Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Civile, Ambientale, Edile e Architettura
----------------------------------------------------------------------------------------

# Outdoor-Indoor tracking systems through geomatic techniques: data analysis for marketing and safety management

Ph.D. Dissertation of:

**Alban Gorreja**

Advisor:

**Prof. Eva Savina Malinverni**

Co-advisor:

**Ing. Roberto Pierdicca**

Curriculum supervisor:

**Prof. Francesco Fatone**

XXXV cycle

Università Politecnica delle Marche
*Dipartimento di Ingegneria Civile Edile e Architettura*
Via Brecce Bianche — 60131 - Ancona, Italy

# Abstract

In the last decades, the use of information management systems in the building data processing led to radical changes to the methods of data production, documentation and archiving. In particular, the possibilities, given by these information systems, to visualize the 3D model and to formulate queries have placed the question of the information sharing in digital format. The integration of information systems represents an efficient solution for defining smart, sustainable and resilient projects, such as conservation and restoration processes, giving the possibilities to combine heterogeneous data. Given the ever-increasing interest in data and their management, the aim of this thesis is to create an effective and clear workflow starting from geomatic surveys in a perspective of improving the collected data on the territory, surrounding buildings and those related to human behaviour so they can be better exploited and integrated into smart management models.

As first step this thesis aims to understand the limits of data interoperability and integration in GIS filed. Before that, the *data* must be collected as raw data, then processed and interpret in order to obtain *information*. At the end of this first stage, when the information is well organized and can be well understanded and used it becomes *knowledge.*

To promote the interoperability of GIS data, it is necessary at first to analyse methods of conversion in different data storage models such as CityGML and IndoorGML, defining an ontological domain. This has led to the creation of a new enriched model, based on connections among the different elements of the urban model in GIS environment, and the possibility to formulate queries based on these relations. The second step consists in collecting all data translated into a specific format that fill a graph database in a semantic web environment, while maintaining those relationships. The semantic web technology represents an efficient tool of interoperability that leaves open the possibility to import BIM data in the same graph database and to join both GIS and BIM models. The outcome will offer substantial benefits during the entire project life cycle. This methodology can also be applied to cultural heritage where information management plays a key role. (Malinverni et al. 2020)

In GIS is not trivial but it is possible to redefine and add new information to the model in order to fulfil new achievements. To demonstrate it, research and studies were conducted on historical villages to preserve and safeguard cultural heritage. The aim of the first work of this kind was to develop a SMART management system for the preservation activities of historical villages through the management of heterogeneous types of data, from the survey to the technical documentation. The workflow was structured as follows: (i) Data acquisition: the survey of a small village was carried out by combining geomatic techniques necessary to produce a complete point cloud; (ii) 3D modelling: data extrapolation from the post-processed point cloud and subsequent generation of a GIS-based on 3D model thanks to the creation of DTM and DSM of the area of interest; (iii) Knowledge modelling: a geospatial information model is necessary to put in order and together all the information collected for the whole village and each building composing it; (iv) SMART management modelling: improvement of the information management system that guarantees the possibility to enrich

and update data at any time. This research paves the way to develop a web platform where GIS data would be imported for a digital twin approach (Gorreja et al., 2021).

All the research done up to this point was to understand the capability of creating smart information models and systems in order to understand the feasibility of hosting heterogeneous data that may be included in the future.

The next step in terms of achieving this result consist of understanding human behaviour in a space. Starting from the definition of space itself, a research was conducted in order to assess the reliability of such data starting from GNSS signal towards indoor mapping and localization. Regarding mapping and positioning systems, a lot is already available in terms of commercial tools and research papers, either outdoor or indoor. But only a few research papers are addressed towards mapping and positioning systems considering both outdoor and indoor spaces. This topic, even though it has few research articles, represents a crucial aspect for many reasons, especially when it comes to safety management of damaged buildings. For instance, knowing the location of emergency team workers such as civil protection or firefighters while they operate may have a significant impact on their job. (Angelats and Navarro 2017) have been working on this aspect providing a system able to track people from outdoor to indoor areas in real-time and vice-versa. The aspects to be taken into account are many, for such reason in their publications they highlight the factors contributing to the risks of these workers, such as the lack of knowledge about their physical environment especially indoors. The use of GNSS sensors combined with Visual Inertial Odometry provide a continuous trajectory without losing the path followed by the monitored user. A part of this research focused on enhancing the final trajectory obtained with the described system above. It was done by carrying out tests on the outdoor trajectory of GNSS in order to understand from the GNSS data only the behaviour of the trajectory when it gets close to buildings or when the user moves indoors.

The last aspect this thesis will focus on is the tracking of people indoors. People live more than 80% of their lives indoors, which makes it a very interesting topic in terms of data analysis for safety, emergency purposes and marketing purposes. Human behaviour is at the centre of several studies in different fields such as scientific subjects, social and economics. Differently from the previous case study of tracking people in outdoor/indoor areas, the scope was to collect information about the dynamic indoor positioning of people based on the WiFi signal, as the use of the GPS signal alone, in today's smartphones is not sufficient for indoor geolocation. Subsequently, a brief analysis of the data will be made to demonstrate the correct functioning of the system, to emphasise the importance of data knowledge and the use that can be made of it. In literature there are various techniques of localizing users in an indoor space, but almost always must equip the person concerned with a wireless tool (such as RFIDs) or have the person interact with a specific application capable of communicating with the equipment installed in the place of interest via Bluetooth. However, the localisation of people via WiFi can take place both without the downloading of applications on smartphone and without a connection to the WiFi system. This allows the geolocation of people while not collecting sensible information on them, thus guaranteeing total anonymity and respecting privacy regulations.

This thesis is the result of a PhD co-funded by Wifiproject s.r.l., within the framework of the Marche Region “Eureka - borse di studio per dottorato di ricerca per l’innovazione” (Eureka - PhD scholarships for innovation). The aim of this project is to promote and link academic research with the regional economic scene, in order to promote innovation and strengthen the competitiveness of companies operating in the region and to help increase the employability of highly educated people.

# Contents

# List of Figures

# Chapter 1

# 1. Introduction

## 1.1. Research topic

In the data-dominated world, humans play a central role, not only in the management of data, but they represent the core data and certainly the most important ones. Data take even more meaning and importance when it comes to managing people's security. There are an infinite number of types of data that can be collected and catalogued in order to produce models and manage information in the best possible way to achieve concrete results in everyday life, such as improving conditions of well-being or increasing the level of security of the environment around us. Data is also of great importance in marketing decisions because it can guide a person to make targeted choices. This causes a strong interest in companies all around the world and brings large money shifts to this industry.

In today's world, it is hard to think of an industry that is not revolutionised by data science. In engineering, data in order to be usable must meet minimum requirements, acceptable errors, redundancy and many other factors so that the world being represented or measured is actually representative of a very complex reality.

Collected data begins as a set of “raw facts”. It is necessary to go through a process of interpretation and data structure in order to obtain what is known as information and becomes knowledge when the information is refined.

One of the aims of this thesis is to understand how to manage data and the information derived from it in order to produce knowledge in a field that has human as its main datum, both in terms of safety and for the marketing-oriented sectors. This type of combination is optimal because it covers two very different interests but which often refers to the same database.

To do this, a large amount of data will be taken into account in environments where there are large flows of people more in general in indoor places, since we spend more than 80% of our time in such spaces.

The study of the behaviour of humans, both as individuals and as a group of people, represents an essential aspect in today's research because the data that is produced and collected every day is becoming more and more available thanks to modern technology. Almost all of us use smartphones and computers to manage information both at work and in our free time. For this reason, geomatics has shifted its attention to ever newer fields and applications, thus also accepting the monitoring of people and proposing to manage the resulting information. One of the main purposes of this thesis is in fact to monitor the flow of people in large indoor environments, such as shopping centres, but this idea can be extended to airports, stadiums and other large spaces where there is a great need to manage the flow of people, depending on peak hours, extraordinary events or environmental events, as well as exceptional events related to fires and sometimes even terrorism, as we have sadly encountered in some European cities in recent years. But as already mentioned, this data can also be used and play a fundamental role for large companies that want to improve product sales through targeted advertising or simply the study of data collected on people indoors leads to a better understanding of behaviour, so more effective strategies can be reasoned out.

Before going into these aspects, it is strictly necessary to understand that generating useful data is not a trivial matter! Even more complex can be transforming data into information and then managing it in such a way that it can be considered knowledge and thus obtain results that improve our everyday life or our work. In order to be able to collect data, geomatics is the science that brings with it a great deal of experience since it includes disciplines and techniques such as topography, cartography, geodesy, photogrammetry, satellite positioning, laser scanning, computer science and more. Some of these have existed for millennia and technological progress has helped to refine each of these disciplines.

Thus, by surveying outdoor and indoor spaces, we have been able to collect information that we are able to manage in information systems. In turn, the models collected and represented in 3D can be managed by manipulating their data structure in such a way that additional information can be included to what is normally provided by a GIS (Geographic Information System) software. This type of manipulation is done in order to fill gaps in GIS, so that more and more information can be included within a single information system or several systems where necessary. Hence, the information is not only representative of the data collected, but can be managed in a "Smart" manner and with a view of laying the foundations for the future digital twin of our cities. This type of digital model requires that information is managed in real-time and that there is a strong and direct interaction between different elements of a city. Through this research, we also want to make a contribution in this sense. Using GIS systems, open standards such as CityGML, creating automatic connections between data through software such as FME (Feature Maniluplation Engine), further paving the way towards a more complex and detailed BIM model, and managing connections quickly through graph databases. This thesis aims to make a contribution to better data management by looking at a broad spectrum of possibilities on an urban scale

With this in mind and as a starting point, data was collected on people's movements, starting with outdoor areas and then indoor areas, such as in shopping centers, universities or airports. How to incorporate information about the users of a space with the space itself is still a big question mark when we talk about spatial positioning and tracking. There are many tools available to capture this kind of information, but it is still not easy to manage it on an urban scale and make it accessible.

This thesis aims to create a workflow to help those working at an urban scale to be able to manage information, both spatial and tracking, and thus produce knowledge in an easy way and able to fill the gaps that exist today in terms of data interchange without neglecting the acquisition phase and the techniques for surveying such heterogeneous data.

## 1.2. Specific research questions

The research questions this dissertation tries to address are many and range from a broader perspective to specific issues. As can be deduced in the previous paragraph, the main objectives are to manage complex information within a surveyed area that represents the real world and to make this as user-friendly as possible by automating most of the process.

Some of the questions we tried to answer may seem trivial, but they are not when it comes to finding a new way of improving what can already be done in terms of time and costs, making the use of data more accessible and how all this can be done. After this general consideration is necessary to answer those questions in terms of technical issues per each work. Although creating a digital twin model was never the aim of this work, it was helpful to interrogate about it, as it improves the general comprehension of how the research on these topics should be addressed. This means it is important to always look at the future knowing where, generically, other research papers and efforts are being made in order not only to give a contribute in terms of research but also to take advantage of what has already been done.

Below a list of question that addresses this thesis towards new approaches and to its final results.

- How can geomatics give a contribute to developing more accurate and smart models of digital twins?
- How can we improve 3D GIS models, how do we represent them and how to manage data when it comes to external information?
- Can GIS communicate with BIM in order to work with a variable scale of detail and what has been done in terms of interoperability?
- Which aspect of life can be improved by tracking people outdoor/indoor?
- Which are the limits of outdoor and indoor positioning and tracking?
- What kind of information can be collected from people indoor tracking?
- How can we use these information to improve people’s life?

## 1.3. Objectives and main research contributions

Given the ever-increasing interest in data and their management, the aim of this thesis is to create an effective and clear workflow (Figure 1) starting from geomatic surveys with a perspective of improving the collected data on the territory, surrounding buildings and those related to human behaviour (Hu et al. 2004; Moeslund, Hilton, and Krüger 2006) so they can be better exploited and integrated into smart management models. By human behaviour is meant how people move and act in relation to their surroundings and stimuli from that environment. However, it is important to define a reference scale in order to be able to understand the acceptable errors in the survey phase. Generally, we will refer to an urban scale, with information on the territory derived mainly from aerial photogrammetry, close range and laser scanner surveys. We will therefore be able to study neighbourhoods, or small villages that are of great interest because they have often been damaged by the earthquakes in recent years that have hit central Italy, up to moving inside large buildings monitoring the flow of people inside them so as to be able to manage this information as a whole and give an accurate and rapid response in the event of emergency situations. Further purposes are to manage and group the information by means of graphs and open to the possibility of studying Key Performance Indicators (KPIs) for design marketing models suited to the space.

There are many articles in the literature explaining and deepening the use of semantic web technologies (Pauwels, Zhang, and Lee 2017), especially complementary to BIM software. The reasons driving the research in this direction are related to the desire to fill the gaps in terms of interoperability between software in the domain of AEC (Architecture, Building and Construction) then exploit the logical foundations of these technologies and connect valuable information where it would not be possible. This thesis addresses this issue in chapter 2.5. by attempting to provide a rapid response both in terms of realising interoperability between software and in terms of utilising these processes, thus enriching the scientific literature with an aspect of interoperability in which information is remodelled in order to create connections between objects and spatial elements using GML formats and the now well-known RDF graphs. One of the objectives is to make this process as simple as possible and the input of new data as automatically as possible (Chapter 2.6.).

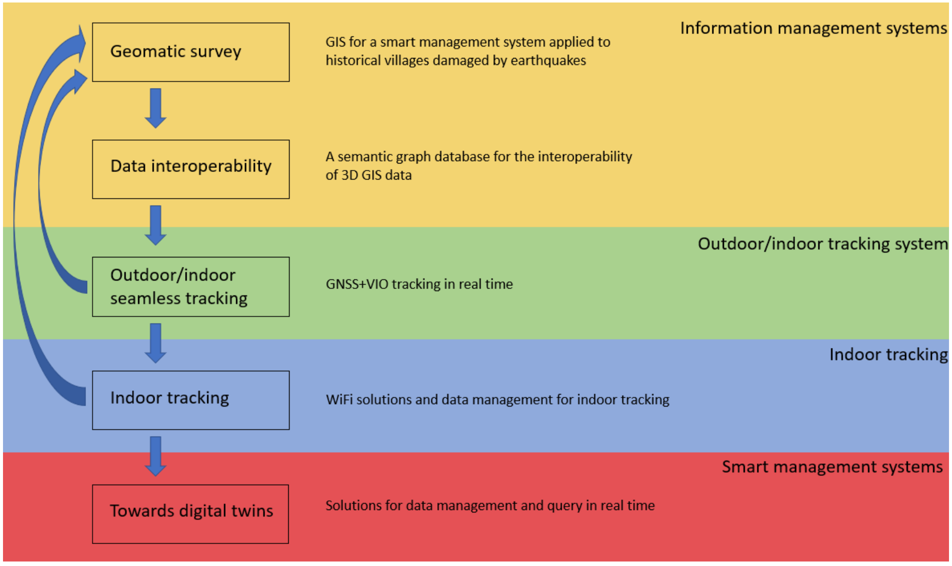


*Figure 1 Workflow and methodology proposed by this thesis*

This flow leads to the questioning of how spatial information is related to data of a new and different nature. Once the space has been defined, both in terms of urban scale and individual buildings, it has been identified that data related to human beings is of great interest. Data relating to the number of people, i.e. density with respect to an area of interest and human behaviour represent a further challenge, because they must be collected with such accuracy that they can be related to the reference scale of the building model. A further barrier encountered in the acquisition of data with regard to geomatic techniques to be addressed is that of data relating to people's trajectories when moving from an outdoor to an indoor space and vice versa. To these considerations, this thesis proposes to provide innovative solutions compared with those in literature and to improve, in terms of output, research works already conducted by other authors (Chapter 3). Again with regard to data collection, this time in the indoor sphere, the scientific literature presents many solutions that can be applied to buildings such as those we will see in this thesis, but it is almost always necessary to equip users with tools capable of interacting in an "active" manner with a predisposed infrastructure, such as applications for smartphones that can be intercepted via Bluetooth or cameras combined with artificial intelligence software for recognising people, thus capable of reconstructing a trajectory for each user in such a way that users do not have to interact directly with a technological infrastructure but can be monitored in a "passive" manner. But these types of solutions may not be feasible in a real-world context because it’s not possible to ask all users in a space (e.g. airport or shopping mall) to download a special application in order to interact with them, or it may be too expensive in terms of passive path recognition. The solution proposed here is a passive tracking model via WiFi installed in the indoor space capable of detecting users without asking the user to interact with a smartphone application in total respect of privacy.

In particular, some of the data collected and analysed in this thesis refer to the beginning of the COVID-19 pandemic. It was interesting to observe in such a specific case the trend of attendance in shopping centres before and during the pandemic. Many analyses have produced great knowledge in this specific period. Still, it is not easy to understand how to make the best use of this information, especially in cases where the collected data cannot be managed in real-time. A real-time approach is undoubtedly a front that needs a lot of investigation so that this information yields effective and immediately usable results. Unfortunately, due to certain technological and economic limitations, the processing of data collected in indoor environments has a delay of several hours. This allows us to make considerations and studies on those data, but not to make real-time decisions. On the other hand, the approach used in Chapter 3 concerning tracking and positioning passing in outdoor/indoor areas must, by its very nature, be able to allow real-time tracking given the dangerous situation in which the operators to whom this device is addressed are subjected.

As a concluding remark, it is worthwhile to note that the proposed methodology fulfils with the European SENDAI framework for disaster risk reduction (Nations Office for Disaster Risk Reduction 2020) in at least two of four basic principles: (i) Understanding disaster risk and (iv) enhancing disaster preparedness for effective response, and to "Build Back Better" in recovery, rehabilitation and reconstruction.

This is not trivial, and the introduction of Geomatic survey combined with human behaviour information for increasing the knowledge represents a fundamental step forwards.

