an evaluation between densities should be carried out, imposing thresholds in their relationship.

To summarising, the evaluation of the MMS train survey could be expressed in terms of the identification of DQE and the related DQM modelling. The choice to adopt statistical theories models drastically impacted in designing the validation process for those actions not yet explained in data product specifications. Moreover, the implementation of processing requires attention to the BIG data characteristics to obtain results in useful times.

8.5.6 The validation of the database

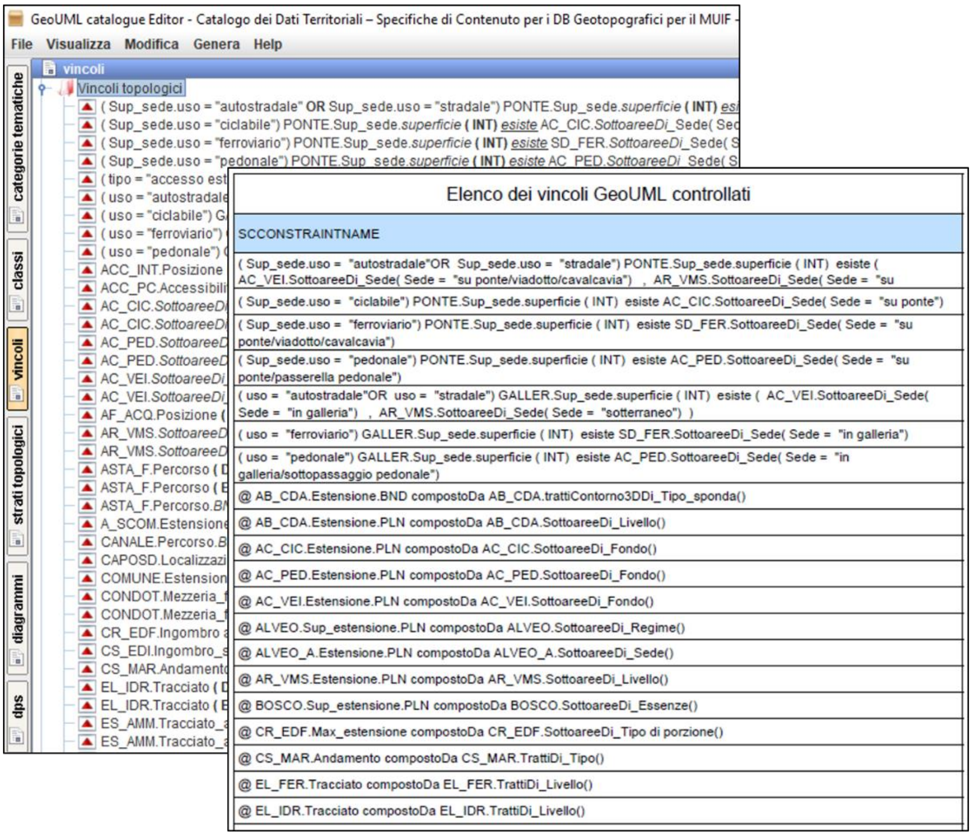
About the DB validation, specific aspects have been considered together with all the process. The first one refers to the data catalogue formalised to acquire from different sources, at different times, by collecting existing and monitoring data. Secondly, the harmonisation requirements between reference basic objects (GTDB) and Railways assets have been taken into account.

According to the international standards (ISO/TC 211 19131, 2018; ISO/TC 211 19157, 2013), the formal validation and main quality certification are drawn up by following the Data Product Specifications.

These specifications define those features, attributes, and relationships considered essential in the data set. As detailed in Chapter 5, the case study GTDB have been supplied following the GeoUML Methodology (http://geo.spatialdbgroup.polimi.it/en/), and as described in Chapter 2. For the validation process, the GeoUML Tools have been used:

- GeoUML Catalogue for the definition of the Conceptual Schema of the adopted specification and the definition of some parameters used in generating physical schemas according to the chosen implementation models.
- GeoUML Validator to perform the conformity check of the case study datasets according to the adopted Specification (Conceptual Schema) produced by the GeoUMLCatalogue.

Some basic relationships and topological constraints embedded into the GeoUML catalogue have been recalled from the national specifications and applied to the case study. Such phase has been carried out by the Company that supplying data added also delivered a self-certificated report using the GeoUML Validator tool (Fig. 70).



Elenco dei vincoli GeoUML controllati

SCCONSTRAINTNAME
(Sup_sede.uso = "autostradale"OR Sup_sede.uso = "stradale") PONTE.Sup_sede.superficie (INT) esiste (AC_VEI.SottoareeDi_Sede(Sede = "su ponte/viadotto/cavalcavia") , AR_VMS.SottoareeDi_Sede(Sede = "su
(Sup_sede.uso = "ciclabile") PONTE.Sup_sede.superficie (INT) esiste AC_CIC.SottoareeDi_Sede(Sede = "su ponte")
(Sup_sede.uso = "ferroviario") PONTE.Sup_sede.superficie (INT) esiste SD_FER.SottoareeDi_Sede(Sede = "su ponte/viadotto/cavalcavia")
(Sup_sede.uso = "pedonale") PONTE.Sup_sede.superficie (INT) esiste AC_PED.SottoareeDi_Sede(Sede = "su ponte/passerella pedonale")
(uso = "autostradale"OR uso = "stradale") GALLER.Sup_sede.superficie (INT) esiste (AC_VEI.SottoareeDi_Sede(Sede = "in galleria") , AR_VMS.SottoareeDi_Sede(Sede = "sotterraneo"))
(uso = "ferroviario") GALLER.Sup_sede.superficie (INT) esiste SD_FER.SottoareeDi_Sede(Sede = "in galleria")
(uso = "pedonale") GALLER.Sup_sede.superficie (INT) esiste AC_PED.SottoareeDi_Sede(Sede = "in galleria/sottopassaggio pedonale")
@ AB_CDA.Estensione.BND compostoDa AB_CDA.trattiContorno3DDi_Tipo_sponda()
@ AB_CDA.Estensione.PLN compostoDa AB_CDA.SottoareeDi_Livello()
@ AC_CIC.Estensione.PLN compostoDa AC_CIC.SottoareeDi_Fondo()
@ AC_PED.Estensione.PLN compostoDa AC_PED.SottoareeDi_Fondo()
@ AC_VEI.Estensione.PLN compostoDa AC_VEI.SottoareeDi_Fondo()
@ ALVEO.Sup_estensione.PLN compostoDa ALVEO.SottoareeDi_Regime()
@ ALVEO_A.Estensione.PLN compostoDa ALVEO_A.SottoareeDi_Sede()
@ AR_VMS.Estensione.PLN compostoDa AR_VMS.SottoareeDi_Livello()
@ BOSCO.Sup_estensione.PLN compostoDa BOSCO.SottoareeDi_Essenze()
@ CR_EDF.Max_estensione compostoDa CR_EDF.SottoareeDi_Tipo di porzione()
@ CS_MAR.Andamento compostoDa CS_MAR.TrattiDi_Tipo()
@ EL_FER.Tracciato compostoDa EL_FER.TrattiDi_Livello()
@ EL_IDR.Tracciato compostoDa EL_IDR.TrattiDi_Livello()

Figure 70 - Self-certified topological constraints of the GTDB

The Annex E of quality standard (ISO/TC 19157, 2013) has been considered for what concerns the choice of sampling methods in the evaluation phase. In particular, regarding the quality evaluation process, the following steps have been considered and implemented in ad hoc procedures:

- about data quality measures: logical consistency, completeness
- about data quality evaluation procedures: specify the type of the implemented procedures
- about the data quality evaluation outputs: description and error type identification, definition of the logical consistency evaluation type, completeness and thematic accuracy (both quantitative and qualitative) description

According to Data Product Specifications and specific designing reports, the specific validation processes and related informatic procedures have been implemented to evaluate the following aspects:

1. Massive validation procedures:

   a. geometrical aspects: according to the level of details of each zone/object, minimal dimensions of areas and lines evaluation according to be compared with those previewed in the Data Product Specifications

   b. Validation of Topology of an object that belongs to different zones (Figure 5) with 3D congruency connection along boundaries

2. Check considering Sampling evaluation methods:

   a. Thematic and completeness accuracy: by considering that classification of errors has omission/commission and completeness, or thematic accuracy is subjective. For example, the misclassification of a residential building instead of an industrial building could alternately be considered as an omission error of the one and commission of the other;

   b. Measurement accuracy: evaluate the different positions of an object according to different accuracies, i.e. 0.40 m for 1:1000 scale and 0.60 m for 1:2000 scale.