## 1.4. Structure of the thesis

Following the introduction of Chapter 1 this thesis is developed as the described structure below:

*Chapter 2. The importance of Information management systems*: will focus on the role of survey, GIS and data management. The survey for context awareness and digital environment development are the first two steps which will lead to a complete 3D GIS model. At this point, the role of GIS and 3D GIS will be introduced in order to understand if and when interoperability is necessary in the process of data management or information management. For such interoperability and data integration, this thesis will show what in terms of ontology should be modified and how software can be "malleable or ductile" whether they easily accept external information or they don't and how to get over such problems. Among the others, two main works will be discussed:

- A semantic graph database for the interoperability of 3D GIS data
- 3D GIS for a smart management system applied to historical villages damaged by earthquakes

The results of these two works represent the first phase of the proposed workflow of this thesis.

*Chapter 3.1. Outdoor indoor tracking system*: GNSS comprehension and use from previous surveys led us to think about the GNSS signal carried on by people on their phones and how this simple information can be used in terms of people's safety. It came out that was a lack of research activity from outdoor to indoor people tracking using GNSS signal combined with other solutions. A research centre in Barcelona, the Center of Technology and Telecommunication of Catalunya (CTTC) was the only one who carried out such kind of research activity. In collaboration with the team which developed this technical approach of combining GNSS and Visual Inertial Odometry data, efforts were made in order to improve the final trajectory obtained by tracking people in emergency scenarios.

*Chapter 3.2. Indoor tracking*: After moving from outdoor into indoor areas, also major importance has to be given to people who occupy these areas, especially when it is about high-density areas. This consideration, which are made in order to better understand the risks and the behaviours of people indoor led to study approaches where was no need of interacting directly with the user of an indoor area. The idea was to collect information without providing users with an additional tool or asking them to interact with any device but to collect data only by the WiFi signal of their devices, even if not connected to the WiFi network. In this chapter will be discussed briefly the technology and the data analysed in order to make observations of the general behaviour.

*Chapter 4. Discussions*: In this chapter, the questions posed in chapter 1.1. are addressed giving not only the solutions proposed on this thesis but also clarifying the limits due to technological development.

*Chapter 5. Concluding Remarks*: The thesis concludes with a summary of the results and proposes what future developments this work will lead to.

# Chapter 2

# 2. The importance of Information management systems

As discussed in the previous chapter, the first part of the work was to give a structured basis, which is intended as a digital hub or platform, where will be placed or linked new data. The first part of the process has been done through geomatic techniques, starting from the survey and data management in GIS and further expanded by considering the current limits of the software used. After a deep literature research necessary to understand what needs to be overcome in order to allow such data integration and data interoperability, a few works were carried out and will be discussed in this chapter, analysing since the beginning what led us to investigate those aspects and which solution this thesis proposes. Before going into these aspects it is mandatory to give a rapid overview on some features and tools which will be recalled in later in this chapter, necessary to understand the works carried out. Those are: semantic web, CityGML, IndoorGML, 3D GIS and graph databases.

## 2.1. Semantic web

The main obstacle of the integration between different information systems is the lack of interoperability across domains, which often in literature is solved using the semantic web. The semantic web is a set of technologies used for the representation, publication and browsing of structural data on online platforms. It is used in chapter 2.6 to convey meaning, which is interpretable by construction project stakeholders as well as BIM and GIS applications processing the transferred data (Karan and Irizarry 2015).

The main elements belonging to a semantic web platform (Hor 2015) are as follows:

- Uniform Resource Identifiers (URIs), a string of characters that identifies a particular resource;
- Web Ontology Language (OWL), a type of knowledge representation languages form authoring ontologies for representing the conceptual schema;
- Resource Description Framework (RDF), a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata model for defining the data according to the schema;
- SPARQL, a SQL-type language for carrying out queries in data.

RDF, RDFS (Resource Description Framework Schema) and OWL are languages with clearly defined semantics or mathematical basis for the meaning of each construct. Since concepts in RDFS and OWL ontologies are expressed formally, they can be processed by computer programs.

The ontology is the key element of the semantic web. It classifies objects, data, etc. (entities), with their attributes and the relationships among them inside a domain of knowledge. In other words, it can be defined as a data structure that represents a model of semantic representation of reality. The shared language used to describe the semantics of the data is the uniform way to facilitate the communication among different users to understand each other (Taye 2010). Ontologies are used to overcome the barriers to heterogeneous semantic data sharing. They are commonly used for many purposes such as network management, data exchange on the World Wide Web and information retrieval (Hor 2015). There are several examples in literature where ontology is required before the data conversion (El-Diraby, Lima, and Feis 2005) (Hor, Jadidi, and Sohn 2016). Also, it remarks the importance to maintain the ontology of a model in order not to lose the meaning of each feature during the data format conversion.

As is described in (Gröger et al. 2012), main elements of RDF are triples composed by three elements: subject, predicate and object. They can be represented in a graph (upper of Figure 2) with three linked nodes.

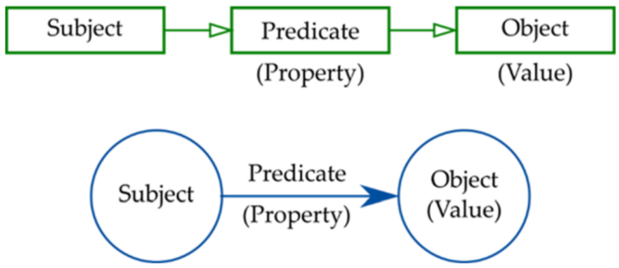



*Figure 2 RDF triples*

The meaning of predicate could be interpreted as property and the meaning of object as value. The final interpretation of the triple will depend on the data stored, e.g., if the predicate is “hasChild”, the meaning of the object could be a value if a boolean data (yes or no) is stored, or another object if all data regarding the child is linked. Other manner to create this RDF graph is using two nodes (subject and object) connected by an edge (predicate) (bottom of Figure 2). The latter structure is being used in chapter 2.6 to maintain CityGML ontology and to keep semantic web techniques. Semantic interoperability is crucial for its ability to exchange information between existing standards. It is also capable to automatically interpret the exchanged information in meaningful and accurate way. This process allows users to combine 3D models and get information from them. To do so, these models must converge to a common information domain. In this way, the exchanged information using semantic interoperability is unambiguously defined. This means that the sent data are the same as what is understood (Wiki.GIS 2019).

## 2.2. CityGML

Nowadays there is a variety of international standards for each field of application. Open Geospatial Consortium (OGC 2019) is an international organization which develops and maintains open standards (more than 60 have been published). Among all them available, the most extended and used in city modelling is CityGML, issued also by the ISO/TC 211 regulation (ISO/TC 211 2019).

CityGML is a common semantic information model for the representation of 3D urban objects that can be shared over different applications. It is an open source data model and eXtendible Markup Language (XML) – based format for the creation and the exchange of city models. The geometry is stored using the Geography Markup Language version 3.1.1 (GML3), which is also XML-based format and is usually employed in geographical information archive. Furthermore, CityGML enables lossless information exchange between GIS software and users. It defines classes and relations regarding their geometrical, topological, semantical and appearance properties (Gröger et al. 2012). It is applicable for large areas and small regions and can represent the terrain and 3D objects in different levels of detail (LOD) simultaneously. For example, simple models without topology and few semantics in one LOD, instead of very complex detailed models with topology and fully semantical, can be represented in different LODs. LODs in CityGML indicate the accuracy of geometries and the potential elements that are included in the model. LODs range from 0 to 4. Let us use building objects for illustrating them: LOD0 is the coarsest model, and it is mainly a 2D model with a DTM; e.g., building would be represented as a 2D polygon laying on the DTM. LOD1 includes buildings with its height. LOD2 defines the structure of roofs and building installations. LOD3 represents the real geometry with accuracy, and LOD4 is the realist model in which all details of the building are modelled. CityGML specification (Gröger et al. 2012) defines for each object model the information needed for each LOD.

## 2.2. IndoorGML

IndoorGML (Li et al. 2019) is a OGC standard, XMLbased, and it is implemented for defining routes inside buildings, linking different rooms through doors and so, navigable spaces. Among all OGC standards, it is the only one that establishes an ontology for navigable routes and has a possible connection with CityGML.

An important difference of indoor space from outdoor is that an indoor space is composed of complicated constraints such as corridors, doors, stairs, elevators, etc., like a road network space is composed of road constraints. It means that proper representations of indoor constraints are key issues of indoor spatial information modelling and standards (Ki-Joune Li 2022).

- Cellular space: indoor space as a set of cells, which are defined as the smallest organizational or structural unit of indoor space [Wordnet, Princeton University, 2010]. Cellular space has important properties. First, every cell has an identifier (namely c.ID) such room number. Second, each cell may have common boundary with others but does not overlap with other cells. Third, postion in cellular space can be specified by cell identifier, although we may employ (x, y, z) coordinates to specify a position for more precise location.

- Semantic representation: Semantic is an important characteristic of cells. In IndoorGML, semantics is used for two purposes: to provide classification and to identify a cell and determines the connectivity between cells. Semantics allows to define cells which can be of importance for navigation. For example, the most commonly used classification of cells in topolographic space is into navigable (rooms, corridors, doors) and non-navigable (walls, obstacles) cells.

- Geometric representation: The geometric representation of 2D or 3D feature in indoor space does not belong to the major focus of IndoorGML, since they are clearly defined by ISO 19107, CityGML, and IFC. However, for the sake of self-completeness, the geometry of 2D or 3D object may be optionally defined within IndoorGML according the data model defined by ISO 19107. There are three options to represent the geometry of a cell in IndoorGML as Figure 3.

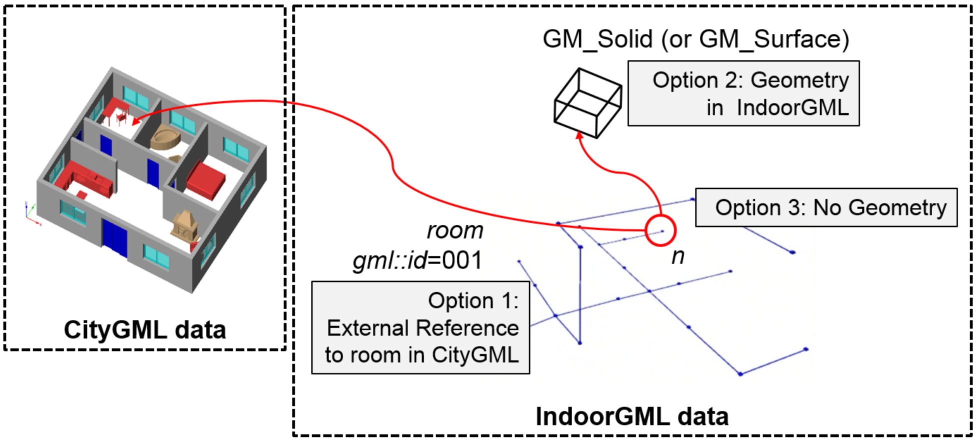



*Figure 3 Geometry in IndoorGML*

External Reference (Option 1): Instead of explicit representation of geometry in IndoorGML, an IndoorGML document only contains external links (namely c.xlink, where c is a cell in IndoorGML) to objects defined in other data sets such as CityGML, where the referenced objects in external data set include geometric information. Then there must be 1:1 or n:1 mappings from cells in IndoorGML to corresponding objects in other dataset.

Geometry in IndoorGML (Option 2): Geometric representation of cell (namely c.geom, where c is a cell in IndoorGML) may be included within an IndoorGML document. It is GM_Solid in 3D space and GM_Surface in 2D space as defined in ISO 19107. Note that solid with holes or surface with holes are allowed in this standard.

No Geometry (Option 3): No geometric information is included in IndoorGML document.

- Topological representation: Topology is an essential component of cellular space and IndoorGML.The Node-Relation Graph (NRG) (Lee 2004) represents topological relationships, e.g., adjacency and connectivity, among indoor objects. The NRG allows abstracting, simplifying, and representing topological relationships among 3D spaces in indoor environments, such as rooms within a building. It can be implemented as a graph representing the adjacency, connectivity relationships without geometrical properties. It enables the efficient implementation of complex computational problems within indoor navigation and routing systems. The Poincaré duality [JR Munkres 1984] provides a theoretical background for mapping indoor space to NRG representing topological relationships. A given indoor space can be transformed into a NRG in topology space using the Poincaré duality. It simplifies the complex spatial relationships between 3D by a combinatorial (or logical) topological network model. According to Poincaré duality, a k-dimensional object in N-dimensional primal space is mapped to (N-k) dimensional object in dual space. Thus solid 3D objects in 3D primal space, e.g., rooms within a building, are mapped to nodes (0D object) in dual space. 2D surfaces shared by two solid objects is transformed into an edge (1D) linking two nodes in dual space.

- Multi-Layered Representation: A single indoor space is often semantically interpreted into different cellular spaces. For example, an indoor space is represented as a topographic cellular space composed of rooms, corridors, and stairs, while it is also represented as different cellular spaces with WiFi coverage cells and RFID sensor coverage cells. For this reason, IndoorGML supports multiple representation layers with different cellular spaces for an indoor space. Each semantic interpretation layer results in a different decomposition of the same indoor space, where each decomposition forms a separate layer of cellular space.

## 2.3. 3D GIS

Geographic information systems (GIS) are a type of database that collects geographic information. GIS software allows to view, analyze and manage the information contained in the database. GIS are utilized in multiple technologies, processes, techniques and methods.

They are attached to various operations and numerous applications, that relate to: engineering, planning, management, transport/logistics, insurance, telecommunications, and business. For this reason, GIS and location intelligence applications are at the foundation of location-enabled services, which rely on geographic analysis and visualization.

Three-dimensional GIS data incorporates and extra dimension (a z-value) into its definition (x,y,z). Z-values have units of measurement and allow the storage and display of more information than traditional 2D GIS data (x,y). Even though z-values are most often real-world elevation values. Such as the height above sea level or geological depth. There is no rule that enforces this methodology. Z-values can be used to represent many things, such as chemical concentrations, the suitability of a location, or even purely representative values for hierarchies.

There are two basic types of 3D GIS data: feature data and surface data. Feature data represents discrete objects, and the 3D information for each object is stored in the feature's geometry while surface data represents height values over an area, and the 3D information for each location within that area can be either stored as cell values or deduced from a triangulated network of 3D faces. Surface data is sometimes referred to as 2.5D data because it supports only a single z-value for each x,y location. For example, the height above sealevel for the surface of the earth will only return a single(Wikipedia contributors. 2022).

## 2.4. Graph database

A graph database (GDB) is a database that uses graph structures for semantic queries with nodes, edges, and properties to represent and store data. A key concept of the system is the graph (or edge or relationship). The graph relates the data items in the store to a collection of nodes and edges, the edges representing the relationships between the nodes. The relationships allow data in the store to be linked together directly and, in many cases, retrieved with one operation. Graph databases hold the relationships between data as a priority. Querying relationships is fast because they are perpetually stored in the database. Relationships can be intuitively visualized using graph databases, making them useful for heavily inter-connected data (Angles and Gutierrez 2008) (Wikipedia contributors 2022a).

## 2.5. 3D GIS for a smart management system applied to historical villages damaged by earthquake

In recent years the use of three-dimensional (3D) modelling software and open-source platforms for SMART data management of urban and building projects has become increasingly frequent. The constantly updated regulations suggest and, in some cases, enforce the adoption of new sustainable, advanced and more accessible digital information systems. (Nam and Pardo 2011) . This SMART approach is not only intended to represent optimal management of the data of one or more projects after a seismic event, but it also serves to define preventive measures (real-time data visualisation, disasters prediction, etc) that can be taken to deal with this type of emergency.

As a case study for this research project, one of the many villages damaged by the seismic events that hit central Italy in 2016 is chosen. Villages that represent a full-fledged historical and cultural heritage. Among these, there is a small village called "Gabbiano", located in the municipality of Pieve Torina, in the Marche region in Italy. It is characterized by a historic centre with a limited number of buildings that have suffered damage as a result of the earthquake that has affected this area.

The aim of this research activity is to create a robust information system for the management of various types of data, from the survey to the technical documentation up to the following phases of the project through an information system such as GIS. The first step is to create a 3D model obtained through a combination of geomatic techniques. The entities composing the 3D geometric model become the object features of an organized knowledge modelling in order to have a precise connection with geometries, defined by a semantic representation following the CityGML schema. This leads to a SMART approach to the information management system that collects not only spatial data but also topological and not-geometrical data. The enriched knowledge modelling linked to the 3D object may represent a starting point for the digital twin approach, useful to manage the monitoring and restoration activities (Ketzler et al. 2020). This solution can be replicated in other similar case studies to safeguard these small villages that risk losing their historical, architectural, urban and social identity as a result of these accidental events.

The GIS platform is well suited for data management at the urban scale in a given geographical area. Moreover, 3D GIS software gives a three-dimensional model of the topographic representation of the terrain with the geolocalization of the buildings in the form of volumes at the various LODs (Level of Details) (Almeida et al. 2016). It incorporates 3D territorial and buildings information giving a stronger, richer and clearer visual impact compared with a simple two dimensional GIS mapping. A 3D GIS can be used as a reference 3D map for any urban planning management activities and among them also in case of seismic emergency (Lenticchia and Coïsson 2017). In particular, GIS is suggested as a suitable tool for the risk mapping or better the spatial extent of risk in which document the effects of hazard situations, according to the ISO/IEC 31010:2009 and the for National Risk Assessment (Commission et al. 2019). An example of risk assessment in GIS was implemented for the data management of expected seismic scenario analysing a building complex in Lisbona (Ilic, Bento, and Cattari 2020). Other case studies to assess and manage the seismic damages and vulnerability in GIS were adopted for Scarperia (Cardinali et al. 2020) or to monitor the recovery plan after earthquake in L'Aquila (Contreras et al. 2016).