Procedures about point 1) have been set up by the development of specific Phyton language[22] scripts (GUI IDLE - Python's Integrated DeveLopment Environment v.2.7.15). Procedures about point 2) have been carried out manually, thus re-acquiring objects of a sample chosen according to the same statical method define for terrestrial surveys (cfr. Paragraph 8.5.5).

Acceptance Quality Limit has been defined based on the experience in evaluating the quality of cartography products and by considering the partially existing standard for inspection by samplings, such as Annex F of (ISO/TC 211 19157, 2013) and (ISO 3534-2, 2006).

Moreover, as concerns the validation of 3D geospatial aspects, the lack of implemented 3D topological tools, both in opensource and commercial software, a specific validation of the 3rd coordinate has been performed and implemented, with attention paid to the consistencies among different objects by visual or, in some cases, manual implementation of the procedures. This aspect significantly impacted the resource planning and the automation validation processing on implementing the geospatial infrastructures.

As Example, one of the main critical topological constraints to be validated into the GTDB referred to evaluating congruencies between different topological models, such as spatial topology versus network topology. As announced in Chapter 6 the first is based on the spatial topological operators (Egenhofer et al. 1994, Clementini et al. 1996), while the second one is based on Linear Reference System and network topology (INSPIRE D2.10.1 2013, ISO/IS 19148 2012). In the context of railways, the railway track centreline must belong to the railway ballast (considering exceptions, i.e. rail crossings). the validation has been performed checking if the 3rd coordinate of the graph was conformant to 3rd coordinates of the boundary of the ballast. Specifically, an ad hoc procedure has been implemented in Python language (GUI IDLE - Python's Integrated DeveLopment Environment v.2.7.15). and tested to validate the case

---

[22] https://www.python.org/

study DB Hence, the 3rd dimension track centre line must have congruent values with the 3rd dimensions of the railway ballast (Fig. 71).

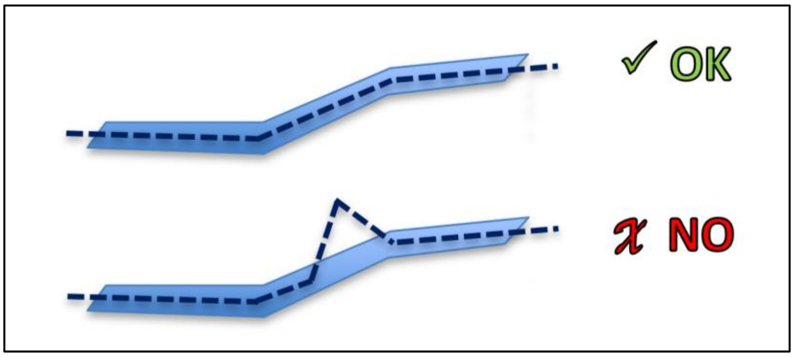


Figure 71 – Example of a 3D topological constraint: the vertices of railway track centreline must be on the ballast surface as well as the vertices of road network must be on the street area

8.6 Some critical aspects of the validation process

Based on the validation phase results, some aspects could be discussed and taken into account for future developments.

8.6.1 Amount of data, BIG data validation

One of the main critical aspects that significantly impacts the validation process is the amount of data to be processed with the same methodology and with the capacity to pursue useful results in a sustainable time. In the BIG data digital age, the availability of tools within product processes suggests some new validation activities.

According to the awareness that the lower limit of BIG data is continually moving upwards, some aspects have been analysed. In general, BIG data includes unstructured information coming from different sources, frequently near real-time supplied, making it possible to evaluate phenomena' statistical behaviour on time. The massive amount of data is not easily managed by traditional processes and requires specific algorithms and methodologies to give the results reasonably.

For a single supply of geospatial data, if the processing is not required in real-time, the association of BIG data appears inappropriate. However, because the case study is based on a multisensory survey for an extended territory, the issues can be considered similar.

The extension of the transportation infrastructure survey could be sizeable. Meantime, the acquisition plan needs to be punctual, rigorous, identifying as short as possible times for each topographic survey and to plan the update campaign. The data production is frequent, and even though each lot cannot be classified as BIG data, it becomes so given the supply frequency. Nevertheless, a single lot could include a comprehensive amount of data.

About the quality certification, these premises suggested a strategy to balance large amounts of data with the need to obtain results in a short time. Therefore, statistical methods have been considered for identifying the smallest sample's consistency to provide a reliable evaluation of the whole dataset (perhaps with a confidence greater than 95%).

8.6.2 Different semantic/thematic definition of object validation

Existing railway network models represent utilities high detailed, however, sometimes their components can not directly interact with urban geospatial features. Since the mentioned networks are very detailed and filled with semantics, several applications can be provided such as those dedicated for daily use in utility companies or to exchange models, or just representing utility networks in urban spaces. Each network represents an abstraction of the real world.it doesn't provide links to the detailed context of urban surfaces thus being not feasible for analysis or simulation purposes in terms of urban analysis, risk and disaster management, and city life-cycle management. A suitable model for those purposes should represent an eligible generalisation of subsets that meet the following requirements::

- The model must highlight both functional and structural relationships from one element to another
- The model must represent independent elements connecting them by relationship classes in order to enable specific simulations and complex analysis
- The model must be valid for different, heterogeneous types of transportation networks
- The model must be able to use simplified structures to reduce the complexity, meantime, preserving the required information to simulate, analyse, calculate, and cartographic visualise disaster scenarios (Adolphi et al., 2013)

The need to define a specific environmental model implies designing relevant features and their mutual relations explicitly, thus allowing the 3D topographical modelling of entire networks, sub-networks and network features, as well as their cartographic representations. Consequently, the network geospatial features have been treated as an abstraction of real-world objects (from the topographic point of view) and as network elements in a graph representation, making the model flexible and similar to those used in GIS utility systems (Adolphi et al. 2013).

What is new is about the desire to go beyond various conversion programmes. then the environment will be truly integrated, the modelling framework will be more universal, and data standards will overpass software programs and vendor-specific platforms (Zlatanova & Prosperi, 2005).

In this context, the validation process becomes a strategic step to certificate open and distributed environment, like a Spatial Data Infrastructure (SDI), where a high level of integration and interoperability is required. Therefore, the current gap between the conceptual and physical DB design and the implementation into a GIS system makes crucial the definition of spatial constraints addressed automatic validation and not only for documentation purposes (Pelagatti et al., 2009).

8.7 Some conclusions about the validation process

An in-depth analysis has been carried out on provided data to investigate the methodological aspects and the issues.

The highlighted examined aspects are listed below:

- The considerable variety of the provided data (images, topographic surveys, GPS tracks, aerial triangulation, numerical models, vector acquisition, GTDB, panoramic pictures and high-resolution laser telemetry), each with its peculiarities, has required high-level skills and competencies to be analysed;
- different evaluation processes have been harmonised considering some well-known best practices of a traditional photogrammetric survey and sampling/massive methods applied both to validate MMS and GTDB datasets;
- The huge amount of data refers to each delivered lot;
- The release frequency of the lots required provided validation outputs in a short time and high-level of confidence;
- The impossibility of carrying sequentially out the check operations for formal reasons;
- The impossibility of carrying out on-the-field surveys and validations due to the overall survey project’s spatial vastness.

A multidisciplinary working group has been set up to properly check evaluation steps, bringing together specific expertise for each phase of which the survey consists. Then, an additional step has been identified to validate the connection among different phases and overall outputs have been shared. Thus, each working group focused only on its specific issues, having had at the same time full confidence in the reliability of the control tests carried out on the other processes of the validation system.

Furthermore, automatic procedures have been developed for all those activities where the validation could be conducted by a series of operations in a cascade. related outcomes have been established addressed to a comparison with pre-set thresholds. in case of the automatic procedures would have been too onerous in machine-time, some suitable methodologies have been identified for selecting the minimum significative sample to make the test anyhow reliable.

The sampling methods have been then used to develop the procedures and balance the time necessary for the completion of the check operations.