It's possible to convert the 3D GIS modelling in an open data system thanks to the CityGML standard (OGC, 2020). CityGML is a common semantic information model for the representation of 3D urban objects that can be shared over different applications. It is an open-source data model and eXtendible Markup Language (XML)- based format for the creation of city models(Gröger et al. 2012). CityGML is widely used in GIS city modelling and represents an open recognized standard and well-defined ontology. The hierarchical structure of data is provided by CityGML that thanks to its schema allows having a semantic representation that defines relationships between the various entities composing the 3D city model(Malinverni et al. 2020). CityGML-based approach that leads to defining a knowledge modelling where not only spatial but also external information, like topological, architectonical, technical data, is managed (Agugiaro 2016). The benefits of using a centralized GIS-based information system, developed in CityGML environment, may guarantee an efficient, well-structured and SMART city management system (Banerjee, Chakraborty, and Das 2020). This open-source GIS tool allows to map the damages and support the emergency planning for risk and hazard scenario like an earthquake (Pollino et al. 2012) (Francini et al. 2018).

## 2.5.1. Methodology

Following the data processing of the data acquired through the geomatic survey, which allowed the creation of mesh of the terrain, the volumes of the buildings in a 3D GIS environment can be generated. In parallel to the geometric modelling is defined a CityGML schema identifying the entities, such as roads, buildings and their features, composing the 3D model of the village and their relationships. This knowledge model will represent the starting point for a SMART management modelling of the entire village promoting advanced data management ensuring access by different figures as project experts and private clients who can constantly update with new data at any time.

The workflow shown in Figure 4 will be described below.

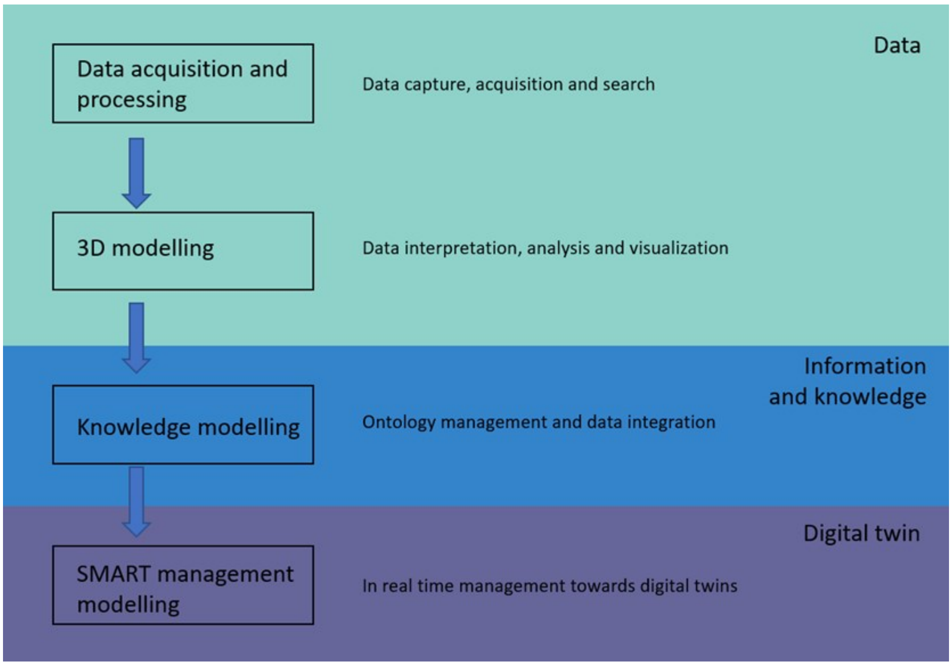



*Figure 4 Research workflow*

The survey of the small village "Gabbiano" was carried out by integrating three geomatic techniques: GNSS (Global Navigation Satellite System), terrestrial LIDAR (Light Detection And Ranging) and aerial photogrammetry from UAV (Unmanned Aerial Vehicle). The aim was to obtain the most comprehensive point cloud possible from which to create a complete 3D model. In fact, the terrestrial LIDAR is very useful to detect building facades facing the narrow alleys of the village, the aerial photogrammetry to complete the survey in detecting roofs and the GNSS to collect the coordinates of specific points through which georeferencing the whole generated point cloud. The terrestrial LIDAR scans registration process has led to a point cloud with a total network RMSE (Root Mean Square Error) of 6.82 cm, while using 10 GCP (Ground Control Points) for the photogrammetric processing of aerial images, a reconstruction with RMSE of 4.82 cm was achieved. These two have been merged into one unified point cloud (Figure 5).

*Figure 5 Unified lidar and photogrammetric point cloud*

The availability of a numerical mapping at 1:10.000 scale by the Marche region, where the equidistance between contour lines is 10 meters, couldn't give a detailed 3D model of the small village. Therefore, the absence of an open-data platform that provides GIS data has led us to obtain the 3D model through the generation of mesh from the processed point cloud. The extrapolation of data from point is useful for 3D modelling and can be conducted through an open-source 3D model management software. In this context CloudCompare (2021) was chosen, in order to ensure that the required data, such as the buildings base plan and their height above ground, are precisely calculated, taking into account the non-constant morphological distribution of the terrain. The "Rasterize" command allows to generate DTM (Digital Terrain Model) and DSM

(Digital Surface Model) raster images. The latter is obtained using the entire point cloud and considering the maximum height of the points in each cell of the grid. For the former instead only the ground points must be used, which can be extracted thanks to the Cloth Simulation Filter (CSF) (Zhang et al. 2016), considering the average height of the points for each cell. In the empty cells, the height is interpolated. Importing the DTM in raster format in GIS it is easy to reproduce the TIN (Triangulated Irregular Network) model of the terrain on which the volumes of buildings are placed. In order to obtain a correct map of the buildings, it has been drawn in CAD software using the point cloud as reference. Then we imported it as a vector layer of polygons in GIS to enrich the data in its attribute table. The mean height values from DTM and DSM on the area of each building is extracted thanks to the "Zonal Statistic" function, while the difference between them can be easily obtained using the field calculator. The latter constitute the extrusion height to use on each polygon to build up the 3D entities with good volumetric representation. In this case LOD 1 (Level of Detail

- OGC, 2020) has been reached (Figure 6). This level will increase once other useful information will be added to enrich the 3D modelling.

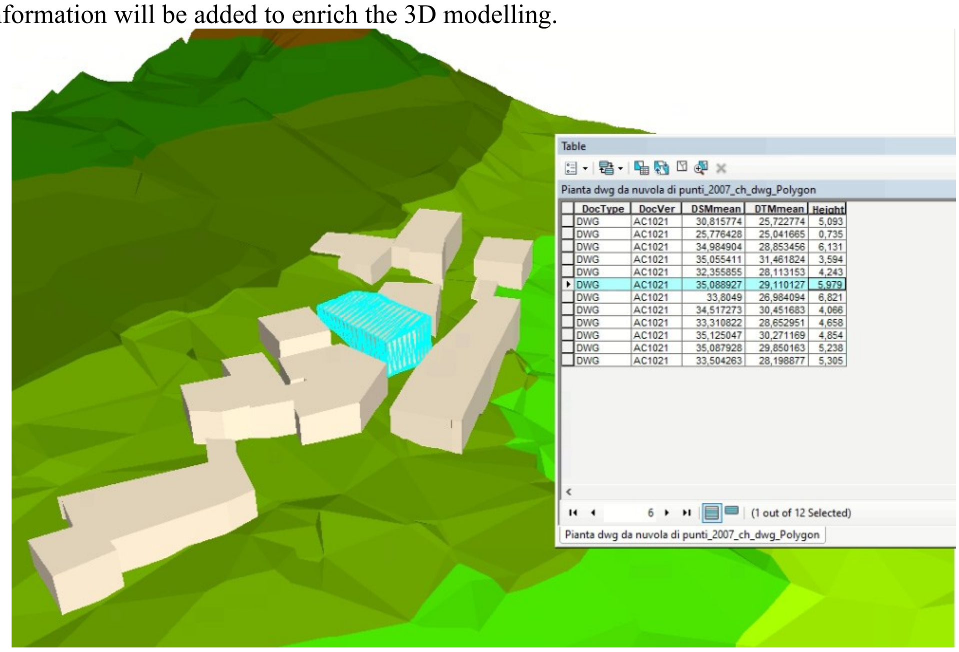



*Figure 6 3D GIS model at LOD 1*

## 2.5.2. Knowledge modelling

The 2016 seismic events made the analysis of the current state of the small village a priority for an adequate and rapid preservation project that would also takes into account the safety works carried out in the aftermath of the earthquake. For this study, the possibility to collect all these data into an information system enriching a 3D model represents a fundamental knowledge step for planning forthcoming interventions. An example of a detailed description of the damage state of these buildings in Italy is outlined on the technical data sheets called AeDES, whose acronym stands for "building operability and damage in post-earthquake emergency" (Baggio et al. 2007), recorded by the Italian Civil Protection. Through GIS it will be possible to create a link to these documents for each building and try to describe the state of damage by enriching the 3D model with thematic mappings of the decay (Malinverni, Giuliano, and Mariano 2018);(Tsilimantou et al. 2020). Once collected the AeDES data sheets, the operation adopted is to store such files in a webserver and link each file with its building in the model. Using any GIS software, it can be done manually by adding to the table of attributes of the project a new column containing the reference path to where the information is stored. For large and complex models, it's not handy doing this for every building manually, an automatic process is suggested.

Since the 3D GIS model was converted into a CityGML model, many modifications, such as further data enrichment or even changes to the elements as well as the topography, can be done. In this step, corresponding between GIS and CityGML is possible to automatize the process of enriching information, such as AeDES sheets. A geospatial information model has been created with the aim to reach a common 3D city model allowing the reuse of data in other application fields of city management. These features make CityGML the best choice to follow. A new 3D model means a new ontological schema, necessary to put in order and together all the information collected for the entire village. In short, each building is represented by a _CityObject of the village. Information related to each building are linked to the _CityObject with other features (subclass) thanks to the CityGML schema. The externalReference, already provided by CityGML, allows to link other types of data to the project object, for example technical documents (Figure 7)

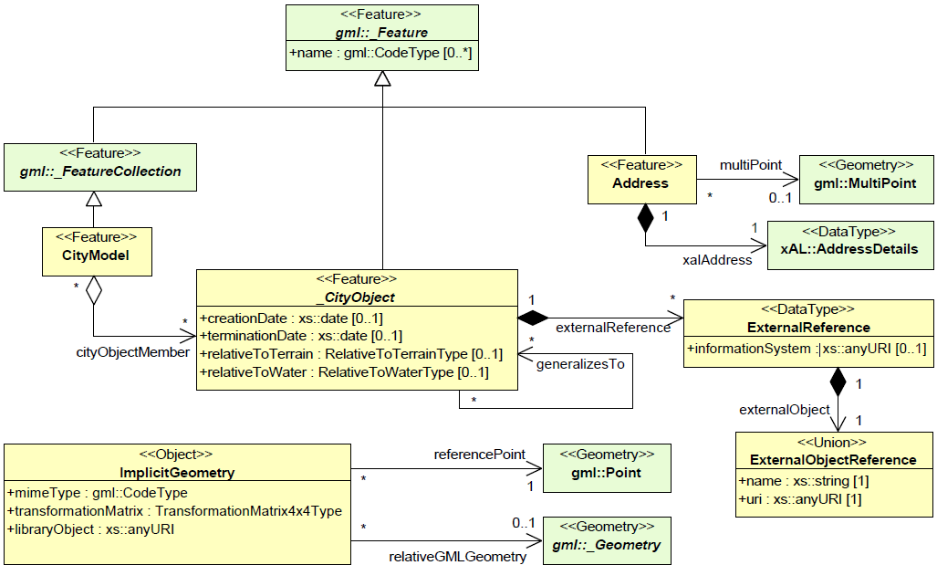



*Figure 7 CityGML schema*

As just outlined, it is necessary to create a relationship between the 3D city object and external documents, but externalReference allows to link these data to the whole 3D city model, not for each single object composing the city model. A solution that can be adopted consists of using another element of the CityGML schema. The Address can be filled with different data for each building. Through FME (Feature Manipulation Engine) software it is possible to create a workflow of transformation tools that can modify the structure of the model itself. FME is a data integration and transformation platform. Leaving an empty column of an attribute within the GIS model, it is possible to fill it with information through FME (2019). It can be done automatically if the data are previously stored in XML format. FME allows redirecting the information stored in the XML to each building using the Address (Figure 8).

This means that the software will get a hypothetic address, write it in the empty attribute previously created. At this point, the attributes of each building in the CityGML model will show a link to the webserver (or local host, depending on where the data is stored) (Figure 9). This ontological approach represents a valuable solution to link information to the 3D GIS model.

### 2.5.3. SMART management modelling

A parametric navigable "virtual" model is very different from the previous systems where the information was limited to something graphic. Today, the advent of new technologies makes it possible to create information systems where information can not only be visualised in real-time in a navigable 3D environment, but can also be entered in real-time, processed and read by a user. In addition, given the complexity of the territory and cities, the information related to them may come from different sources and refer to different fields of application. A system designed in this way is a structure that represents the dual reality for which it is called a "digital twin". One of the aims of this research is to address this challenge by providing features and data from a GIS information system that will form part of, or better still define the basis for a digital twin model. Therefore, a working process as explained in the previous paragraph can help to move towards a detailed representation of reality. Generating a workflow where information is updated in a very short time can be a further development of the information management of an urban area. It may also save time and resources, both in the research project and for figures working in the AEC (Architecture, Engineering and Construction) field and public institutions. The involvement of several figures within the planning of a recovery project such as the reconstruction of a village hit by an earthquake also allows common and optimal choices and decisions to be taken in a short time thanks to this open data platform(Artese and Achilli 2019). These innovative aspects that allow an improvement of data management through an information system like GIS give it a development of the knowledge modelling towards SMART management modelling.

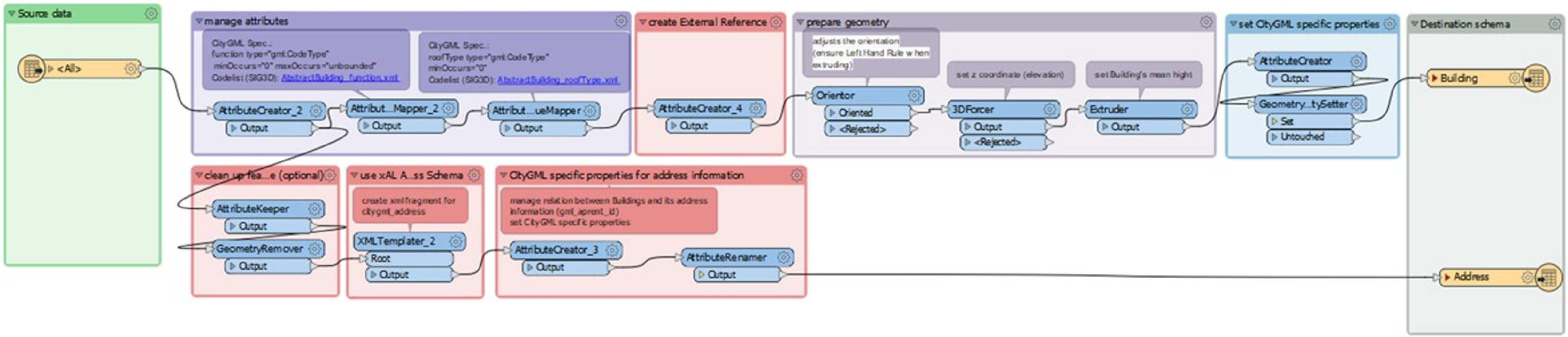



*Figure 8 Workflow showing address and external reference data enrichment*

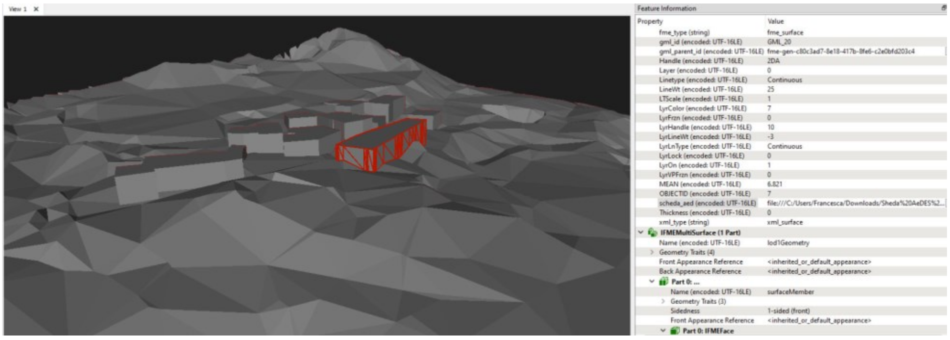


*Figure 9 3D CityGML: model and related attributes to the selected building*

## 2.6. A semantic graph database for the interoperability of 3D GIS data

The use of information technologies in the building data processing led to radical changes to the methods of production, management and archiving of data documentation. In particular, 3D model visualization and query formulation are offered by many information systems at the building and terrain scale for data access (graphics, photos, technical documents, regulations, etc.) linked to the object. Information systems are based on the digital representation with an informative 3D model linked to a relational database, which is an archive where all data of the object, in different formats, can be collected together under the entity relationship model. Therefore, they represent useful tools for project management. In recent years, the research in AEC sector did not focus on the use of a single information modelling for the data management, but it tried to make an integration of different information systems to create a multi-disciplinary, multi-scale, multi-scope, multi-user approach. In order to achieve a model technically richer and more complete of the object, it is necessary to share data among the information systems used and so, where it needs, to proceed to the data conversion or transformation in a digital exchange formats to promote the concept of interoperability. The development of this more advanced environment, supported by the interconnection, the communication, the data transferring and enrichment among the information systems, requires the creation of a common platform. For example, using the web network, where all data from different sources are collected and connected in a unique graph.

In AEC field of work, among the information systems and technologies, GIS is a useful tool for building and terrain data management. As already discussed, the integration of GIS with data coming from other information modelling environment represents an efficient way for defining a smarter, more sustainable and resilient project. It gives the possibilities to combine heterogeneous data: geometric shapes, quantitative analysis, enrichment of semantic knowledge, application of different technologies and multi-scale management (Ma and Ren 2017)(Fosu et al. 2015)(Yamamura, Fan, and Suzuki 2017). This innovative integration between information systems becomes suitable to many building fields of application. Some representative examples are the design and construction stages, the management of the construction sites, the site layout planning and the location of temporary objects (Sebt, Parvaresh Karan, and Delavar 2008). GIS integration with other technologies permits to define, follow and control each step of the building or infrastructure project and its effects on the territory or the elements connected to such as installations and services. Moreover, GIS provides to be an additional support system for the historical building or sites management, offering a new way of co-working for the preservation, conservation, monitoring and restoration activities of cultural heritage. GIS data can be linked with high detailed information models, like those created through Building Information Modelling (BIM) software. For this reason, the integration of GIS and Heritage (or Historical) Building Information Modelling (HBIM) (Murphy, Mcgovern, and Pavia 2009) can be considered as an adequate solution for managing the information for conservation and restoration projects of existing building at architectural scale (Malinverni et al. 2018) (Malinverni et al. 2019)(Matrone et al. 2019).

Although GIS was originally used to manage geospatial data in 2D scale, it provides a robust data storage system (Vacca et al. 2018) based on a sort of hierarchy of classes and subclasses identified by the levels of detail of CityGML schema. The latter permits the definition of topological and semantic relationships between the objects. By the development of 3D GIS it makes possible the creation of three-dimensional geospatial modelling, allowing the data management of specific building, offering a precise visualization of the geographical contextualisation and permitting the formulation of spatial queries (Rinaudo, Agosto, and Ardissone 2007). So, the 3D building model can be raised on a modelled terrain in a relative urban context with its surroundings. While CityGML describes and classifies object in the urban scale, IndoorGML focuses on the connections between the rooms of a building helping to create navigable routes, creating a network. In this research we used IndoorGML to make connection between buildings and outdoor elements such as roads. After this preliminary overview, it is clear that the attention focuses on the in-depth analysis about the management of GIS data. In particular, this work illustrates a research topic based on the conversion of 3D GIS data into an exchange format that can be readable and interpretable by other information systems, such as BIM or HBIM. In this research we tried to find out a possible solution to convert 3D GIS data, giving those data a CityGML ontology and, after a process of elaboration, importing the transformed data in a semantic web database based on RDF (Resource Description Framework) graph. The final result of the work is to identify the benefit in using the semantic web platform to collect GIS data, translated in a unique common exchange format, based on RDF connecting triples(Hor et al. 2018). Then, in the same web environment, it is possible to add and gather other converted data coming from other information systems. This semantic interoperability needs to be provided by a domain for the description and the individuation of the objects (Karan and Irizarry 2015). In computer science, ontologies are adopted as domains providing a formal structure for sharing and managing data defining objects (taxonomies) and their relationships (El-Diraby et al. 2005).

Keeping a standardized ontology can play a key role in a project development (such as reconstruction or building restoration projects). It helps the designers to have a quick overview of infrastructure affected by any work decisions. Furthermore, they allow linking information between many models created under different standards (GIS, BIM, HBIM) to make more complete decisions. This work was structured as follows: the first part is dedicated to the state-of-the-art of the concept of interoperability between information systems ("State-of-the-art"); then after a small bracket about the semantic web, the methodology workflow is described ("Methodology"), defining so the standards and the tools, used for 3D GIS data conversion, analysing their characteristics, properties and functions ("Standardization").

Hence, the process of translation from CityGML to RDF-graph will be outlined ("Implementation of graph database"), and finally, some considerations and results will be discussed

The data integration process, as said before, between information systems represents an innovative approach offering substantial benefits, and this combination has to take into account the strong points from each system (Zhang et al. 2009) Considering for example the most used systems like BIM and GIS, their integration represents an efficient tool for AEC projects. In short, BIM describes geometry, semantic relationships and identifies the building components. GIS provides a well-structured database (level of detail) and a geospatial model with topological and semantic relationships. But there are some dissimilarities between them, such as spatial scale, level of representation of geometric models and structure of database.

By literature, the topic of translation of data between GIS and other information systems, BIM in particular, has been already dealt with, as it can be red in some examples below, and it is still an ongoing research, but it led to a specific solution, not simple to understand or to propose again for different case studies. Especially, it is to be highlighted that some software producers started a collaboration trying to get easier and to overcome the complexity of the data conversion and sharing between information systems, so it is already possible to import BIM data into a GIS environment. Although this has shown good performance, it must be said that this data conversion has a limitation because it only proceeds in unidirectional way, from BIM to GIS. This is not the aim of this research work.

Analysing research activities focusing on the topic of integrated information system, the interoperability between BIM and GIS can be presented in different ways: syntactic interoperability and semantic interoperability (Bishr 1998). The syntactic interoperability refers to use common data format to exchange information between BIM and GIS systems, using a domain of one of the information systems. Examples of syntactic interoperability are systems that combine building data with landscape maps, data formats of BIM object on GIS environment, and solutions to convert BIM data from IFC (Industry Foundation Class) to CityGML (Akinci et al. 2008). The approach of IFC to GIS (IFG) project has been developed to provide geographic information between the frameworks of IFC, in order to get a more efficient planning (Kolbe, Gröger, and Plümer 2005). Other examples of syntactic way are based on the use of CityGML schema. The GeoBIM extension from BIMServer allows conversion from IFC files to CityGML files by defining additional information in CityGML entities (de Laat and van Berlo 2011).

(Nagel, Stadler, and Kolbe 2009) proposed a conceptual method of transitioning from a KML graphics model to BIM through CityGML. The authors developed mapping rules to allow transforming CityGML model to IFC model. Another attempt, tried by (Hagedorn et al. 2009), was to create a conceptual dual graph for representing topological relationships among indoor entities of a building, but not the whole one with the geospatial context. All these examples chosen to explain this type of interoperability show a common characteristic: they follow a unidirectional way of translation and do not consider the semantic information mapping in the process. A bidirectional approach is needed (Deng, Cheng, and Anumba 2016), if we want to get a dual interoperability between BIM and GIS.

The highest level of interoperability is guaranteed by the semantic aspect of data integration. The key point of semantic interoperability is to make sure that features and relations between information management systems are maintained during data conversion (Peachavanish et al. 2006). Objects as entities and their relationships are defined under a domain called ontology that makes possible the representation, the sharing and the management of the knowledge (El-Diraby et al. 2005). The information should be described and classified in a standard way. The semantic web technology represents new efficient online platform to make possible this kind of interoperability. The web ontology language (OWL) expresses the data in terms of classes. A collection of these classes, their attributes and relations can be stored as RDF triples describing each individual object, its properties and features, which can be understood as a graph based on nodes (entities) and vertices (relationships) (Hor et al. 2016). Semantic web technologies have been used by several researchers to facilitate construction project information sharing. (Anumba et al. 2008) explored the use of semantic web technologies to meet the challenges of collaborative project information management. (Akinci et al. 2008) developed a eb-based approach to enable semantic interoperability between CAD and GIS platforms. (Beetz 2009) demonstrated the feasibility of semantic web tool to address information exchange and integration problems in AEC interoperability. Taking this approach into account, it becomes possible to move forward a graph database direction, useful also for a BIM data interaction providing so a complete interoperability in the web environment and so allowing an eventually bidirectional way of data transfer between information systems.

## 2.6.1. Methodology

The improvement of GIS model, by creating connections and data conversion towards another domain, such as BIM, has been already discussed. The aim of this work is to transfer the data from a GIS model to a common web database where all its features remain interconnected. Other models, as BIM one, can be also imported in this web database and linked to the GIS model. The hypothesis is that every model is created using open standardized ontologies and keeping them is the main purpose. The resulting database defines the connections between models in the same information system, where a feature can influence features of other models. Every feature in the model is connected to other thanks to the ontological relationships.

Taking the urban context as example, objects composing the city model can be created and collected in different information systems where, basing on own ontological schema, specific features are managed. A unified model, which gather all the information coming from several information systems, will be built in a semantic web environment where all the figures involved to the project can interact with it. Among the urban services, this approach can be used for analysing the road system, in particular the management of the emergency, security and traffic when an obstruction may compromise the passage. For example, works on a façade building near the adjacent road may imply a traffic diversion. Furthermore, the unified model will be useful to create spatial connections between elements of the urban area. In this work it was considered a simple case, as a demonstration, that recognizes buildings rising in sequence along a street, testing the operation of building-road connection.

In order to test the connection building-road, we have chosen a case of study concerning an urban area of the city of Bologna, characterized by the presence of porticos which need interventions. The methodology is centred on GIS model and its transformation to the common database. The workflow is structured on steps that are outlined as follow. First, it is necessary to prepare the GIS model based on the available data. Shapefiles (.shp) of roads and buildings have been downloaded free from the OpenData, an open source platform provided by the Municipality of Bologna. These objects allowed to realize a 3D GIS model, where first we made the Triangulated Irregular Network (TIN) of the terrain. Then we built the 3D shapes of the buildings and defined the road map, laying all them on the TIN. The second step consists in importing shapefiles and other data into FME (Feature Manipulation Engine) software to create CityGML and IndoorGML models. FME is an ETL (Extract, Transform and Load) tool, which automatically translate the information from one format to another one (FME 2019). Finally, using an algorithm developed in Python, to transform CityGML and IndoorGML models into JSON data, we collected them into a graph database on web environment. We have chosen the solution of a NoSQL database, because this kind of database, thanks to its flexibility, scalability, and high performance, is able to collect a wide variety of heterogeneous data in modern applications. Next paragraphs explain deeply the methodology adopted (Figure 10).


*Figure 10 Research workflow*

### 2.6.2. 3D GIS model

The information, provided by cartographic services, is expressed in bidimensional way where sometimes elevation values are specified. In order to create a 3D GIS model, we need to make modifications in the original 2D data. As the case study represents an urban area, a 3D city model has been created from Digital Terrain Model (DTM) and 2D shapefiles of buildings and roads has been imported. The building shapefiles consist in polygon features including height data, whereas roads are drawn by polyline features. The latter have planar coordinates (x, y) but do not have elevation values (z). As previously indicated, the aim of this research is to create spatial connection between GIS object of buildings and roads,

through a representation based on nodes. In this case, the connection point of a building with a road is assumed to be the nearest point of the road (road nodes) to centre of gravity of the building (building nodes) calculated within ArcGIS (ArcGIS 2018). The roads are splitted at these connection points obtaining a logical network.

Furthermore, for an easy and fast recognition, an univocal identifier (ID code) is assigned to the resulting roads, to the road nodes and to the buildings. All these data, ID codes and number of nodes of the roads, are stored in a specific datatable saved in CSV format. Therefore, to fill the building datatable, a new column will be added to contain the ID code of the corresponding road node (bottom left of Figure 11). In the same way, the ID codes of the starting and ending nodes for the corresponding segment are selected to the resulting road shapefile. The data preprocessing and the code identification help the following creation of the IndoorGML model, built after defined the CityGML schema. However, buildings and roads are still 2D features. In passing to the 3D level, first, the z coordinate (altitude) of the base feature needs to be retrieved from the TIN (Triangular Irregular Networks) of the terrain, which has been obtained from contour maps with accuracy of 2 m. Thus, knowing the height of the buildings, they become 3D polygons. In this way, 3D polygons are represented by solids built extruding the baseline polygon to their height (obtainable from the building datatable) and using multipatch features (upper right of Figure 11). At the same time, lines of the road map follow the three-dimensional orography of the TIN model.

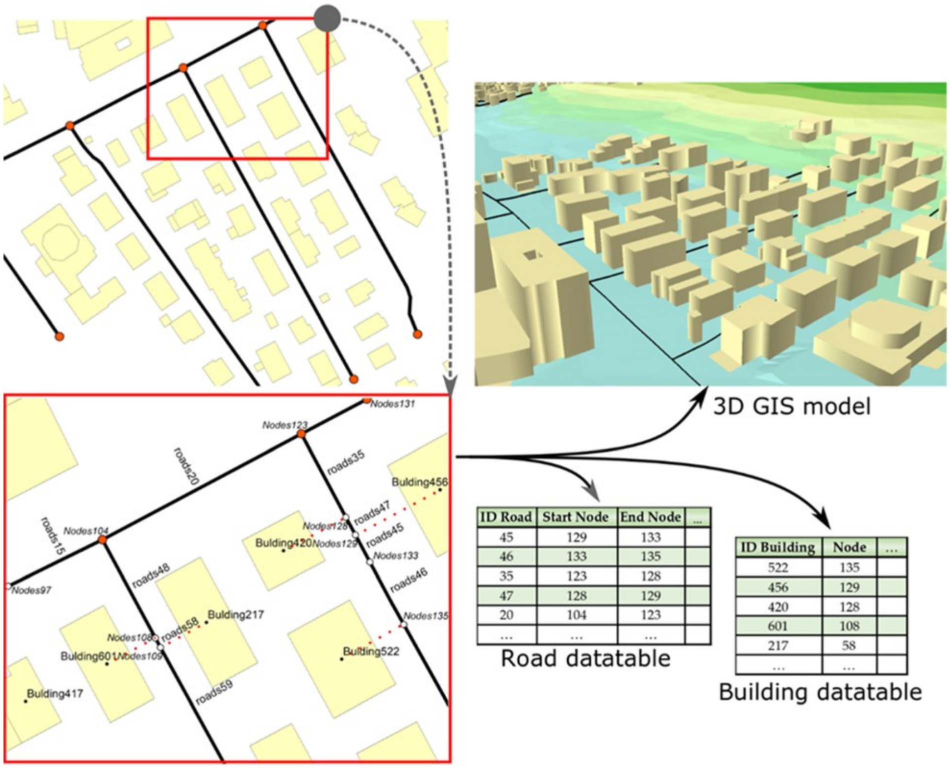



*Figure 11 Operations carried out from the original data, at upper left. A zoom view, in bottom left, shows ID roads nodes, ID buildings and ID roads (linked by a dashed line). The assigned ID codes collected in datatables, in down right. Upper right: the final 3D GIS model*

## 2.6.3. CityGML model

As has been previously indicated, CityGML is chosen because it is widely used in GIS city modelling and represents an open recognized standard and well-defined ontology. In this project, the CityGML model was constructed based on the 3D GIS model using FME software. 3D building and road shapefiles are imported into FME, and they are transformed to LOD1 Building and LOD0 Road objects in the CityGML model. The TIN is also transformed to a 3D relief and incorporated to CityGML model. FME has a set of tools that carries out this conversion with only specifying its input data. In this way, this ETL allows the designer to easily create the CityGML model. A previsualization of CityGML geometric result from FME interface is shown in Figure 12. The main problem of this data is the way in which semantic connection between these independent elements is created. One way to create semantic relationships among objects in CityGML is through geometry, but it would require a full topologically consistent model with high LOD. This is rarely achievable. For example, many building and road shapefiles available do not share any point between them. Consequently, topological connections cannot be carried out and both kind of objects remain independent. Other problem is that FME does not automatically obtain geometrical relations between buildings: if two buildings share a wall, this geometic element is written in both buildings CityGML model. For these reasons, other ontological domain is needed to fill this gap and IndoorGML semantics adjusts to it.

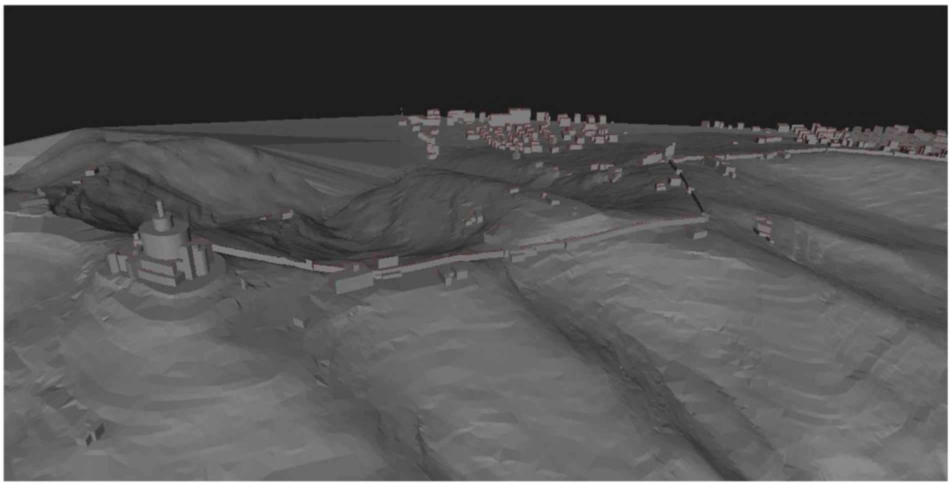


*Figure 12 CityGML model*

## 2.6.4. IndoorGML model

In this research, a slightly different interpretation of IndoorGML has been developed to make connections between CityGML elements. In this case, IndoorGML acts as linking network between outdoor and indoor spaces. An improvement of its ontology could deal with any real space because the paths can be used for several case studies, such as WiFi connectivity (IndoorGML 2019). The proposed general model could have a central navigation module with some predefined extensions in indoor and outdoor environments, but this development is not the scope of this research. The aim of this work is to use IndoorGML ontology with others existing standardized to create connections between buildings, identified by a point in their centre of gravity, and the nearest point belonging to road lines (nodes).