Further refinements have been finally made to the first lots, after implementing the methodologies mentioned above, to make simpler and more robust the procedures in the presence of non-completely conform data.

8.8 The continuous updates

A recent need for many geographic applications concerns the ability to receive continuous updates as they undergo information integrations over time in the systems they serve. This dynamicity implies that the geospatial data, cannot remain static and without update over time. It is worth noticing that the information systems themselves can increase the amount of information that could be used to update datasets in their spatial or not spatial components. For dealing with the problem of continuous updates, some issues have to take into account :

1) “Since the different update processes collect spatial data in different ways and with different accuracy, it is necessary to manage instance-level accuracy information through specific metadata”. (Belussi et al., 2006)

2) “Continuous updates can be based on different kinds of observations: absolute coordinates and relative measurements and logical properties of updated or new objects are observed. In some cases, the different nature of the observations can lead to an observation conflict”. (Belussi et al., 2006)

3) Since continuous updates produce new observations, these new observations should be used to continuously enhance the quality of the database only for particular aspects such as spatial or alphanumerical properties. For instance, a better time resolution could conflict with a bad spatial accuracy than the previous state. So, for each stage of the update, some ranks and congruency rules have to be defined.

The choices made in this thesis allows the definition of an Information System which will continue for a long time, and this process will feed on increasing layer information (and related data management). The SDBMS has been designed founded on rules for sharing data and continuous integration of data. Hence, the integration process is focused on gradually growing contents by harmonising conceptual schemas, not all at the same time.

The update process related to the case study based on the MUIF project is at the moment in the design stage. However basic requirements are following declared. The update process model starts with the identification of changes in the railway infrastructure as the input determining the update flow to be considered.

To manage the different updates and the related database history, the SDBMS has been integrated to host data from different sources, multi-scale, multi-precision and multi-time.

The different source updates and a general flowchart have been preliminarily designed (Fig.72):

1) data coming from constructive/executive projects on the railway infrastructure;

2) endogenous secondary sources based on asset update LRS positioning;

3) changes of the asset type characteristics;

4) cyclical survey by diagnostic trains;

5) exogenous secondary sources.

Data from all update sources will need to be integrated into the SDBMS, allowing the general management of railway assets and the territorial environment as well as medium and long-term historical data analysis. Moreover, the update flow also allows the database manager to monitor the progress of the updates.

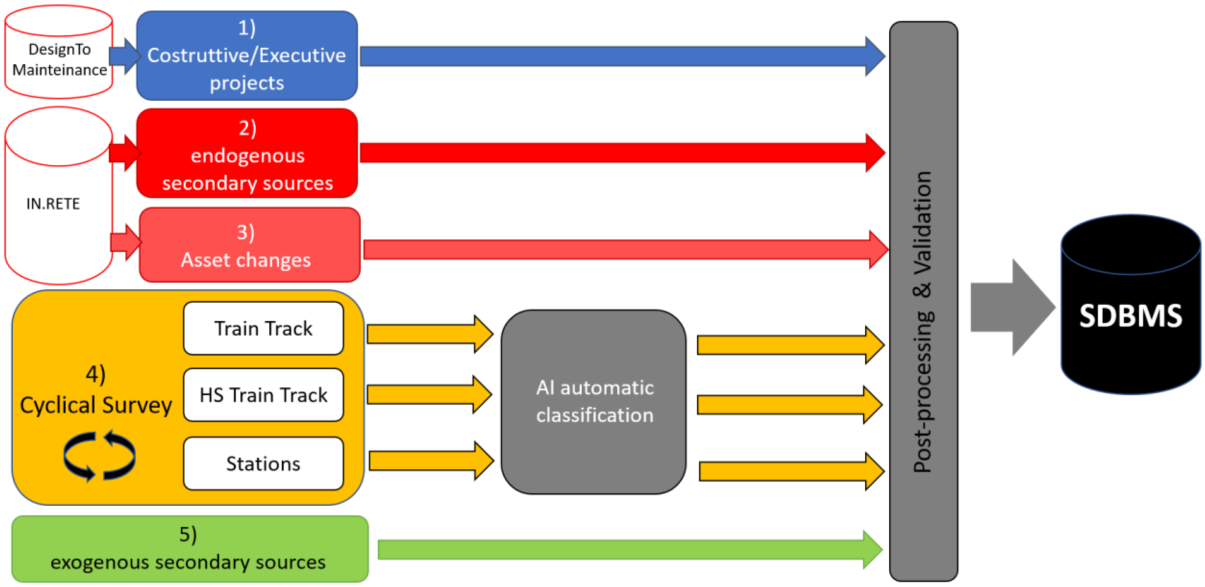


Figure 72 - Basic update flowchart

According to the need to describe metadata at the instance level, relating to the update process's spatial/temporal accuracy, a list of four fields has been added as attributes to each class of the DBMS (Table 4).

Field	Description	DataType	Domain		
			Code	description	category
Source	Source of observation	CodeList	FP	Primary source	Existing
			PE	Executive project	1
			PC	Executive project	1
			PAB	As-Built	1
			FOS_AGEA	exogenous secondary sources (AGEA)	5
			FOS_AE	exogenous secondary sources cadastral data	5
			LRS	endogenous secondary sources	2
			COORD	projection	2
			AC_TR	Track merge	3
			SC_TR	Track split	3
			AG_CI	Cyclical Update	4
Data	Update data	TimeStamp	DD:MM:YYYY		
State	Update State	CodeList	Removed Updated Existing		
3D	Geometry Dim	CodeList	2D 3D		

Table 4 - Metadata at instance level useful to manage the update process

9. CHAPTER 9 – CONCLUSIONS

9.1 Conclusions

The main topic of this thesis is interoperability in geospatial systems across different domains. The obtained results relate to a process rather than a specific case. Different steps are taken into account to reuse and combine the information contents in shared data models, starting from a reference cartographic database. The railway infrastructure domain represents the case study to test the interoperability between different geospatial models. In all the considered research steps, the obtained results always start from a conceptual approach. However, a physical solution has not always been reached. Some specific aspects, as far as they need further investigation, have been treated from a general point of view because out of the thesis scope.

Many research questions have found practical answers across the thesis because the case study refers to a real project at a national level in Italy. Some others remain open issues for future developments because they need to be monitored over time to be aware of their feasibility. Some of the research questions have been joined in a congruent context to correctly answer, referencing the chapters and contributions.

9.1.1 Interoperability levels and related standard (problem 1, paragraph 1.3.1.)

As a general approach, interoperability has been carried out avoiding translation from one format to another. The intermediate models as links between domains to be connected (cartography, Geographic Information Systems GIS, Building Information Modelling BIM, Spatial Data Infrastructures etc.) have been considered, maintaining the original models for each phase. The concept of interoperability has been adopted at a higher level as possible, according to the deepness of the subtopic research and the case study datasets/data models' availability. As a result, about the case study datasets (Chapters 5, 6, 7, and 8) the maximum level of the interoperation (shown in Fig. 2) has been achieved, considering both the modelling and the abstraction of semantics to obtain an integrated Spatial Database Management System (SDBMS). Topographic and railway assets have been modelled according to the same conceptual and application schema. Hence, the interoperability has been carried out following the same approach to connect the geographic reference information and the railway thematic environmental/infrastructure spatial applications (Chapter 6).

Nevertheless, all the connections have been studied firstly considering the conceptual level because an abstraction level can avoid physical implementations and dependent platform solutions.

However, some real data tests have been conducted in the specific tasks (Chapter 7 to connect point clouds with SDBMS, Chapter 8 about 3D modelling) to demonstrate the feasibility of the theoretically proposed solutions. Indeed, procedures described in these chapters should be considered as experimental test and in a preliminary stage, not extended to the whole case study. From the implementation point of view, some procedures had not been optimised, because of this research's scope. The efforts have been addressed to follow the more straightforward solution, although not necessarily the better.

Finally, some research aspects are already at the design stage to allow a continuous update process (Chapter 8), physical implementation is almost ready to start (at the beginning of 2021).

9.1.2 The starting point, 2D-3D evolution (problem 2, paragraph 1.3.2.)

The case study refers to the Italian national coverage, considering different standards with different details, based on the reference cartography integrated with specific railway assets. For this reason, the research did not start from scratch; choices have been contextualised according to the use case project. Nevertheless, always standard compliance has guided the process.

The case study, except for specific railway assets, refers to the geotopographic information according to the Italian national law (MD, 2012) as explicitly acquired for the MUIF project. The acquisition of new homogeneous reference cartography is due to the lack in Italy of a unique national agency of geographic information, that has determined different reference cartographic maps at larger scales (1:1.000 – 10.000 scale cartography) over the years, not useful for the case study scope. Nowadays, the transposition of national law (MD, 2012) by the different Public Administrations that deliver geographic information is already an ongoing process. However, it is worth noting that the national law of geotopographic information and the GeoUML methodology (SpatialDBGroup, 2011) refers to a conceptual data model of content specifications. Hence, their implementation depends on the technological solutions adopted by every single Public Administration that delivers its reference cartography. As a result, many reference maps at a larger scale are not interoperable at the physical level.

One of the main heterogeneous characteristics of Geographic Information is related to the third-dimension modelling because topological constraints have been implemented in the GIS only for the planimetric geometry types. In addition to that, not many experiments of 3D city models in Italy have been carried out in wider territories compared to the case study.

In Chapter 2 the definition of the Italian GIS specifications and the compliance with the GeoUML methodology have been explained in detail to define the state of the art of the cartography reference and prepare the requirements in the integrated GeoTopographic DataBase (GTDB) of the case study.

In Chapter 3 the SienaGTDB has been explained as an example of an intermediate solution of a 3D cartography toward a 3D city model and compliant with national specifications on GTDB, then the same approach has been adopted in the GTDB case study (Chapter 5). The 3D approach has been used both to discuss BIM-GIS connection (Chapter 4) and to extend the 3D modelling to specific assets (Chapter 8).

9.1.3 BIM-GIS connection (problem 3, paragraph 1.3.3.)

The BIM-GIS connection (Chapter 4) has been analysed only from a theoretical level, without any physical transformation. This approach has been addressed to the unavailability of datasets based on either a BIM or 3D City model standard related to railway infrastructures (OGC LandInfra, 2016; OGC InfraGML, 2017). Moreover, the adopted 3D City model could be considered an intermediate solution toward a full B-rep geospatial model (Foley et al., 1995).

The literature mainly describes the translation from BIM to GIS, whereas in the thesis a vice-versa approach has been considered. Despite a predominant BIM to GIS transposition, the orientation on GIS environment has suggested a definition of a metamodel based on the CityGML standard (OGC CityGML, 2012) to be used as a link between the two domains, avoiding any translation and consequently loss of information toward BIM. However, the profiling toward infrastructures themes, despite a strictly building context of the CityGML, suggests that there will be further developments of the Landinfra/InfraGML standards (OGC LandInfra, 2016; OGC InfraGML, 2017) applied specifically to the case of railways.

Similarly, from the perspective of BIM, further developments in the domain of infrastructures will probably happen. As a result, also the BIM-GIS connection will evolve orienting more on territorial/infrastructural aspects. For these reasons, the analysis conducted in Chapter 4 should be considered only in a preliminary stage.

9.1.4 Reference Vs Railway Thematic GIS application (problem 4, paragraph 1.3.4.)

Focusing on the railway context (Chapter 6), in the last decades some reference models have been defined and implemented to manage, in a georeferenced environment, railway assets and the transportation network. The main characteristic of these standards, namely RTM/RailML standards (IRS 30100, 2016) are predominantly oriented on a network point of view, so any spatial object is positioned as a displacement from the track centreline measured by a linear reference system (LRS), despite territorial objects are interlocked following a spatial topology applied to a triplet of coordinates (x, y, z). Spatial and Linear systems are not topologically compatible. So data have to be modelled in different datasets even in the same DB. The chosen solution is addressed to combine assets both belonging to the Topographic and Railway DBs as source objects. Assets have been implemented as events on network routes based on LRS and integrated into the GTDB as spatial objects. Topological constraints between elements/junctions and network routes have been validated, linking then the two models. Again, the dataset link approach has been chosen instead of translating the information in this case.

9.1.5 Point clouds in the SDBMS integration (problem 5, paragraph 1.3.5.)

Thanks to technological evolution, geographical data are continuously updated using different sources most effectively. Sources and sensors related to observations need to be differentiated according to spatial and temporal accuracies. The integrated SDBMS used as a case study could be congruently applied to another thematic spatial database, following the same approach used to integrate the Geotopographic database with railway infrastructures.

The correctness of the chosen model, independently from the specific spatial thematic level, is guaranteed by international standards on geomatics and applicative shared semantic models.

Finally, a test to integrate point clouds data into an SDBMS has been carried out (Chapter 7). This phase has been considered for two main reasons: the first one is addressed to manage through a hyperlink the connection between a vector feature and point clouds related to an asset, the second one focuses on the automatic classification of assets from point clouds to be detected in a near-real-time survey. Both phases aim to support designing informative flows in the continuous updates of an integrated SDBMS (Chapter 8).

9.2 Futures developments

As mentioned above, for the generality in the treatment of interoperability pathways, the thesis is a point of arrival as a starting point for future research developments. Some of these points briefly explained below, are the natural evolution of the thesis topics and already in an ongoing start-up phase.

9.2.1 3D city modelling for infrastructures

One of the first future work will concern the development of 3D modelling in GIS. So far it has concerned 3D city modelling with a predominant focus on building theme, both in GIS and BIM environment. However, in recent years, there has emerged the need to develop detailed 3D geographical elements in vaster geographical contexts as in transport infrastructures; a detailed description of each part must be framed in the territory. Some future studies could be addressed to test the intermediate models considering different thematic applications (Billen et al., 2015) other than railways, such as geology, hydrology, road transportation, hazard events. These datasets could be aimed at Disaster Management (Kemec et al., 2009) and Hazard and Risk validation applications.

The most widely standard for 3D city model implementation is the CityGML (OGC CityGML, 2012), mainly currently limited to urban contexts. The LandInfra/InfraGML standards (OGC LandInfra, 2016; OGC InfraGML, 2017) seem to be more suitable for territorial purposes, but not yet so widespread. Moreover, also on the BIM side, it is necessary to achieve the development of semantics, suitable for infrastructures rather than buildings. For this reason, a probable future work refers to a physical implementation of the railway case study according to LandInfra, LandGML standards.

9.2.2 Artificial Intelligence (AI) and point clouds role

Another future development will concern implementing more performant informatic solutions considering the update of massive information during the time. Some of the research steps have been carried out as a preliminary test of experimental phases to demonstrate the feasibility of some pointed solutions as it has happened for procedures implemented in Chapter 7 about point clouds source data and their automatic segmentation.

The application’s context has been addressed to find a solution to connect point clouds with geospatial features in an SDBMS, intending to interrelate source with processed geospatial information. So a multi-dimensional/multi-source SDBMS has been implemented.

The main advantage of this approach refers to the possibility to continuously update such SDBMS automatically. The AI techniques address this scope, where the algorithm has been trained with a ground truth considering the first SDBMS implementation process. The same algorithm should then be used to classify and automatically segment point clouds data sources of future cyclical updates.

However, optimisation of informatic procedures and the extension of the number and type of assets, need to be considered.

This perspective is aimed to support the management and validation of the update flow related to the cyclical survey as the design in Chapter 8 (Fig. 72).

In general, all the requirements for the integrated SDBMS have been implemented in the case study and then validated (Chapter 8) although they have been designed to support the SDBMS management over time. As future development, it is planned to tune these requirements to support continuous and multi-source update campaign. A specific procedure to validate the correctness of these updates will be designed based on the one applied for the first data delivery (Chapter 8).

9.2.3 SDBMS and relationship with Geoportal Web Services

The thesis's basic approach follows a *data-centric* point of view, focusing on management and updating geospatial contents over time. Therefore, predominant interoperability goes through WEB applications (Geoportals) and WEB services, where the approach follows a mainly *service-centric* point of view. Consequently, in a W3C[23] (World Wide Web) environment simpleness and linearity of datasets are requirements to obtain interoperability, while SDBMS focuses on completeness and structured information through relationship and constraints, complexifying the usability explicitly. For these reasons, it seems that data-centric and service-centric point of views addresses different scopes. Sometimes implemented solutions refer to different geospatial models. By the way, the interoperability on the web impacts the SDBMS data model and vice-versa. Recently some existing experiences such as (Yao et al., 2018), (Jetlund et al., 2020), (Basanow et al., 2008), (OGC W3DS, 2005) just to name a few, make evident an increasing interest in the development of interfaces and geoportals through specific application server implementation. Besides this context, future work should be aimed to make easier access to integrated SDBMSs in a user-friendly approach, as it happens for the INSPIRE-compliant Spatial Data Infrastructures (SDIs) (Directive 2007/2/EC., 2007). In that case, the SDI will be based on OpenGeospatial Web (OGC OWS)[24] Services and a shared metadata catalogue (OGC CS, 2014).