CityGML and IndoorGML have some similar features, in particular: geometry, existing in both models, and CellSpace object in IndoorGML which corresponds with Buildings object in CityGML. In Figure 13 is shown a schematic diagram of CityGML and IndoorGML considered and the proposed connections in geometry and CellSpace.

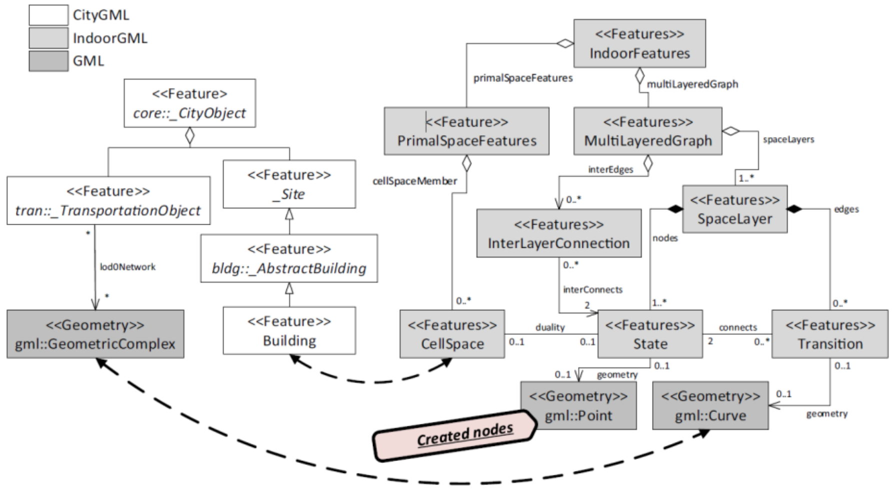



*Figure 13 SimplifiedUML diagrams of CityGML and IndoorGML. The dashed line indicates the connection between both structures and the label*

An useful software to easily create an IndoorGML has not been found. FME specifications indicate that thisXML format has already been incorporated, but there is not a direct processing to generate IndoorGML data. Therefore, a new workflow was created using FME software to write an IndoorGML file, although any other software (or even a script) could be used. This workflow takes as input data building–road connections as specified in “3D GIS”. in CSV format (created from Roads datatable and Building datable represented in Figure 11).

The following explanation illustrates the process of how the UML (Unified Modelling Language) diagrams of CityGML and IndoorGML were interconnected using FME while creating connections between the two shapefiles, output of the 3D GIS data, previously created.

To elaborate the proposed procedure some considerations have to be taken into account. Every building is considered as a unique SpaceLayer element of IndoorGML ontology because they are unrelated. Inside this layer, all buildings are defined as CellSpace and assigned its corresponding ID as is possible through GML specifications. On the other hand, roads form another SpaceLayer element and are defined by their geometry (line entities for Transition elements and point entities for State ones). Connections among elements from different SpaceLayer are made with InterLayerConnection feature.

Furthermore, Transition feature can connect two States features using nodes created in “3D GIS”. It is therefore possible to use for roads connections between stretches of road or buildings and roads. The connection of nodes with road segments and buildings creates a network between elements. It achieves the objective of having one interconnected model where you can navigate from one feature to others. The connection between IndoorGML CellSpace and CityGML Buildings and geometry connections are carried out through its ID, which has to be the same in each element of both

models to connect them. The importance of such process will be better understood in the next paragraphs, where the use of a graph database will better explain how these connections can be used to improve an information model through queries and enriching the data. As a result we obtain the model in GML containing all necessary links for the graph database.

### 2.6.5. Implementation of graph database

The use of open source standard facilitates end users to understand every model. But these standards are usually disconnected among them as is the case of CityGML, IndoorGML or IFC. As previously expressed, some solutions exist and the approach used here is useful to create a common system where an entire model can be incorporated. This common system is represented by a graph database with RDF triples.

Graph databases belong to NoSQL databases. These types of database are useful when they have to store unstructured information. Furthermore, they allow to carry out fast transversal queries. Nowadays, there are a variety of software in this field such as ArangoDB (ArangoDB 2019;, MongoDB (MongoDB 2019), or Neo4j (Neo4j 2019) (Fernandes and Bernardino 2018; Hor et al. 2018). We chose to use ArangoDB for its speed carrying out traversal queries, but any other could be used.

These databases have multi-model structure because they can collect information in different ways. This methodology is based on the use of documents and graphs (ArangoDB 2019). Documents store information, whereas graphs define relationships between data. Both are used by ArangoDB.

ArangoDB uses JSON (JavaScript Object Notation) format to store information in documents. Each one of them can contain different type and quantity of attributes. Documents are stored into collections assigning them, automatically or manually, a univocal key value. There are two types of collections: vertices and edges. The main difference between them is that edge collection has two special attributes that vertex collection does not have: _from and _to. These two attributes are used to create relations among documents of any vertex collection stored in the database.

The diagram of Figure 2 shows that these graph databases can be defined graphically, with vertices and edges, where vertices are documents and edges are relations. It allows to use RDF graphs triples to define the data in ArangoDB where Predicates (from RDF) represent edges, whereas Objects and Subjects are vertices (or documents).

### 2.6.6. From GML to JSON

Both CityGML and IndoorGML are GML formats, but currently, ArangoDB only imports JSON or CSV formats. Consequently, GML models must be converted to one of these formats. The simplest way to create objects is to transform in JSON rather than in CSV.

Indeed, there is an encoding for the OGC CityGML data model called CityJSON (CityJSON 2019). CityJSON mainly describes the geometry, attributes and semantics of different kinds of 3D city objects. While we need to manage only the geometry of the objects and their geographical position to define the spatial relationships, we opted to use the GeoJSON format (Butler et al. 2016), mostly used for the road map management (Gunawan, Ferdinandus, and Setiawan 2017). GeoJSON supports geometry types as point, polygon and multipoint. All things, which are supported by JSON, are also supported by GeoJSON. The difference between JSON and GeoJSON is that the key naming of each array element in GeoJSON has to follow certain guidelines. It is because the structure of GeoJSON follows the international standard published by OCG. GeoJSON has a specific function to support geographical data with a standardized format. It defines several types of JSON objects and the manner in which they are combined to represent data about geographic features, their properties and their spatial extents. GeoJSON uses a geographic coordinate reference system, World Geodetic System 1984 (WGS 84). As just explained, GeoJSON is a JSON format to store and exchange spatial data e so can be imported in ArangoDB.

As an RDF graph database is used, it requires to define both nodes (subject and object) and the linking edges (predicates). First of all, GML format is constituted by elements between one start-tag and one end-tag (Figure 14). Tags represent the kind of element that is being saved between them: for example they are the ontological names of the element. An element could be empty, have information concerning the element or have other nested elements (called child elements). An attribute is a name–value pair placed after the start-tag. In order to create ArangoDB–RDF graph from GML formats (CityGML and IndoorGML) the main hypotheses are listed below:

- An empty initial node is created to which all main elements are linked. It helps to have a connection between all main elements of the model.
- Each tag is translated as an edge that links two nodes.
- Nodes store the information concerning the element or its attributes. If it only has nested elements, it will be an empty node.
- The attributes of a tag are saved in the document or node that acts as object (child node).
- If the attribute “gml:id” exists, it is used to define the document identifiers through the ArangoDB keyword: “_key.” Every document in a collection has a unique identifier.
- If the attribute “xlink:href” exists, it is translated as an edge that links the parent element to the object at which it refers.
- The geometry is translated to GeoJSON format and saved in the child node.

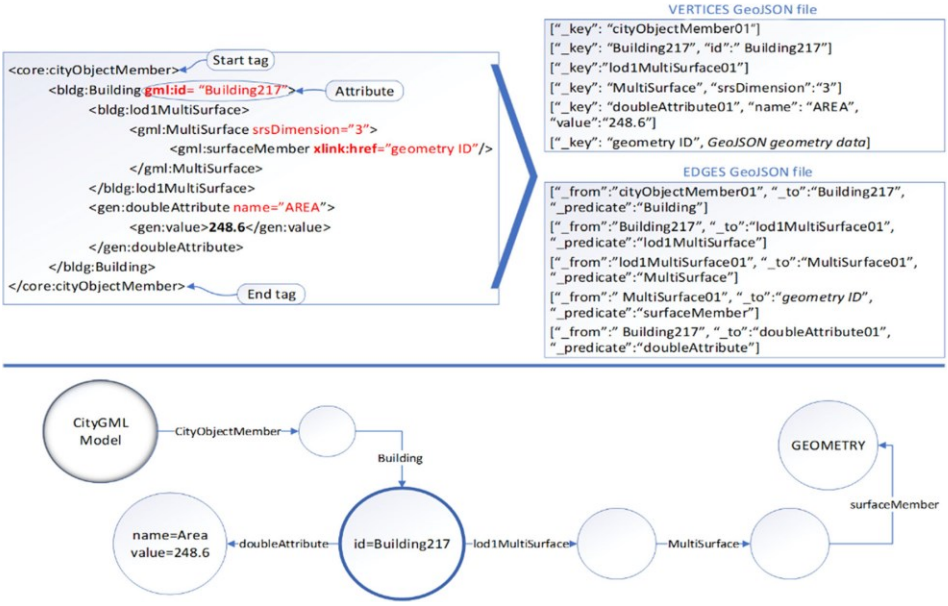



*Figure 14 One CityGML Building object with a generic attribute and its transformation to GeoJSON file. Above the definition of vertices and edges file, below the graphic representation of the relations*

In order to import data in ArangoDB, data must be translated to JSON format. While any useful software has not been found to carry out this transformation, a script in Python (Python 2018) has been developed. It uses XML library for reading the file and OGR library (GDAL 2019) for translating GML geometry to GeoJSON. Its outputs are two collections in JSON format for each model transformed: the edge collection and the document (vertex) collection. In this way, ArangoDB can read them as individual documents and assign them a unique identifier (if it is not defined with "gml:id" attribute). The script is based on schema less conversion. For example, it only reads the data model without checking the schema structure. However, it can become tedious and prone error when GML has many abstract type elements contained in the schema. XML schema defines that abstract types cannot be called from models or other schemas. So, they will not be defined neither in the data model nor in the graph database, while they will be shown in UML diagram. For instance, "_AbstractBuilding" does not appear in the graph diagram in Figure 15.

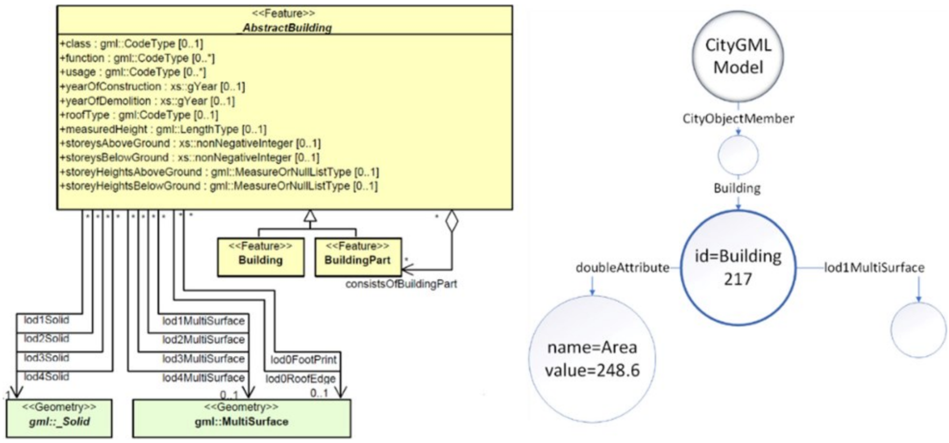



*Figure 15 Comparison betweenUML schema of CityGMLand Graph database. The abstract element AbstractBuilding does not appear in the graph at right because it cannot be created in CityGML model due to GML language specification*

As previously said, the initial node is empty and refers to all CityGML model. In this example, only one building defined by one geometric and one generic attribute composes the model (Figure 15, right). The predicates, or edges, are the names of the schema objects that are not GML abstract elements. Thus, following this representation, firstly appears the predicate “Buildings,” which defines the element type. The object of the first triple contains the building ID (i.e., the vertex after the edge with predicate “Building”). The geometry, inside the building element, is determined by other triple. The UML diagram (Figure 15, left) shows different ways to define the geometry. In this case, there is a “lod1MultiSurface” element that can be identified in the graph. It composes the final triple of this example, where the subject is the node of the building, the predicate is “lod1MultiSurface” and the object will be the geometry itself.

Figure 16 shows the base schema of process used by the Python script. It takes into account the hypothesis shown previously and creates the JSON file following this procedure:

1. Read GML model as if it was any XML document and create the initial empty node to which all the main model elements will be linked.
2. The elements from the file are read and stored in memory. Most of them are processed following the same method, but there are some special cases where they require a different treatment. Mainly, these are included in IndoorGML and are listed below:

    - Cellspace element. It is part of IndoorGML model, and it is referred to a CityGML building through its ID.
    - Duality element. It is part of IndoorGML and it creates a connection between a Cellspace and a State from a linking element using its ID. In this case, it is a vertex from the roads.

• Attributes elements. They are part of CityGML generic attributes and allow creating special attributes. They are saved in a document, linked to the parent element using the tag-name and saving the data in the document using the property name "value." If tag-name attribute is defined, it is also stored.

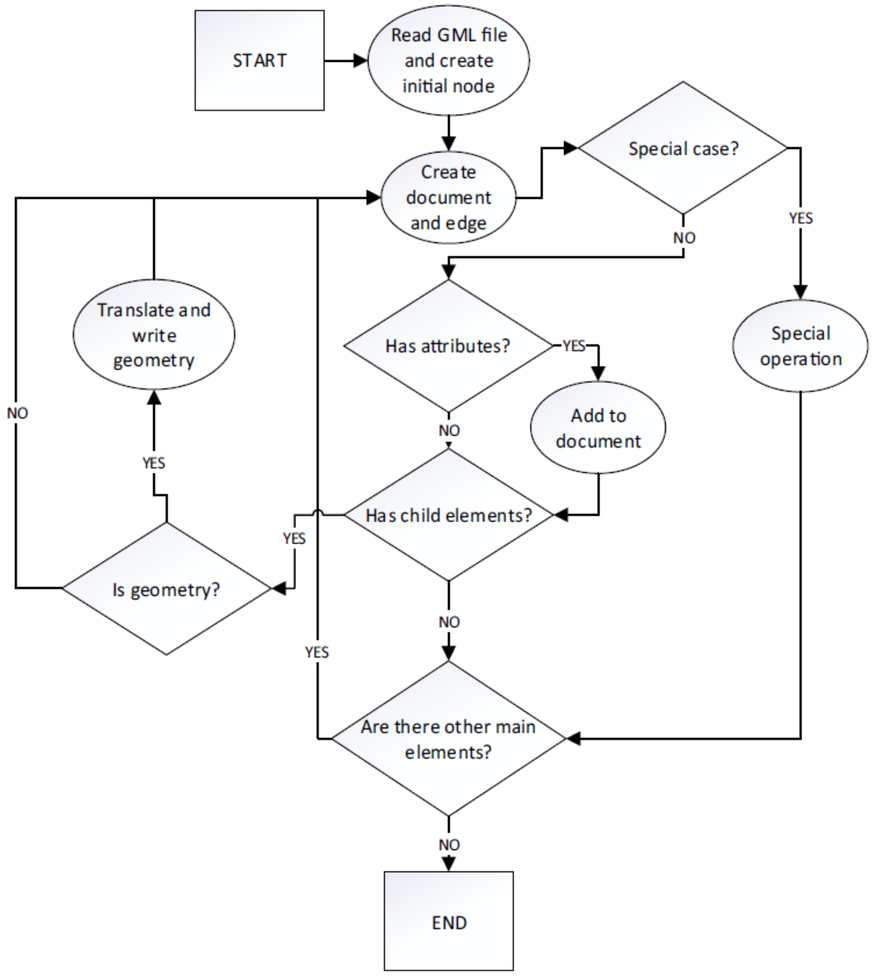



*Figure 16 Script workflow*

If there are not any special cases, the elements pass the same process. Tag attributes (if they exist) are saved in the document that has been previously created. Then, the script checks the existence of nested elements. If the next nested element contains the geometry, then it is translated to JSON and saved in a new document linked by an edge, which its name is the same name of the tag element. The geometry names in CityGML can be seen in its schema and mainly depends of the LOD referred to.

In order to allow linking all elements, a key value is assigned to every document created, using the "gml:id" namespace, if the actual element does not have an ID code. The edges are defined by the key words "_from," which is the ID of the parent document, and "_to," which is the ID of the actual document. The edge also stores the tag name, which represents the predicate in RDF, inside a string with name "_predicate." It creates a new value that is not part of any other ontology. To solve this, the predicate could be the name of the document collection. But this option would create an enormous quantity of files: one file for every kind of element. The visual management of this database would become more difficult and more complicated.

## 2.6.7. ArangoDB graph database

The result of the last procedure described creates two JSON files for each model that can be imported in ArangoDB. One of them refers to a document collection, while the other one makes a reference to an edge collection.

Referring to the urban model of the case study, in order to carry out a simple and explanatory analysis, only buildings and roads were considered. The number of building features in the selected area is 876, whereas the number of road features is 84. One CityGML and one IndoorGML model were constructed based on available cartography and then encoded in JSON format. CityGML occupies 4.3MBof disk space and IndoorGML 0.9 MB.

The CityGML model encoded in JSON creates a total of 48,831 documents in 4 MB of disk space, and 48,832 edges in 5 MB to link that documents. The documents created with IndoorGML case are 12,133 in 0.5 MB and 19,905 edges in 2.2 MB. It shows that IndoorGML is mainly used to create links among CityGML objects because of the higher number of edges in relation with documents.