---

[23] https://www.w3.org/standards/

[24] https://www.ogc.org/standards/owc

9.3 References

Adolphi, T. & Nagel, C. & Heinrich, T. & Kolbe T. (2013). Semantic 3D Modeling of Multi-Utility Networks in Cities for Analysis and 3D Visualization. DOI: 10.1007/978-3-642-29793-9_3 - In book: Progress and New Trends in 3D Geoinformation Sciences Project: BIM and GIS integration.

Aleksandrov, M. & Diakité, A. & Yan, J. & Li, W. & Zlatanova, S. (2019): systems architecture for management of BIM, 3D GIS and sensors data. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-4/W9, 3–10, https://doi.org/10.5194/isprs-annals-IV-4-W9-3-2019.

Arastounia, M. (2017). An enhanced algorithm for concurrent recognition of rail tracks and power cables from terrestrial and airborne LIDAR point clouds. Infrastructures, 2(2), 8.

Arastounia, M. & Oude Elberink, S. (2016). Application of template matching for improving classification of urban railroad point clouds. Sensors, 16(12), 2112.

Arroyo Ohori, K. & Diakité, A. & Krijnen, T. & Ledoux, H. & Stoter, J. (2018). Processing BIM and GIS Models in Practice: Experiences and Recommendations from a GeoBIM Project in The Netherlands. ISPRS Int. J. Geo-Inf. 2018, 7, 311

Basanow, Jens & Neis, Pascal & Neubauer, Steffen & Schilling, Arne & Zipf, Alexander. (2008). Towards 3D Spatial Data Infrastructures (3D-SDI) based on open standards — experiences, results and future issues. 10.1007/978-3-540-72135-2_4.

Belussi, A. & Brovelli, M.A. & Negri, M. & Pelagatti, G. & Sansò, F. (2006). Dealing with Multiple Accuracy Levels in Spatial Databases with Continuous Update. Proceeding on The Seventh International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Spatial Accuracy 2006). Lisbon, Portogallo. 5 - 7 Luglio 2006. pp. 203-212.

Belussi, A. & Liguori, F. & Marca, J. & Migliorini, S. & Negri, M. & Pelagatti, G. & Visentini, P. (2011). Validation of Geographical Datasets against Spatial Constraints at Conceptual Level, Proceeding of The Urban Data Management Symposium (UDMS 2011). Delft, The Netherlands, Sept. 28-30, 2011.

Bensalah, M. & Elouadi, A. & Mharzi, H. (2018). Integrating BIM in railway projects: Review & perspectives for Morocco & Mena. Int. J. Recent Sci. Res. 2018, 9, 23398–23403.

Ben-Shabat, Y. & Lindenbaum, M. & Fischer, A. (2018). 3DmFV: Three-dimensional point cloud classification in real-time using convolutional neural networks. IEEE Robotics and Automation Letters, 3(4), 3145–3152.

Biljecki, Filip & Stoter, Jantien & Ledoux, Hugo & Zlatanova, Sisi & Coltekin, Arzu. (2015). Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information. 4. 2842-2889. 10.3390/ijgi4042842.

Biljecki, Filip & Ledoux, Hugo & Stoter, Jantien. (2016). An improved LOD specification for 3D building models. Computers Environment and Urban Systems. 59. 25-37. 10.1016/j.compenvurbsys.2016.04.005.

Biljecki, Filip & Ledoux, H. & Stoter, Jantien. (2016b). Generation of multi-LOD 3D city models in CityGML with the procedural modelling engine Random3Dcity. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. IV-4/W1. 51-59. 10.5194/isprs-annals-IV-4-W1-51-2016.

Biljecki, F. & Kumar, K. & Nagel, C. (2018). CityGML Application Domain Extension (ADE): overview of developments. Open geospatial data, softw. stand. 3, 13. https://doi.org/10.1186/s40965-018-0055-6

Billen, R. & Cutting-Decelle, A.F.& Métral, C. & Falquet, G. & Zlatanova S. & Marina O. (2015). Challenges of semantic 3D city models. Int. J. 3-D Inf. Model, 4, 68–76, (2015)

Bitelli, G. & Dubbini, M. & Zanutta, A. (2004). Terrestrial laser scanning and digital photogrammetry techniques to monitor landslide bodies. International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences, 35(B5), 246–251.

Boreggio, M. & Bernard, M. & Gregoretti, C. (2018). Evaluating the differences of gridding techniques for Digital Elevation Models generation and their influence on the modeling of stony debris flows routing: A case study from Rovina di Cancia basin (North-eastern Italian Alps). Frontiers in Earth Science, 6, 89.

Bosschaart, M. & Quaglietta E. & Janssen, B. & Goverde, R. M.P. (2015). Efficient formalization of railway interlocking data in RailML. Information Systems, Volume 49, Pp. 126-141. ISSN 0306-4379. https://doi.org/10.1016/j.is.2014.11.007.

Bleiholder, J.& Naumann, F. (2009). Data fusion. ACM Comput. Surv. (CSUR) 2009, 41, 1.

bSI IFC Railway. (2015). Railway BIM Data Standard (version 1.0). CRBIM1002 China Railway BIM Alliance. https://www.buildingsmart.org/wp-content/uploads/2017/09/bSI-SPEC-Rail.pdf

Dong, J.; Zhuang, D.; Huang, Y.; Fu, J. Advances in multi-sensor data fusion: Algorithms and applications. Sensors 2009, 9, 7771–7784.

Carrion et al. (2008) Metodi e modelli per il controllo di qualità di data base topografici multiscala. Atti 12a Conferenza Nazionale ASITA, 21 -24 ottobre 2008, L'Aquila, IT, ISBN 978-88-903132-1-9.

Cecchini, C. (2019). From data to 3d digital archive: a GIS-BIM spatial database for the historical centre of Pavia (Italy). Electronic Journal of Information Technology in Construction. 24. 459-471.

Che, E. & Jung, J. & Olsen, M.J. (2019). Object Recognition, Segmentation, and Classification of Mobile Laser Scanning Point Clouds: A State of the Art Review. Sensors 2019, 19, 810.

Chen, S.-E. (2012). Laser scanning technology for bridge monitoring. Laser Scanner Technology, 71.

Clementini, E. & Di Felice, P. (1996). A model for representing topological relationships between complex geometric features in spatial databases. Inform. Syst. 90, 1–4, 121–136.

Colucci, E. & De Ruvo, V. & Lingua, A. & Matrone, F. & Rizzo, G. (2020). HBIM-GIS Integration: From IFC to CityGML Standard for Damaged Cultural Heritage in a Multiscale 3D GIS. Appl. Sci. 2020, 10, 1356.

Corongiu, M. & Cannafoglia, C. & Desideri, M. & Rossi M. (2004). Intesagis: The Basis for a National Spatial Data Infrastructure. Proceeding of the 10th EC-GI&GIS Workshop. June 23-25, 2004, Warsaw, Poland.

Corongiu, M. & Galetto, R. & Rossi, M. & Spalla, A. (2006). Cartografia numerica per i database Topografici e il 3D city model dei centri storici. Bollettino SIFET, 1/06: 45-68 SIFET, Bollettino della Società Italiana di fotogrammetria e topografia.

Corongiu, M. & Tucci, G. & Santoro, E. & Kourounioti, O. (2018). Data integration of different domains in geo-information management: a railway infrastructure case study. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-4, 121-127, https://doi.org/10.5194/isprs-archives-XLII-4-121-2018

Corongiu, M. & Masiero, A. & Tucci G. (2020). Classification of railway assets in mobile mapping point clouds. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B1-2020, 219–225. DOI: 10.5194/isprs-archives-XLIII-B1-2020-219-2020

Directive 2007/2/EC. (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002

Döllner, J. & Baumann, K. & Buchholz, H. (2006). Virtual 3D city models as foundation of complex urban information spaces. In: 11th international conference on Urban Planning and Spatial Development in the Information Society (REAL CORP). Essen. Germany. (pp. 107–112).

Dong, J. & Zhuang, D. & Huang, Y. & Fu, J. (2009). Advances in multi-sensor data fusion: Algorithms and applications. Sensors 2009, 9, 7771–7784.

Egenhofer, M. J. & Clementini, E. & Di Felice, P. (1994). Topological relations between regions with Holes. Int. J. Geograph. Inform. Syst. 8, 2, 128–142.

Elberink, S. O. & Khoshelham, K. (2015). Automatic extraction of railroad centerlines from mobile laser scanning data. Remote sensing, 7(5), 5565.

Ellul, C. & Boyes, G. & Thomson, C. & Backes, D. (2017). Towards Integrating BIM and GIS – An End-to-End Example from Point Cloud to Analysis. In: ©Springer International Publishing AG 2017. Advances in 3D Geoinformation. A. Abdul-Rahman (ed.). Lecture Notes in Geoinformation and Cartography. DOI 10.1007/978-3-319-25691-7_28.

Foley, J. & Van Dam, A. & Feiner, S. & Hughes, J. (1995). Computer Graphics: Principles and Practice. Addison Wesley, 2nd Ed.

Gézero, L. & Antunes, C. (2019). Automated Three-Dimensional Linear Elements Extraction from Mobile LIDAR Point Clouds in Railway Environments. Infrastructures 2019, 4, 46.

Govardanan, C. S. & Gnanapandithan, L. P. G. (2020). SMSS: does social, mobile, spatial and sensor data have high impact on big data analytics. International Journal of Intelligent Enterprise 2020 7:1-3, 215-233. https://doi.org/10.1504/IJIE.2020.104657

Gröger, G., Reuter, M., & Plümer, L. (2004). REPRESENTATION OF A 3-D CITY MODEL IN SPATIAL OBJECT-RELATIONAL DATABASES. In proceedings of the 20th Congress of ISPRS. Istanbul. Turkey.

Guarnieri, A.& Masiero, A. & Vettore, A. & Pirotti, F. (2015). Evaluation of the dynamic processes of a landslide with laser scanners and Bayesian methods. Geomatics, Natural Hazards and Risk, 6(5-7), 614–634.

Hlubuček, A. (2017). RailTopoModel and RailML 3 In Overall Context. Acta Polytechnica CTU Proceedings. 11. 16. 10.14311/APP.2017.11.0016. INSPIRE D2.8.I.7. (2014). Data Specification on Transport Networks, Technical Guidelines INSPIRE Infrastructure for Spatial Information in Europe https://inspire.ec.europa.eu/id/document/tg/tn

INSPIRE D2.10.1 (2013). Data Specifications – Base Models – Generic Network Model https://inspire.ec.europa.eu/documents/Data_Specifications/D2.10.1_GenericNetworkModel_v1.0rc3.pdf

IRS 30100 (2016). RailTopoModel - Railway infrastructure topological model. IRS International Railway Solution.

Isikdag, U. & Zlatanova, S. (2009). Towards Defining a Framework for Automatic Generation of Buildings in CityGML Using Building Information Models. In 3D Geo-Information Sciences. Springer. Berlin Heidelberg. pp. 79—96. https://doi.org/10.1007/978-3-540-87395-2_6.

ISO/TC 211 19103. (2015). ISO/IS 19103:2015 Geographic information — Conceptual schema language. International Organization for Standardization (ISO)

ISO/TC 211 19125-1. (2004) ISO 19125:2004 Geographic information — Simple feature access — Part 1: Common architecture. International Organization for Standardization (ISO)

ISO/TC 211 19107. (2002). ISO/IS 19107:2002 Geographic information—Spatial schema. International Organization for Standardization (ISO)

ISO/TC 211 19109. (2019). ISO/FDIS 19109:2005(E) Geographic information—Rules for application schema. International Organization for Standardization (ISO)

ISO/TC 211 19125. (2004). ISO 19125-1:2004 Geographic information — Simple feature access — Part 1: Common architecture. International Organization for Standardization (ISO)

ISO/TC 211 19131. (2018). ISO/CD 19131:2018 Geographic information—Data Product Specifications. International Organization for Standardization (ISO)

ISO/TC 211 19148. (2012). ISO/FDIS 19148:2012. Geographic information—Linear referencing. International Organization for Standardization (ISO)

ISO/TC 211 19157. (2013). ISO/FDIS 19157:2013. Geographic information— Data Quality. International Organization for Standardization (ISO)

ISO/TC 211 19166. (2018). ISO/CD 19166:2018. Geographic information – BIM to GIS conceptual mapping (B2GM). International Organization for Standardization (ISO).

ISO/TC 59/SC 13. (2018). ISO 16739-1:2018. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries—Part 1: Data schema. International Organization for Standardization (ISO)

ISO/TC 59/SC 13. (2018). ISO 16739-1:2018. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema. International Organization for Standardization (ISO)

ISO 14825. (2011). Intelligent transport systems - Geographic Data Files (GDF) - GDF5.0. International Organization for Standardization (ISO). https://www.iso.org/obp/ui/#iso:std:iso:14825:ed-2:v1:en

ISO 3534-2 (2006) Applications of statistical methods Statistics — Vocabulary and symbols — Part 2: Applied statistics. International Organization for Standardization (ISO)

ISO 10303. (2004). ISO 10303-11:2004. Industrial automation systems and integration — Product data representation and exchange — Part 11: Description methods: The EXPRESS language reference manual. International Organization for Standardization (ISO)

Jaakkola, T. & Haussler, D. (1999). Exploiting generative models in discriminative classifiers. Advances in neural information processing systems, 487–493.

Jetlund, K. & Onstein, E. & Huang, L. (2020). IFC Schemas in ISO/TC 211 Compliant UML for Improved Interoperability between BIM and GIS. ISPRS Int. J. Geo-Inf. 2020, 9, 278.

Kavisha, K. (2020). Modelling and managing massive 3D data of the built environment. https://doi.org/10.4233/uuid:47218911-c93d-4295-a3de-231d023c1743.

Kemec S. & Zlatanova S. & Duzgun S. (2009). Selecting 3D Urban Visualisation Models for Disaster Management: a Rule-based Approach, In: Proceedings of TIEMS 2009 Annual Conference (Sahin, Drager, eds.), Istanbul, pp. 99-110, (2009)

Kumar, K. & Labetski, A. & Ohori, K. & Ledoux, H. & Stoter, J. (2019). The LandInfra standard and its role in solving the BIM-GIS quagmire. Open Geospatial Data, Software and Standards. 4. 5. 10.1186/s40965-019-0065-z.

Kolbe, T. (2009). Representing and Exchanging 3D City Models with CityGML. 3D Geo-Information Sciences pp. 15-31. Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg. DOI 10.1007/978-3-540-87395-2_2.

Kolbe, T., Burger, B.S., & Cantzler, B. (2015). CityGML goes to Broadway.

Krizhevsky, A. & Sutskever, I. & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 1097–1105.

Kuhn, W. (2001). Ontologies in support of activities in geographical space. International Journal of Geographical Information Science. 15. 613-631. 10.1080/13658810110061180.

Kurwi, S. & Demian, P., & Hassan, T. (2017). Integrating BIM and GIS in railway projects: A critical review. Chan, P.W. & Neilson, C. J. (Ed.). 33rd Annual ARCOM Conference, Cambridge, UK. pp. 45-53.

Li, W. & Zlatanova, S. & Diakite, A. A. & Aleksandrov, M. & Yan, J. (2020). Towards Integrating Heterogeneous Data: A Spatial DBMS Solution from a CRC-LCL Project in Australia. ISPRS Int. J. Geo-Inf. 2020, 9(2), 63. https://doi.org/10.3390/ijgi9020063.

Liu, X. & Wang, X. & Wright, G. & Cheng, J.C.P. & Li, X. & Liu, R. (2017). A State-of-the-Art Review on the Integration of Building Information Modeling (BIM) and Geographic Information System (GIS). ISPRS Int. J. Geo-Inf. 2017, 6, 53.