A graph centred in one building is shown in Figure 17. This graph shows in vertices the value of _key field and along the edges the value of the created _predicate field. At the centre of this graph is represented the vertex with Building217 as _key value (which it is the same ID in the GIS model). Most of the vertices around it are of the same collection, i.e. from CityGML model, but those placed at right most are from IndoorGML collection. This central node represents a building because it only acts as object in one triple considering only *CityGMLcollection*. Also is easily visible that it comes from a predicate *cityObjectMember*, which is the first element of a CityGML.

Every element in the model is always connected to others, so the graph is more extended than shown in Figure 17. Navigating through the graph model and reaching any searched element is possible using ArangoDB Query Language (AQL). In addition, it allows to carry out queries to see, for instance, which buildings are in a determined route without using any geometrical operation.

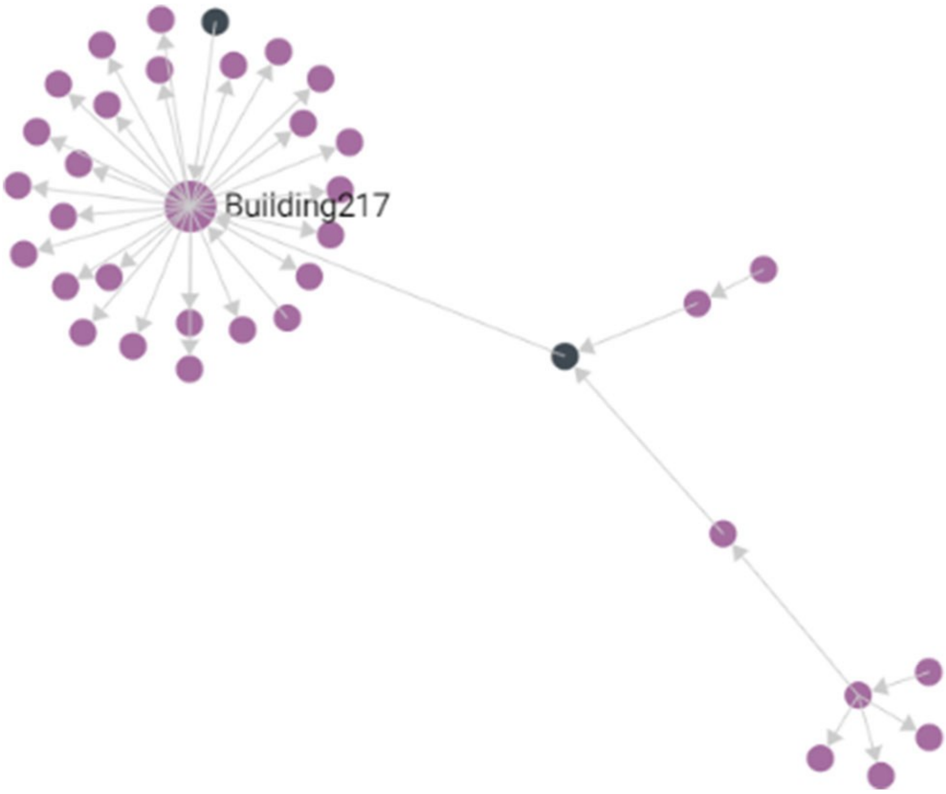



*Figure 17 Part of the ArangoDB representation of the CityGML*

To move from one element to another all paths must be previously defined. These paths would be based on predicates and logical relations between elements. Although there are solutions in ArangoDB to search using the shortest path, but sometimes it could report a not correct one. For instance, in this model there are always a minimum of two paths to arrive from one point to other one.

### 2.6.8. Discussion of the results

Data modelling and management software, related to geomatics, urban planning, project management and many others, represent today a fundamental tool in both working and research fields. However, the exchange of data between information systems is still a huge challenge, while, if possible, it could bring enormous benefits to those who work with them.

It becomes necessary to find a solution where data interoperability is lacking, that can be replicated by everyone. This could be possible using tools already standardized and accessible to everyone, which means basing on defined and known schemas and ontologies. For this reason, it becomes essential to go through CityGML, as it helps the designer to build a standardized ontology. Different tests, to prove the interoperability between information management system, have already been presented by other researchers moving in both directions, especially focusing from GIS to BIM. They try to integrate both spatial and non-spatial data basing on the definition of domains or ontologies, and sometimes using semantic web technology.

The work carried out in this work aims to create a new enriched model, compared with the standard GIS one and demonstrates that it could be possible to interoperate those modified data towards a new unified data model without losing any information. To do so, we decided to create connections between different elements of the urban model in GIS allowing the interaction among them. This implicates that the formulation of queries is based specifically on these relations, and not only the geometrical elements. We adopt the use of CityGML and IndoorGML schema, but also other standardized format could be selected and added. The graph database represents a useful tool to collect all kind of data, converted in JSON formats or similar, creating RDF triples that connect all the elements and their attributes. The graph database preserves the ontologies during the various data conversions, and it turns out to be a powerful tool in order to improve considerably the performance of data management. Graph databases can query the files contained in them through the new defined ontological schema and basing on the relations created between the nodes.

The work develops the possibility to converge everything that was originally a GIS system in a graph database, based on NoSQL technology. This methodology improves the way in which the identification of affected areas by restoration works of a building is made. It allows making better analysis of operational decisions such as traffic affections, space requirements, affected neighbors and planning alternative routes. Keeping a standardized ontology helps to teammates in planning, managing and operate the building site. Or even to all stakeholders associated to the project as could be the public administration.

Moreover, this research test leaves open the possibility to import BIM data in the same graph database and so to join both GIS and BIM models in the same semantic web environment. To do so, there are many related works which show how to create a unified model which allows to link data (Song et al. 2017; Volk, Stengel, and Schultmann 2014) for instance how to build a common ontology BIM-GIS and eventually how to export data in IFC format that can be successively import in BIM environment.

# Chapter 3

# 3. Tracking systems

## 3.1 Geomatics techniques for Outdoor-indoor systems

When it comes to safety, it is mandatory to keep in consideration not only buildings and surrounding as just discussed in the previous chapter, but especially the people working and living those areas. In fact, geomatic researchers found a lot of interest on people tracking outdoor, indoor and while transiting from the twos.

Outdoor tracking technologies offer the possibility of obtaining information on people's location at low cost and in real-time. The integration of GNSS in smartphones, in combination with applications such as Google Maps or similar, gives us a very clear idea of what a digital twin is. Although the error of these types of GNSS is of the order of a few meters, it represents a more than satisfactory model for the purpose it aims to achieve, namely to help the user reach a place within a city, whether it is a strategic site or any building, we can recognise it very well even if the accuracy offered by this system is not centimetric. It applies to all people moving in an outdoor space, especially when it comes to monitoring positioning in terms of safety and risk management(Cheng and Teizer 2013). We do not need to know the centimetric position in this type of survey, as it is not an architectural survey where often centimetric or sub-centimetric accuracies are required.

This means that low-cost systems can be used with excellent results for this type of data acquisition, which is now affordable from everyone, so no great economic effort is required to implement these solutions. The problem arises when people move from outdoors to indoors, because GNSS-based systems do not work in these environments (Mautz 2009). This is due to the fact that GNSS antennas are not reached by the satellite signal in indoor environments, and they also encounter difficulties in outdoor environments where there are obstacles of various kinds. This issue was addressed in the literature by Angelats and his team at the research centre “Centre Tecnològic de Telecomunicacions de Catalunya” (CTTC). The aim of their research was to track members of civil protection or emergency teams continuously from outdoor to indoor areas and vice versa.

Although several approaches for indoor positioning exist in the literature and on the market, these are not appropriate for tracking in real-time emergency staff because these approaches rely on existing infrastructure and there are no examples of wearable devices which can track in real-time nor combined solutions to solve the outdoor/indoor problem. (Angelats et al. 2021)

The authors believe that current technologies, both in terms of software and hardware, are powerful enough to build a system capable of meeting this type of demand based on IMUs, RGB-D cameras and GNSS receivers (Angelats and Navarro 2018). The proposed device is meant to improve the decision making process based on reliable and real-time information. (Angelats, Navarro, and Parés 2020). All these devices were installed on a helmet and will capture two trajectories in real-time, one based on the GNSS signal when outdoor and the other one on the camera, more precisely we may say based on the Visual Inertial Odometry (VIO) for indoor areas (Figure 18).

*Figure 18 Positioning system mounted on a helmet designed and built by CTTC*

Although this work has been proven to increase the safety of workers, such as firefighters and civil protection staff in emergency scenarios, the final result in terms of produced trajectory can still be improved. This is due to the fact that switching device, for instance, from outdoor to indoor is based on a simple standard deviation calculation on the GNSS signal. As a result, the overall trajectories will be accurate in outdoor areas and in indoor areas, but the connection between the two is not as precise and this can lead to a final trajectory with a big shift in indoor areas. Here comes the contribute of this thesis work.

Since one of the objectives of this thesis is to improve life and work of emergency staff in dangerous urban areas, such as those cities or villages hit by earthquakes, it was useful to understand the approach explained so far so we could collect more information and an attempt to improve this approach to get a better final trajectory is proposed in this chapter.

### 3.1.1. GNSS signal related to environment

The borders between indoor and outdoor are not simple to define, in her review (Zlatanova et al. 2020) makes a clear collection and exposure of definitions and space comparisons in different domains and from other authors in spatial science and urban applications. As stated in the same review: an increasing number of standards and applications are working towards three-dimensional notions to be able to represent parts of space.

These aspects were investigated mainly because the key to linking the outdoor/indoor trajectories is a function of the space due to the fact that the GNSS signal is influenced by the space. For instance, even outdoors, tall buildings may create the so called urban canyons. Narrow streets between tall buildings will interfere with signals from satellites and will cause big errors in the tracking of the GNSS device. Also, other elements interfere with such signal. A very dense canopy is another case where the GNSS signal may not be reliable, but there are also a lot of intermediate cases among open outdoor area and urban canyons (Xie and Petovello 2015) or dense forest areas(Næsset and Gjevestad 2008). So, it is natural to ask which is the limit, in terms of environment, we may consider an outdoor area with acceptable signal noise.

There are many studies presented in literature regarding the definition of spaces in terms of navigation. Some define the indoor space simply as a place bounded by physical boundaries (Zlatanova et al. 2014; Zlatanova, Liu, and Sithole 2013) such as walls, doors and windows. Also underground spaces may be classified as indoor space (Yang, 2011) but other studies consider indoor space only when it's about building, excluding all natural environments (Yang et al. 2011) depending on the use case. Only a few papers treat semi enclosed areas that are not clear and simple to define as outdoor or indoor. (Kray et al. 2013) proves that using only two types of space are not enough. It must be considered at least a third one known as transitional space.

From the point of view of this thesis, what matters is the signal of a GNSS receiver. In terms of reception of GPS signal (Ortiz et al. 2015) considered whatever was GPS-denied environment as indoor or semi-indoor, while (Wang et al. 2016) provided another classification of the areas: open outdoors, semi-outdoors, light indoors, and deep indoors as shown in Figure 19. Areas such as urban canyons or wooded areas, are classified as semi-outdoors. Light indoor is similar to semi-outdoor but inside the building. These are areas around windows which still have some satellites availability. Deep indoors refer to places without any satellite coverage.

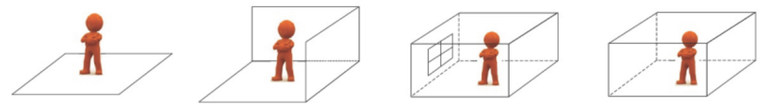
| Environment | Open Outdoors | Semi-Outdoors | Light Indoors | Deep Indoors |
|---|---|---|---|---|
| Definition | Outside a building | Near a building | In a room with windows | In a room without windows |
| Example |  | | | |

*Figure 19 Space definition according to the reception of satellites signal* (Zlatanova et al. 2020)

As a first attempt, this approach seemed to be the one that better fits this study since, in terms of GNSS signal used for this calculation, it summarises and gives a quick idea of how the signal could be affected by the surrounding space. Several considerations though must be made to find the limits of the so called semi-indoor and semi-outdoor areas. For example, is any building affecting the GNSS signal? Or must it be taken into account only tall buildings? How tall it has to be to influence the signal? How far from the building it will be considered still a semi-outdoor area? And the same for semi-indoor: how near to the window should the device be in order to be considered semi-outdoor? Shall canopy spaces be considered in this classification? How dense must a forest be to be considered light indoor or deep indoor? How to treat a single tree?

Obviously, some of these questions may seem to be trivial and others without a proper answer. Others already have an answer in literature. For instance, it is well known how the GNSS will behave close to a building or in dense urban areas. This is a function not only of the distance from the building but also of the building's height. The reason the GNSS receiver will give a higher error is due to the fact that some satellites are not directly "visible" from the GNSS (Figure 20 left), in other words, are Not in the Line Of Sights (NLOS) of that device. When such a situation occurs a multipath effect is expected. It is the combination of the direct and reflected path signal as shown on right of Figure 20 (Feriol, Vivet, and Watanabe 2020).

In order to understand how to treat several environments, we must know how the GNSS receiver will respond by testing it in such spaces. Only after it can be defined a limit for semi-indoor and semi-outdoor. Context detection for navigation purposes starts with testing the GNSS in the above mentioned areas. Observing the Rician K-factor and the C/N0 (dB-Hz) values given by the GNSS device in relation to the counts of the satellites it is possible to make an accurate estimation and set limits and parameters for the investigated semi-indoor and outdoor.

Nevertheless, the limits set are prone to many factors and variables which also depends from the hour of the day because of the position of the satellites and the weather. This causes the limits to be a vanishing area rather than a proper line in each situation. It came out that finding a proper rule in context detection using only GNSS signal in order to define areas is a complex achievement. The approach used so far for the detection of the indoor areas moving from outdoor and vice versa, was to use the standard deviation on the trajectory collected by the GNSS. As stated at the beginning of this paragraph this led to some issues in terms of joining the GNSS trajectory with the one form VIO. The key to improve the combined trajectory is still hidden in the trajectory itself, so the trajectory data must be studied in order to obtain a trend which explains the behaviour of the GNSS in different spaces and the algorithm must recognise the trend changes in time, otherwise linking a new trajectory to the GNSS one, will result in big errors. These aspects will be explained in detail in the next paragraphs.

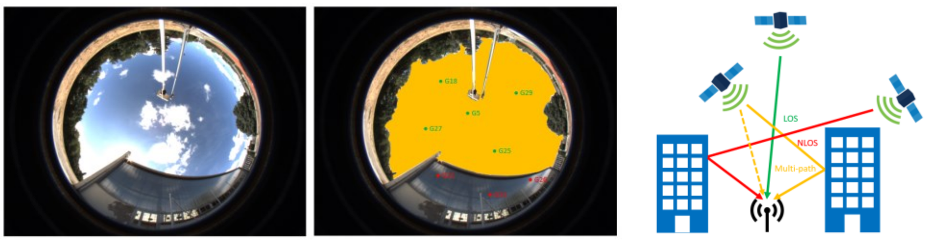



*Figure 20 Left side: sky representation with projected satellites (green = LOS; red = NLOS). Right side: schemes of the multipath and NLOS effect.* (Feriol et al. 2020)

### 3.1.2. GNSS trajectories and tests

After the consideration made above and before going into the trajectories collected with the helmet shown in the previous chapter, some areas were chosen to make the consideration mentioned on the space. Not only for simplicity but also because it represents a good scenario where there are several elements where the GNSS can be tested due to tall buildings and open outdoor spaces. Every kind of scenario which has both outdoor and indoor spaces is a possible emergency scenario, for this reason the trajectories will be collected in different places: the CTTC campus, a small town with narrow streets named Garrag, canopy areas, tunnels, etc.

A first GIS model of the campus was produces thanks to the data downloaded from the "Centro de Descargas" from the Geographical National Institute and "Minister de trasportes, movilidad y agenda urbana" of Spain. Thanks to Open Steet Maps plugin in QGIS the shapefile of the building was easy to generate, therefore the height and any information of the building was not provided. Here it comes in help the data downloaded from the Centro de Descargas, where Digital Elevation Models (DEM) and Digital Surface Models (DSM) of the interested areas were downloaded. Such data are in raster or lidar format. So, it was necessary a simple query to get the elevation of each building and of the terrain.

*Figure 21 QGIS model with the raster of the area at CTTC campus*

Based on the experience of the literature, so knowing the effects of the buildings on the GNSS signal when outdoor, a first area of influence was set in order to define a very rough semi-outdoor area. As the first attempt, it was set by taking into consideration an angle of 30 degrees from the top of each building (Figure 22). Lately it will be reduced to 20 degrees since the empirical test showed that the trajectory trend was not influenced that much and such a small variation in the trajectory can be caused by any small fluctuation that could influence the signal. Another reason whereby 30 degrees was too much is due to the admitted error of this system. As already discussed, the acceptable errors are a function of the final aim, which was to track a person in an emergency scenario. It means that an error of 2 or 3 meters is acceptable as long as the trajectory remains coherent.

Figure 23 shows the result of the influence area with an angle of 30 degrees and the collected GNSS trajectory was placed.

It must be noted that the path followed by the user carrying the GNSS + VIO system mounted on the helmet is going from outdoor to indoor and then outdoor again. At a certain moment, the user is also walking very close to the building in the bottom left in Figure 23 in a straight line to observe then the behaviour of the trajectory due to multipath and after is entering two different buildings on the right side of the same figure. Here is shown only the GNSS trajectory. The latter is the one which will be used to understand the spaces based on its trend. Since it is only visible the GNSS trajectory, when the user enters a building it is clear that the data is not reliable anymore.