López, F.J. & Lerones, P.M. & Llamas, J. & Gómez-García-Bermejo, J. & Zalama, E. (2018). A Review of Heritage Building Information Modeling (H-BIM). Multimodal Technologies Interact., 2, 21. doi.org/10.3390/mti2020021.

Löwner, M.-O. & Benner, J. & Gröger, G. & Häfele., K.-H. (2013). New concepts for structuring 3D city models - An extended level of detail concept for CityGML buildings. In Computational Science and Its Applications – ICCSA 2013, pages 466–480. Springer.

Löwner, M.-O. & Gröger, G. & Benner, J. & Biljecki, F. & Nagel, C. (2016). Proposal for a new lod and multi-representation concept for citygml. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4.

Lu, L. & Becker, T. &, Löwner, M.-O. (2017). 3D Complete Traffic Noise Analysis Based on CityGML. In: ©Springer International Publishing AG 2017. A. Abdul-Rahman (ed.). Advances in

3D Geoinformation. Lecture Notes in Geoinformation and Cartography. DOI 10.1007/978-3-319-25691-7_15.

Maalek, R. & Lichti, D. D. & Ruwanpura, J. Y. (2018). Robust segmentation of planar and linear features of terrestrial laser scanner point clouds acquired from construction sites. Sensors, 18(3), 819.

Masiero, A. & Guarnieri, A. & Pirotti, F. & Vettore, A. (2015). Semi-Automated Detection of Surface Degradation on Bridges Based on a Level Set Method. ISPRS - International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(3), 15–21.

Matrone, Francesca & Colucci, Elisabetta & De Ruvo, Valeria & LINGUA, Andrea & Spano, Antonia. (2019). HBIM IN A SEMANTIC 3D GIS DATABASE. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.. XLII-2/W11. 857-865. 10.5194/isprs-archives-XLII-2-W11-857-2019.

Ministerial Decree of 10 November 2011. (2012). DECREE of 10 November 2011 (2012) Regole tecniche per la definizione delle specifiche di contenuto dei database geotopografici. (12A01800) (GU Serie Generale n.48 del 27-02-2012 - Suppl. Ordinario n. 37).

Nash, A. & Huerlimann, D. & Schuette, J. & Krauss, V. (2004). RailML - A standard data interface for railroad applications. 10.2495/978-1-84564-500-7/01.

Neves, J. & Sampaio, Z. & Vilela, M. (2019). A Case Study of BIM Implementation in Rail Track Rehabilitation. Infrastructures. 4. 8. 10.3390/infrastructures4010008.

Nuttens, T. & De Breuck, V. & Cattor, R. & Decock, K. & Hemeryk, I. (2018). Using BIM models for the design of large rail infrastructure projects: Key factors for a successful implementation. Int. J. Sustain. Dev. Plan. 2018, 13, 73–83.

Neubert, M. & Hecht, R. & Gedrange, C. & Trommler, M. & Herold, H. & Krüger, T. & Brimmer, F. (2008). Extraction of railroad objects from very high resolution helicopter-borne LIDAR and ortho-image data. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38, 25–30.

Noardo, F. & Ohori, K. & Biljecki, F. & Ellul, C. & Harrie, L. & Krijnen, T. & Eriksson, H. & Liempt, J. & Pla, M. & Ruiz, A. & Hintz, D. & Krüger geb. Amiri, N. & Leoni, C. & Leoz, L. & Moraru, D. & Vitalis, S. & Willkomm, P. & Stoter, J. (2020). Reference study of CityGML software support: the GeoBIM benchmark 2019 -- Part II.

OGC CityGML (2012). Open Geospatial Consortium - City Geography Markup Language (CityGML) Encoding, Standard 2.0.0. https://www.ogc.org/standards/citygml

OGC CS (2014). Open Geospatial Consortium - Catalogue Services 3.0 - General Model https://www.ogc.org/standards/cat

OGC GML (2016). Open Geospatial Consortium - Geography Markup Language (GML) Encoding Standard, standard 3.2.1. https://www.ogc.org/standards/gml

OGC InfraGML (2017). Open Geospatial Consortium - InfraGML 1.0: Part 0 – LandInfra Core - Encoding Standard https://www.ogc.org/standards/infragml

OGC LandInfra (2016). Open Geospatial Consortium - Land and Infrastructure Conceptual Model Standard (LandInfra). https://www.ogc.org/standards/landinfra.

OGC SLD (2007). Open Geospatial Consortium - Styled Layer Descriptor Profile of the Web Map Service Implementation Specification. https://www.ogc.org/standards/sld.

OGC SE (2006). Open Geospatial Consortium - Symbology Encoding Implementation Specification. https://www.ogc.org/standards/se.

OGC W3DS (2005). Open Geospatial Consortium - Web 3D Service. https://portal.opengeospatial.org/files/?artifact_id=8869.

Ohori, K.A. & Biljecki, F. & Diakité, A. & Krijnen, T. & Ledoux, H. & Stoter, J. (2017). TOWARDS AN INTEGRATION of GIS and BIM DATA: WHAT ARE the GEOMETRIC and TOPOLOGICAL ISSUES? In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, (Copernicus GmbH), pp. 1–8.

Oosterom, P. & Martinez Rubi, O. & Ivanova, M. & Horhammer, M. & Geringer, D. & Ravada, S. & Tijssen, T. & Kodde, M. & Goncalves, R. (2015). Massive point cloud data management: design, implementation, and execution of a point cloud benchmark. Computers & Graphics. 49, June 2015, Pages 92-125. https://doi.org/10.1016/j.cag.2015.01.007

Oosterom, P. & Martinez-Rubi, O. & Tijssen, T. & Gonçalves, R. (2017). Realistic Benchmarks for Point Cloud Data Management Systems Advances in 3D Geoinformation. In: ©Springer International Publishing AG 2017. A. Abdul-Rahman (ed.). Advances in 3D Geoinformation. Lecture Notes in Geoinformation and Cartography. DOI 10.1007/978-3-319-25691-7_1.

Park, H. S. & Lee, H. & Adeli, H. & Lee, I. (2007). A new approach for health monitoring of structures: terrestrial laser scanning. Computer-Aided Civil and Infrastructure Engineering, 22(1), 19–30.

Pastucha, E. (2016). Catenary system detection, localization and classification using mobile scanning data. Remote Sensing, 8(10), 801.

Pelagatti, G. & Negri, M. & Belussi, A. & Migliorini, S. (2009). From the conceptual design of spatial constraints to their implementation in real systems. 448-451. 10.1145/1653771.1653841.

Peters, S. & Jahnke, M. & Murphy, C.E. & Meng, L. & Abdul-Rahman, A. (2017). Cartographic Enrichment of 3D City Model - State of the Art and Research Perspectives. In: ©Springer International Publishing AG 2017, A. Abdul-Rahman (ed.). Advances in 3D Geoinformation. Lecture Notes in Geoinformation and Cartography. DOI 10.1007/978-3-319-25691-7_12.

Pocobelli, D. P.& Boehm, J.& Bryan, P.& Still, J.& Grau-Bové, J. (2018). BIM for heritage science: a review. Heritage Science 6:30. doi.org/10.1186/s40494-018-0191-4.

Psomadaki, S. (2016). Using a Space Filling Curve for the management of dynamic point cloud data in a relational DBMS. Master's thesis, TU Delft.

Qi, C. R. & Su, H. & Mo, K. & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 652–660.

Rodríguez-Cuenca, B. & García-Cortés, S. & Ordóñez, C. & Alonso, M.C. (2015). Automatic Detection and Classification of Pole-Like Objects in Urban Point Cloud Data Using an Anomaly Detection Algorithm. Remote Sens. 2015, 7, 12680-12703.

Ross, L. (2011). Virtual 3D City Models in Urban Land Management - Technologies and Applications. 10.14279/depositonce-2744.

S´anchez, J. & Perronnin, F. & Mensink, T. & Verbeek, J. (2013). Image classification with the Fisher vector: Theory and practice. International journal of computer vision, 105(3), 222–245.

Shapiro, V. (2002). Boundary representation: a compromise. In Handbook of Computer Aided Geometric Design, Chapter 20 - Solid Modeling. Pages 473-518, ISBN 9780444511041, https://doi.org/10.1016/B978-044451104-1/50021-6

Song, Y.& Wang, X. & Tan, Y. & Wu, P. & Sutrisna, M. & Cheng, J.C.P. & Hampson, K. (2017). Trends and Opportunities of BIM-GIS Integration in the Architecture, Engineering and Construction Industry: A Review from a Spatio-Temporal Statistical Perspective. ISPRS Int. J. Geo-Inf. 2017, 6, 397.

SpatialDBGroup. (2011). GeoUML Methodology and Tools. An Overview. Documents on https://spatialdbgroup.polimi.it/?page_id=41

SpatialDBGruop. (2011). GeoUML Model – Geometric Model and OCL Constraints Templates. Documents on https://spatialdbgroup.polimi.it/?page_id=41

Stadler, A. & Kolbe, T. (2007). Spatio-semantic coherence in the integration of 3D city models. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XXXVI-2/C43, pp. 8.

Stoter J. et al. (2008). A Data Model for Multi-scale Topographical Data. In: Ruas A., Gold C. (eds) Headway in Spatial Data Handling. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68566-1_14

Tack, F. & Buyuksalih, G. & Goossens, R. (2012). 3D building reconstruction based on given ground plan information and surface models extracted from spaceborne imagery. Isprs Journal of Photogrammetry and Remote Sensing - ISPRS J PHOTOGRAMM. 67. p.52-64. 10.1016/j.isprsjprs.2011.10.003.

Thaduri A. & Galar D. & Kumar U. (2015). Railway assets: A potential domain for big data analytics. Procedia Computer Science, 2015 INNS Conference on Big Data, Volume 53, 2015, Pages 457–467 doi: 10.1016/j.procs.2015.07.323

Tolk, A. & Turnitsa, C. & Diallo, S.Y. (2006). Ontological Implications of the Levels of Conceptual Interoperability Model. Computational Modeling and Simulation Engineering Faculty Publications. 33. https://digitalcommons.odu.edu/msve_fac_pubs/33

Tucci, G. & Betti, M. & Conti, A. & Corongiu, M. & Fiorini, L. & Matta, C. & Kovačević, C. & Borri, C. & Hollberg, C. (2019). BIM for museums: an integrated approach from the building to the collections. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W11, 1089–1096, https://doi.org/10.5194/isprs-archives-XLII-2-W11-1089-2019.

Tucci, G. & Conti, A. & Fiorini, L. & Corongiu, M. & Valdambrini, N. & Matta, C. (2019b). M-BIM: a new tool for the Galleria dell'Accademia di Firenze. DOI:10.4995/var.2019.11943. pp.40-55. In VIRTUAL ARCHAEOLOGY REVIEW - ISSN:1989-9947 vol. 10.

Tucci, G. & Corongiu, M. & Flamigni. F. & Comparini, A. & Panighini, F. & Parisi, E.I. & Arcidiaco, L. (2020). The validation process of a 3D multisource/multiresolution model for railway infrastructures. Applied Geomatics Volume 12, 1 April 2020, Pages 69-84, DOI: 10.1007/s12518-019-00286-3, ISSN: 18669298 http://hdl.handle.net/2158/1179504

TU Delft. Cities/regions around the world with open datasets. https://3d.bk.tudelft.nl/opendata/opencities/, 2019.

Tupper, C. D. (2011). Object and Object/Relational Databases. In Data Architecture, Pages 369-383, ISBN 9780123851260, https://doi.org/10.1016/B978-0-12-385126-0.00021-8

Tutcher J. (2014). Ontology-driven data integration for railway asset monitoring applications. in: Big Data 2014 IEEE International Conference on, IEEE, pp. 85–95.

Uitermark, H. & Oosterom, P. & Mars, N. & Molenaar, M. (2005). Ontology-based integration of topographic data sets. International Journal of Applied Earth Observation and Geoinformation - INT J APPL EARTH OBS GEOINF. 7. 10.1016/j.jag.2005.03.002.

Verbree, E. & Zlatanova, S. (2005). 3D-modeling with respect to boundary representation within geo-DBMS. Onderzoeksinstituut OTB. GIS report No.29. TU Delft.

Verbree, E. & Zlatanova, S. (2007). Positioning LBS to the third dimension. 10.1007/978-3-540-36728-4_8.

Visser, U. & Stuckenschmidt, H. & Schuster, G. & Vo¨gele, T. (2002). Ontologies for geographic information processing. Comput. Geosci. 28, 103–117.

Weinmann, M. & Jutzi, B. & Hinz, S. & Mallet, C. (2015). Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing, 105, 286–304.

Wendel, J., Simons, A., Nichersu, A., & Murshed, S.M. (2017). Rapid development of semantic 3D city models for urban energy analysis based on free and open data sources and software. Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics.

Yao, Z. & Nagel, C. & Kunde, F. & Hudra, G. & Willkomm, P. & Donaubauer, A. & Adolphi, T. & Kolbe, T. (2018). 3DCityDB - a 3D geodatabase solution for the management, analysis, and visualization of semantic 3D city models based on CityGML. Open Geospatial Data, Software and Standards. 3. 10.1186/s40965-018-0046-7.

Yu, Y. & Li, J. & Guan, H. & Zai, D. & Wang, C. (2014). Automated Extraction of 3D Trees from Mobile LIDAR Point Clouds, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XL-5, 629–632, https://doi.org/10.5194/isprsarchives-XL-5-629-2014.

Zhu, L. & Hyyppa, J. (2014). The Use of Airborne and Mobile Laser Scanning for Modeling Railway Environments in 3D. Remote Sens. 2014, 6, 3075-3100.

Zlatanova, S. & Prosperi, D. (2005). Large-scale 3D Data Integration. CRC Press Reference. ISBN 9780849398988 - CAT# 9898

Zlatanova, S. (2006). 3D Geometries in Spatial DBMS. In: Abdul-Rahman A., Zlatanova S., Coors V. (eds) Innovations in 3D Geo Information Systems. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-36998-1_1.

9.4 List of PhD publications

Corongiu, M. & Masiero, A. & Tucci G. (2020). Classification of railway assets in mobile mapping point clouds. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B1-2020, 219–225. DOI: 10.5194/isprs-archives-XLIII-B1-2020-219-2020

Tucci, G. & **Corongiu, M**. & Flamigni. F. & Comparini, A. & Panighini, F. & Parisi, E.I. & Arcidiaco, L. (2020). The validation process of a 3D multisource/multiresolution model for railway infrastructures. Applied Geomatics Volume 12, 1 April 2020, Pages 69-84, DOI: 10.1007/s12518-019-00286-3, ISSN: 18669298 http://hdl.handle.net/2158/1179504

Tucci, G. & Betti, M. & Conti, A. & **Corongiu, M.** & Fiorini, L. & Matta, C. & Kovačević, C. & Borri, C. & Hollberg, C. (2019). BIM for museums: an integrated approach from the building to the collections. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W11, 1089–1096, https://doi.org/10.5194/isprs-archives-XLII-2-W11-1089-2019.

Tucci, G. & Conti, A. & Fiorini, L. & **Corongiu, M.** & Valdambrini, N. & Matta, C. (2019). M-BIM: a new tool for the Galleria dell'Accademia di Firenze. DOI:10.4995/var.2019.11943. pp.40-55. In VIRTUAL ARCHAEOLOGY REVIEW - ISSN:1989-9947 vol. 10

Corongiu, M. & Tucci, G. & Santoro, E. & Kourounioti, O. (2018). Data integration of different domains in geo-information management: a railway infrastructure case study. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-4, 121-127, https://doi.org/10.5194/isprs-archives-XLII-4-121-2018

9.5 Curriculum Vitae

Manuela Corongiu was born in 1972 in Cernusco sul Naviglio (MI) Italy. In 1997 She obtained her master's degree in Civil Engineering at the University of Cagliari-Italy. Until 2006 She worked on the "IntesaGIS" project for Italian specifications on geotopographic databases. Since 2002 she had collaborated with the "LAboratory for Monitoring and Environmental Modelling for the sustainable development" project on geomatic aspects and standards (ISO TC/211 19118 and 19144-2 editing Committees). Since 2011 she has a permanent position as a researcher at the Consorzio LAMMA (a public consortium between Tuscany Regional Government and Italian National Council Research) with main tasks on spatial database design and spatial data infrastructure management.