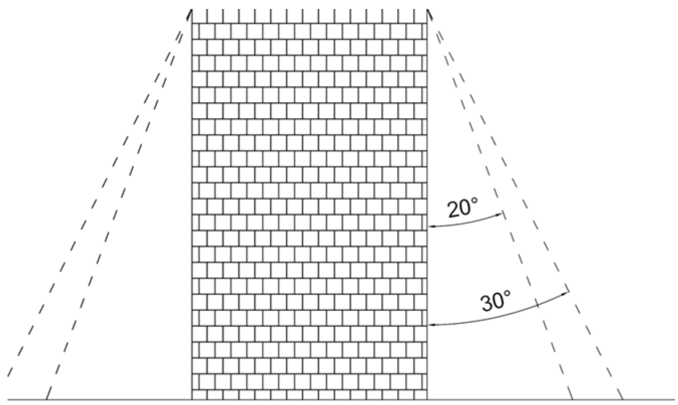


*Figure 22 Proposed area of influenced*

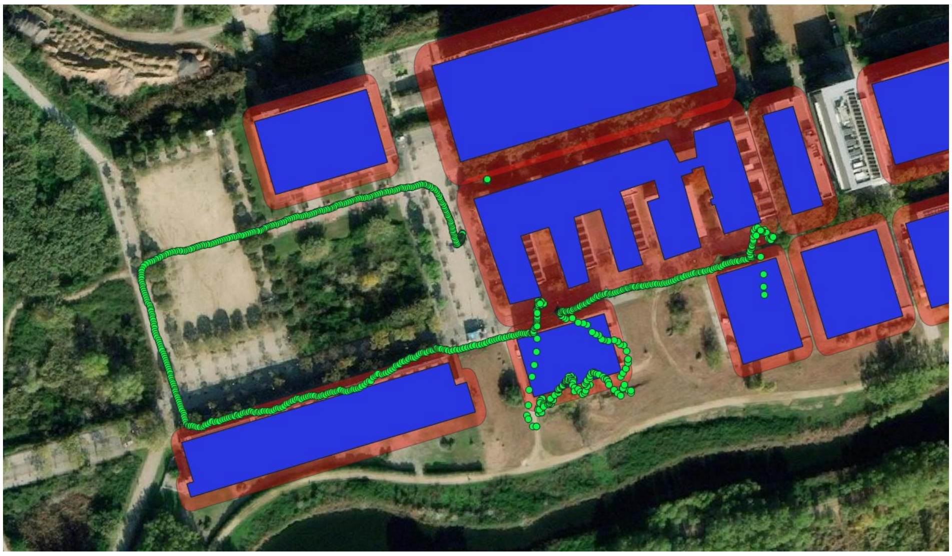


*Figure 23 Behaviour of the GNSS signal at the campus*

The trajectory has been investigated with the aim of understanding how it changes not only when going into the buildings and in outdoor areas but also when coming from the inside of a building. The data available produced by a GNSS and used in this thesis are the following:

- Time of each point collected
- Longitude, latitude and height of each point collected
- Horizontal error
- Vertical error
- Number of satellites

Note that each point of the trajectory was collected nearly every second. It means that for a trajectory of 100 seconds, were collected around 100 points. To prove this time interval was appropriate some tests were made running instead of simply walking and the results, in terms of precision and accuracy appear to be the same as the previous. This aspect would guarantee good performance in real case scenarios where it is most likely to have situations where firefighters or civil protection staff may need to move quickly through the environment.

Starting from these simple data, the first idea was to plot simply the horizontal and vertical error of each point with the time in order to observe their behaviour. The horizontal error (H_err) was without any doubt the most reliable data to investigate (Figure 24). The behaviour of the Vertical error (V_err) is likely to be less prone to changes in a short time. This shows smooth changes on the slope compared to the Horizontal error which means harder to encounter peaks and drastic increases or decreases of the general error in the trajectory. More generically, a plot of this kind doesn't underline considerable variations when it should and is not deductible what the trajectory is encountering along its path, whether multipath areas or a simple tree that can cause noise on the signal (Figure 25).

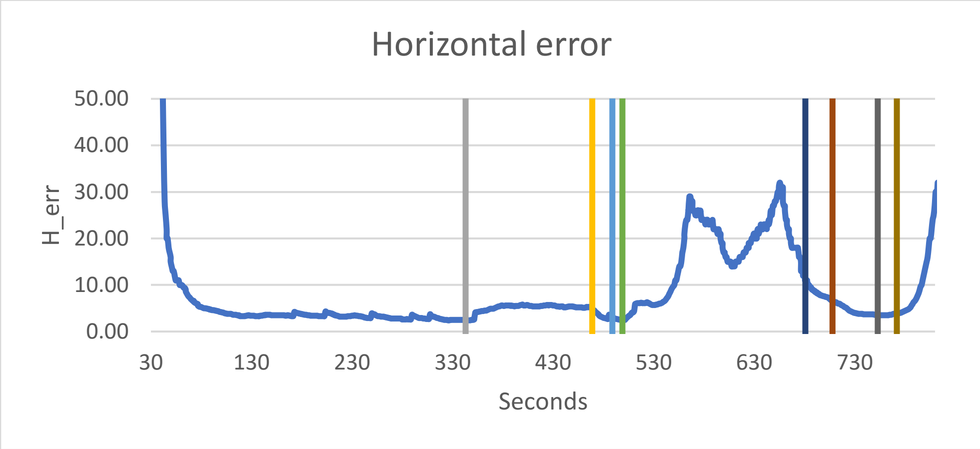


*Figure 24 GNSS trajectory horizontal error vs time*

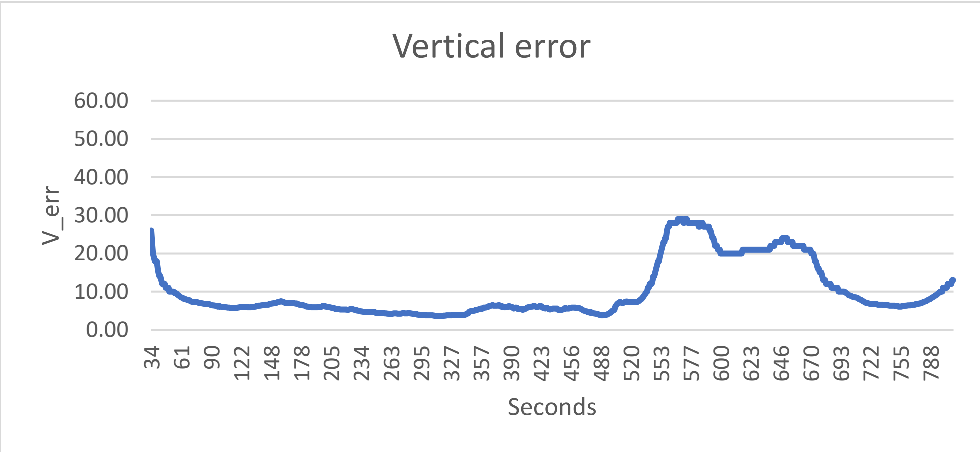


*Figure 25 GNSS trajectory vertical error vs time*

An analysis of Figure 24 is necessary to better understand the overall trajectory collected at the CTTC campus. Starting from the beginning, when the GNSS receiver is turned on, it needs a few seconds to locate itself in the world while looking for the satellites. Usually, this aspect requires a few seconds, but the error remains very high even though the coordinates are quite precise. After the stabilization of the error, the graph follows an almost flat curve setting its value around 3.5 meters. This part of the graph though, it is shown with some peaks on it. Those are due to the trees in the path, which does not interfere with the trajectory. In fact, it can be verified, by observing Figure 23 that the trajectory has no variation due to these trees, mainly because the variation in the error variation is very small: it is actually increasing around 1 meter.

The first interesting variation on this plot is around second 342, right after the grey line, where is shown an increase of the error slightly above 5 meters, settling for about 130 seconds until the orange border. All this area, checking the trajectory, shows the path near the building. As expected, the multipath area results in an increase in the error and after some seconds also the trajectory doesn't follow a straight line, as the user actually moved, but starts to move out of the real path. Must be noted that the error and the trajectory responded with some difference in time to such a situation. The horizontal error began to deteriorate immediately once the device was close to the building while the trajectory followed after some seconds. It doesn't mean that it will always act the same way. In this particular case, the horizontal error, but also the vertical one, are still admissible and the trajectory didn't change significatively. It will be different in those cases where the user of the device will go directly inside a building, as it will be shown ahead in this paragraph.

The yellow line delimitates the end of this first obstacle. As expected, there is an improvement in the horizontal error now that the receiver is in an open space and the H_err is improving and stabilizing at the same values it had before encountering the first building.

Following the trajectory, there are two new lines (blue and green) very close to each other. The blue one is at the border of the supposed influence of the next building following the criteria of 30 degrees from the top of the building as area of influence. Whilst the green line is at the moment when the user approaches the entrance of such a building. This kind of approach shows here a weakness. It is expected at this point to see a major variation in order to tell that the device is actually going inside a building and is not reliable anymore. This should be the moment to switch to the visual inertial odometry. Unfortunately, the horizontal has only a small increase, comparable to the one observed earlier where the multipath occurred. Since an error of 5 meters is still acceptable for such an application, using an algorithm with such settings, will not guarantee a continued and realistic trajectory. The H_err values are increasing drastically when the device is inside the building after several seconds. So, using a simple approach such as the H_err only or its standard deviation in a chosen area will attach the VIO trajectory only when it is too late. Of course, this will result in a wrong positioning of such part of the trajectory, losing important information on the indoor location of the user carrying the device. Furthermore, the same problem will occur when the user gets outside the building. The dark blue line in Figure 24 describes the moments the user gets out the building. As soon as the GNSS signal is reliable must be used instead of the VIO trajectory Although, at this point is clear that the H_err is not stable yet. This represents another problem since is being used a shifted VIO trajectory, which has to be considered wrong. The GNSS trajectory needs to be used as soon as possible once the signal is good enough. These limits still need to be understand and set. Observing the plot in Figure 24 is clear that this graph alone is not enough.

### 3.1.3. First tests on the trajectory

To overcome this issue, some considerations were made on the possible trend using all the possible information given by the GNSS. Despite plotting H_err against time gives a good overview of the overall state of the trajectory, it was already explained why this information is not enough alone.

At first were created clustered areas of interest and calculated for each the mean values and the standard deviation, for horizontal, vertical errors and number of satellites (Figure 26). This knowledge was useful for the definition of outdoor, indoor and transition areas, but still the issue wasn't solved. This cluster doesn't help in any way to find the starting point of each area. So far the beginning and end of each area were set manually. For this reason, the calculation of the so called Z score or Standard score ((Wikipedia contributors 2022b) seemed to be a possible indicator. In statistics, the standard score is the number of standard deviations by which the value of a raw score is above or below the mean value of what is being observed or measured. Raw scores above the mean have positive standard scores, while those below the mean have negative standard scores. It is calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. This process of converting a raw score into a standard score is called standardizing or normalizing. Unfortunately, as shown in Figure 27 the result cannot be used for detection of changes in the navigated space because there are not clear information in the suspected areas. On the contrary, there are higher peaks in the clean outdoor area which makes impossible any identification in space changes.

| Area 1 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 1507.05 | 766.95 | 3.38 |
| dev st | 492.50 | 80.98 | 0.49 |

| Area 2 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 5.11 | 6.86 | 12.74 |
| dev st | 5.09 | 2.88 | 3.53 |

| Area 3 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 2.72 | 3.94 | 19.79 |
| dev st | 0.32 | 0.23 | 0.52 |

| Area 4 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 5.16 | 5.71 | 19.77 |
| dev st | 0.58 | 0.42 | 0.95 |

| Area 5 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 2.95 | 4.47 | 21.00 |
| dev st | 0.28 | 0.39 | 0.69 |

| Area 6 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 2.74 | 3.99 | 22.00 |
| dev st | 0.34 | 0.19 | 0.00 |

| Area 7 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 17.68 | 19.66 | 17.04 |
| dev st | 7.28 | 6.83 | 2.66 |

| Area 8 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 7.56 | 10.31 | 20.08 |
| dev st | 1.01 | 1.32 | 0.27 |

| Area 9 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 4.03 | 6.81 | 20.98 |
| dev st | 0.63 | 0.54 | 0.65 |

| Area 10 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 3.84 | 6.34 | 21.00 |
| dev st | 0.26 | 0.18 | 0.32 |

| Area 11 | H_err | V_err | Number of sats |
|---|---|---|---|
| mean | 17.90 | 9.15 | 15.36 |
| dev st | 11.52 | 1.93 | 3.34 |

*Figure 26 Mean and standard deviation calculated by areas of the trajectory in relation to H_err V_err and number of satellites*

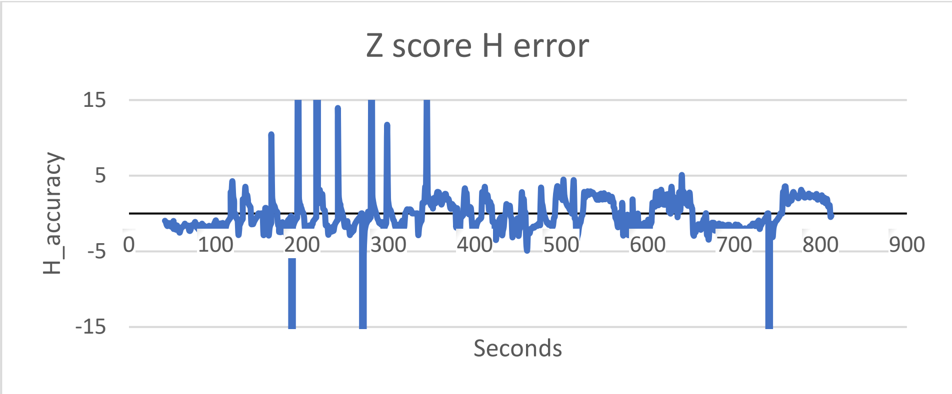


*Figure 27 Z score plot*

The next attempt was to consider the slope of the horizontal error in the last 10, 5 and 3 seconds. The slope calculated in a time frame of a few seconds would show where there is a drastic increase or decrease of the error and it will be smoother if calculated on a larger amount of time. Examples of 10, 5 and 3 seconds are shown in the next Figure 28 Figure 29 Figure 30, plotting both horizontal and vertical error.

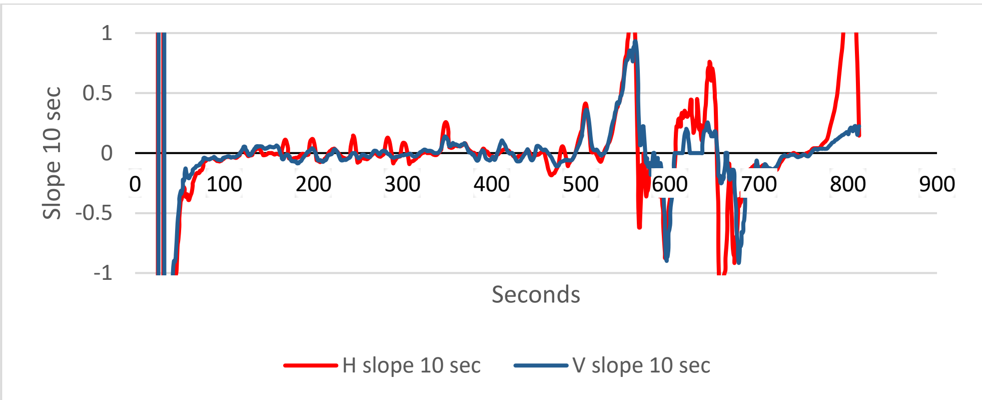


*Figure 28 Slope calculated in a timewindow of 10 sec*

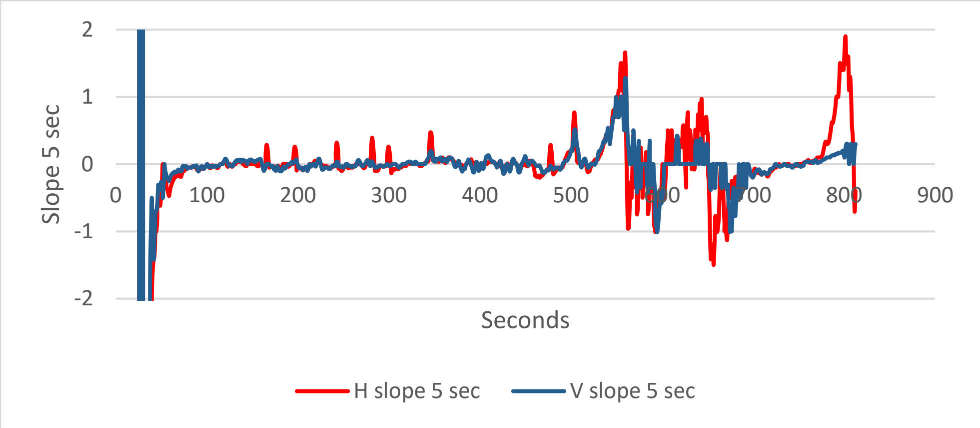


*Figure 29 Slope calculated in a timewindow of 5 sec*

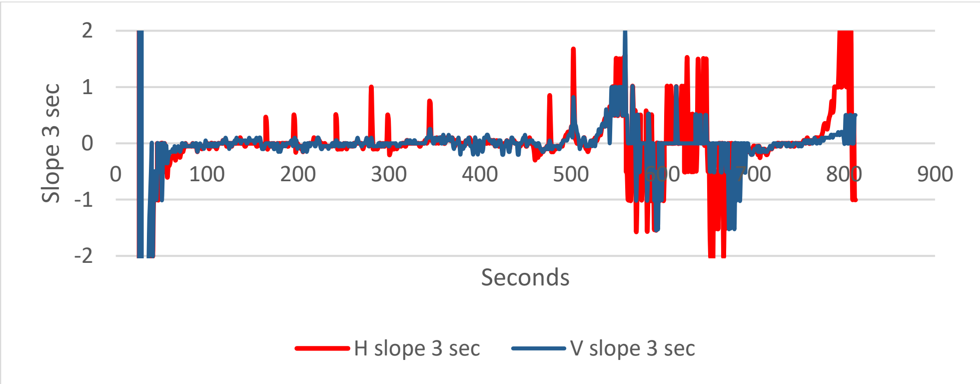


*Figure 30 Slope calculated in a timewindow of 3 sec*

The red peaks that emerge in each graph are signs of an obstacle encountered in the environment. This is promising since drastic changes are well visualized especially in the graphs of 5 and 3 seconds. In fact, the graph taken into account cannot be the one calculated over 10 seconds where these peaks are not prominent enough.

Looking for the rule to set when the system must change from GNSS to VIO, the plot of the slope at 3 seconds was chosen as it has the sharpest peaks. As observed, the first peaks encountered are due to the trees and they have a value of around 0.5. The rule must be set in order to exclude these peaks and to keep those in the proximity of a building or when going inside. Since small obstacles, such as a single tree, influence the horizontal and vertical errors only for a small amount of time, the rule will take into account the time necessary to avoid them. For this reason, just setting horizontal or vertical slopes higher than 0,5 will not be enough to understand whether the GNSS is approaching an indoor area or a small obstacle that will disappear in a moment. Of course, this is still not trustworthy for switching to VIO at that moment. It can be understood from this the necessity of a transition area. In fact, setting the previous rule true in an interval of [0;2] seconds and an horizontal error higher than 6 meters will guarantee of being in an area that is not going to deteriorate the signal furthermore. Whereas, setting a value of H or V slope higher than 1 in an interval of [2;10] seconds combined with an accuracy higher than 10 meters, will ensure the device is moving into an indoor area, and not only moving through a transition area. If these rules are not satisfied it means that the user is still in an open outdoor area.

However, it is very hard to determine the transition area when coming from indoor. In this situation, when the slope is lower than -1, which means that the horizontal error is improving quickly and the general error is already lower than 6 meters, it can be considered that the outdoor areas have been reached and the GNSS signal is reliable again. The explained algorithm is summarised in Figure 31.

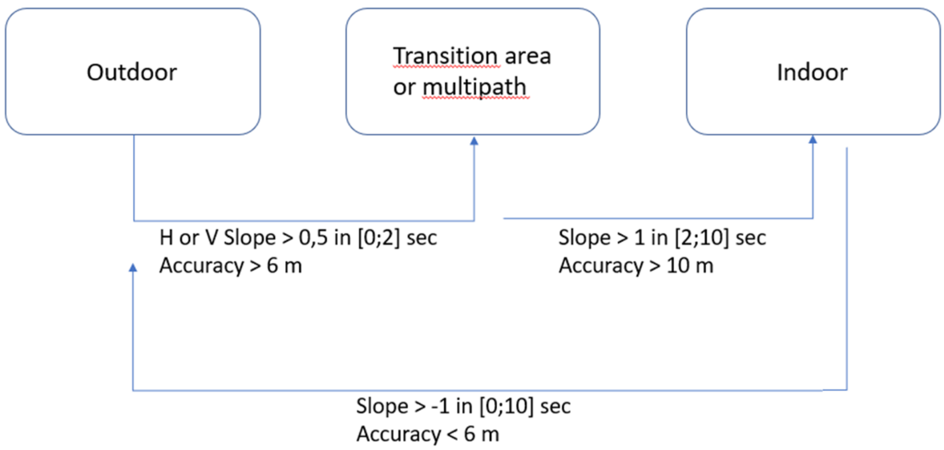



*Figure 31 Algorithm using the calculated slope*

Once the rules were defined, they were set in QGIS to visualise how the trajectory would appear. The green colour should show the moment of switching from GNSS to VIO. It is visible at the beginning of each building encountered. Analysing this results, it may seem appropriate to switch to VIO in all the cases suggested. However, there is not a great improvement compared with the previous results obtained from Angelats team. This approach may seem appropriate at the first building but there is a notable delay especially when entering the last building in the figure below (Figure 32). This is why another approach must be found and followed. The algorithm of identification of outdoor, indoor and transitional areas must work in any of these spaces. This led to investigate using a new approach. In the next paragraph will be explained how the Mann Kendall test works and why it turned out to give the best results.

*Figure 32 Plotted result of the algorithm based on the slope*

### 3.1.4. Mann-Kendall test

The Mann-Kendall statistical test (MK Test) is a non-parametric test (Kendall 1948; Mann 1945) used to assess the significance of trends (Yue, Pilon, and Cavadias 2002). Mainly used for hydro-meteorological time series, we believe that such kind of statistical test, even though there are no cases in the literature where this test has been used for similar scopes, can be appropriate for such dataset as they are thought to be suitable for non-normally distributed data.

Although similar tests, such as Pettitt's test (Pettitt 1979) were performed in the same dataset, they did not produce any significant result, for this reason will not be discussed in this thesis.

The following assumptions underlie the MK test:

- When no trend is present, the measurements (observations or data) obtained over time are independent and identically distributed. The assumption of independence means that the observations are not serially correlated over time.
- The observations obtained over time are representative of the true conditions at sampling times.
- The sample collection, handling, and measurement methods provide unbiased and representative observations of the underlying populations over time.

(https://vsp.pnnl.gov/help/index.htm#vsample/design_trend_mann_kendall.htm)

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn\,(X_j - X_k)$$

$$sgn(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & if\ x = 0 \\ -1 & if\ x < 0 \end{cases}$$

The mean of S is $E[S] = 0$ and the variance is

$$\sigma^2 = \left\{ n(n-1)(2n+5) - \sum_{j-1}^{p} t_j(t_j - 1)(2t_j + 5) \right\} / 18$$

where p is the number of the tied groups in the data set and $t_j$ is the number of data points in the $j_{th}$ tied group. The statistic S is approximately normal distributed provided that the following Z-transformation is employed:

$$Z = \begin{cases} \frac{S-1}{\sigma} & if\ S > 0 \\ 0 & if\ S = 0 \\ \frac{S+1}{\sigma} & if\ S > 0 \end{cases}$$

The calculation of the MK test would not make sense if applied to all the dataset at once. A single value calculated on the whole dataset cannot tell where or when the trajectory is being influenced by the surrounding area. Moreover, it cannot be used in small areas because this kind of test must work in real-time or following the succession of data acquired along the path. The idea was to apply the MK test on the last 10, 5 and 3 seconds acquired. The 10 seconds test gave the best results and is plotted in the trajectory in Figure 33. The legend shows how the points are coloured from a range of -4 to +4. The values between -2 and 2 show small variation or no variation on the trend if in the proximity of 0.

*Figure 33 Mann Kendall test calculated on a timewindow of 10 sec and plotted on the trajectory*

This graph is not immediately clear, but some areas are showing some recognizable trends as expected. For example the deep blue area while approaching all buildings and the red points when getting out or when the trajectory is improving. In other words, this plot needs to be refined.

At this stage, we applied a similar rule as the previous one used for the slope to the algorithm. Detecting the transition area is still of great importance, but the GNSS + VIO system can only switch between the two trajectories, there is no need for a rule between outdoor and indoor but only one that is comprehensive of indoor and transitional areas with great effects on the GNSS data. The new rules applied to such trajectory are applied as indicated in Figure 34.

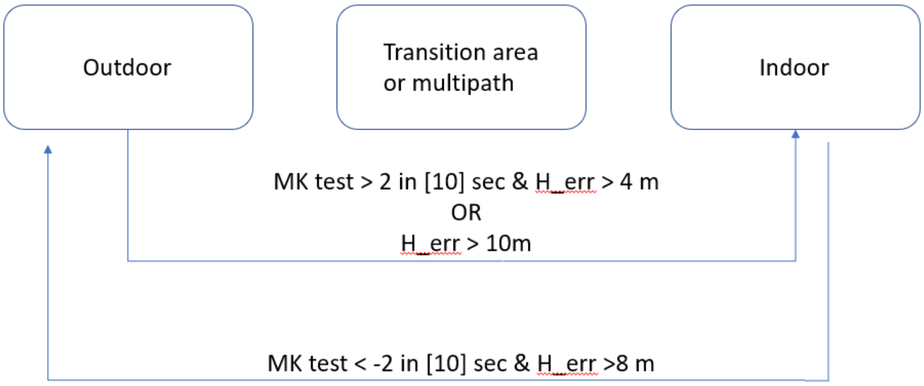



*Figure 34 Algorithm using the MK test*

This resulted into a new coloured trajectory. Indoor is recognisable before getting inside a building and outdoor is found in less than 10 seconds when coming from indoor. Also, the transition areas are detected when big multipath affects the receiver and all small disturbs are now disappeared. It is a significant improvement that can change the time of switching from one device to another only using the last 10 seconds of data collected.


*Figure 35 MK test applied to the trajectory of CTTC campus. Green points are good GNSS areas, orange are transitional areas from outdoor to indoor and red are from indoor to outdoor*

Observing the second building from the bottom left, the user spent some time inside of it. Inside buildings there can be an improvement of the GNSS when the user passes near a window for example. For this reason can be seen some red areas indoor the building, suggesting to switch to GNSS but this does not represent a problem because to switch from VIO to GNSS again it must be waited until the GNSS location is classified as green (outdoor).

At this point, the VIO trajectory was positioned at the moment suggested by this set of rules depending from the MK test results. The output is shown below in Figure 36. In blue the final combined trajectory GNSS +VIO and in red is shown the old positioning from the VIO. It is clearly visible, especially when it goes indoor that the trajectory is much more accurate. Also, when going back outdoor, the trajectory moves quickly to the GNSS one with much higher accuracy.


*Figure 36 Comparison of old (red) trajectory with new one (blue)*

Using this approach, it is proved that in an urban context the combined GNSS + VIO system will be significatively improved. But how it will behave in different scenarios? To have a reliable system the same approach must be tested in several areas, considering all the cases mentioned before. In other words, any scenario can be a place where emergency staff might work. Will be tested then areas with narrow streets, dense canopy, tunnels etc. As many different scenarios from open outdoor to deep indoor and all that's in the between.

All the tests were conducted in Spain. Below are presented four more sites where few tests were carried out per each site. All the tests are very similar in terms of trajectory and the final result of the new proposed approach give significant improvement to the seamless outdoor indoor tracking system. Figure 37 Figure 38 Figure 39 and Figure 40 are showing the results the new approach using MK test.

*Figure 37 MK test applied to the trajectory in Garraf town*

*Figure 38 MK test applied to the trajectory of Falconera near Garraf town*

*Figure 39 MK test applied to the trajectory of Collsuspina*

*Figure 40 MK test applied to the trajectory of CatuAV BCN Drone Center*

### 3.1.4. Results

Comparing the old combined trajectory with the one from the proposed approach based on the MK test, it is visible a great change in the positioning of the VIO trajectory. The distance between the two is about 2 to 8 meters. Based on the knowledge of the interiors of the buildings it is possible to tell that the new method is giving as a result a trajectory very close to reality.

The following figures shows the result indoor in CatuAV, falconera and Garraf.

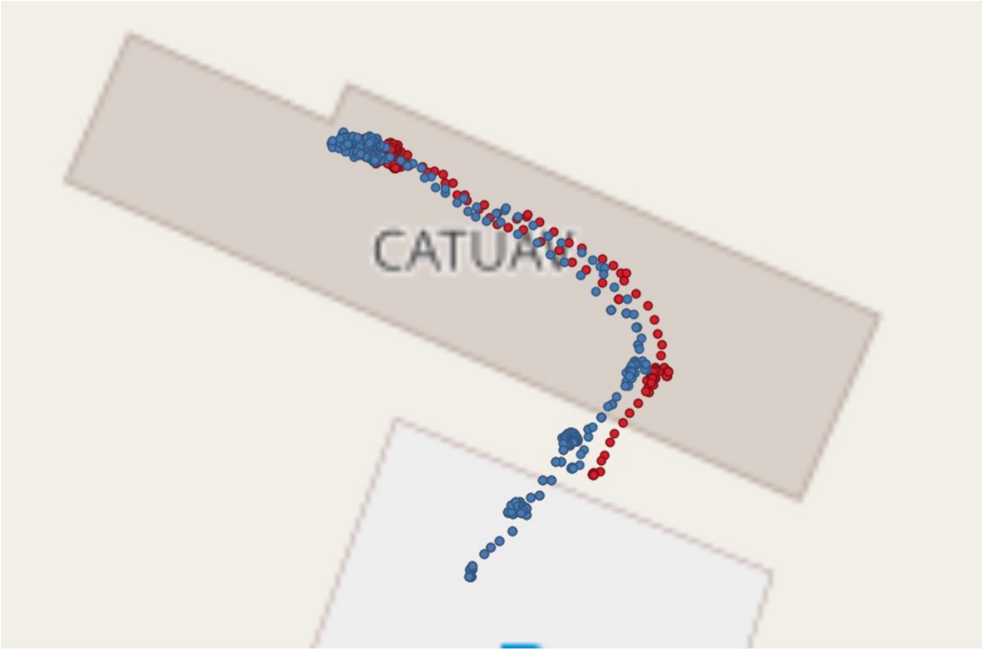


*Figure 41 Comparision of old (red) trajectory with new one (blue) in indoor CatuAV trajectory*

*Figure 42 Comparision of old (blue) trajectory with new one (orange) in Falconera*

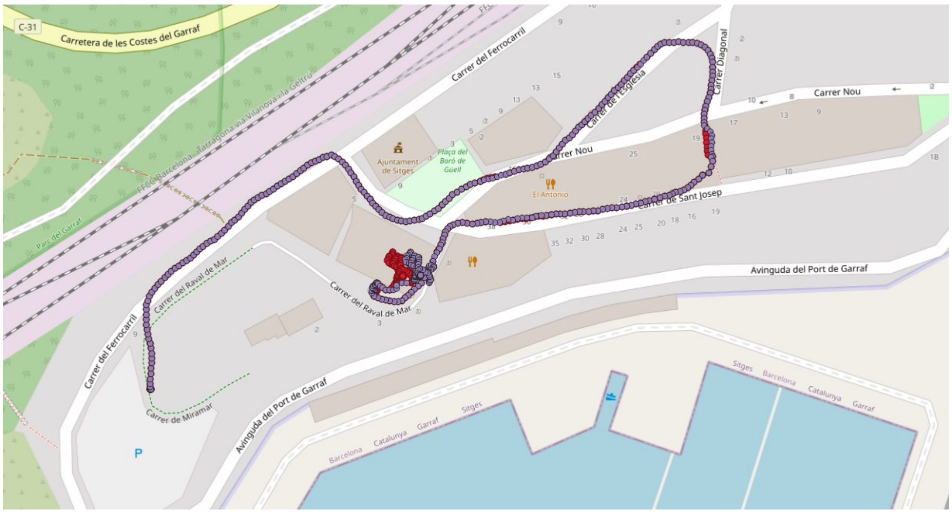



*Figure 43 Comparision of old (purple) trajectory superimposed with new one (red) in Garraf trajectory*

To stress this method in a more sever way, two different trajectories were covered having as finish line the indoor of the same building. The trajectories were also shown previously in Figure 37 and Figure 38. One is in Faclonera and the other is in Garraf town. Since the areas covered for Falconera is quite big, below will be shown the ending parts of the two trajectories with both, old and new location of the VIO.

Knowing that the indoor path was the same for the ending of the Falconera and Garraf trajectories, it is possible to compare the ending part of each trajectory. In Figure 44 there is quite a big distance between the twos while using the old approach based on the standard deviation. While using the new approach is the two trajectories are very close to each other as it can be verified in Figure 45. There is still a little error even using the new approach, but for the requirements set at the begging it is more than acceptable. In fact, it can be said that this approach guarantees to know which room is occupied by the user.

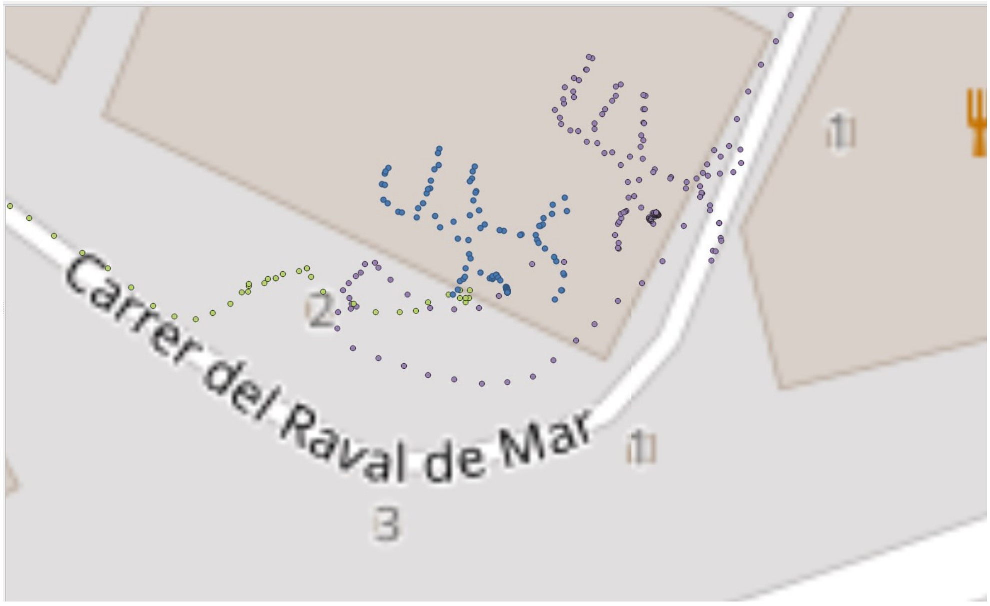


*Figure 44 Last parts of Falconera and Garraf trajectories ending indoor the same building before the new algorithm proposed*

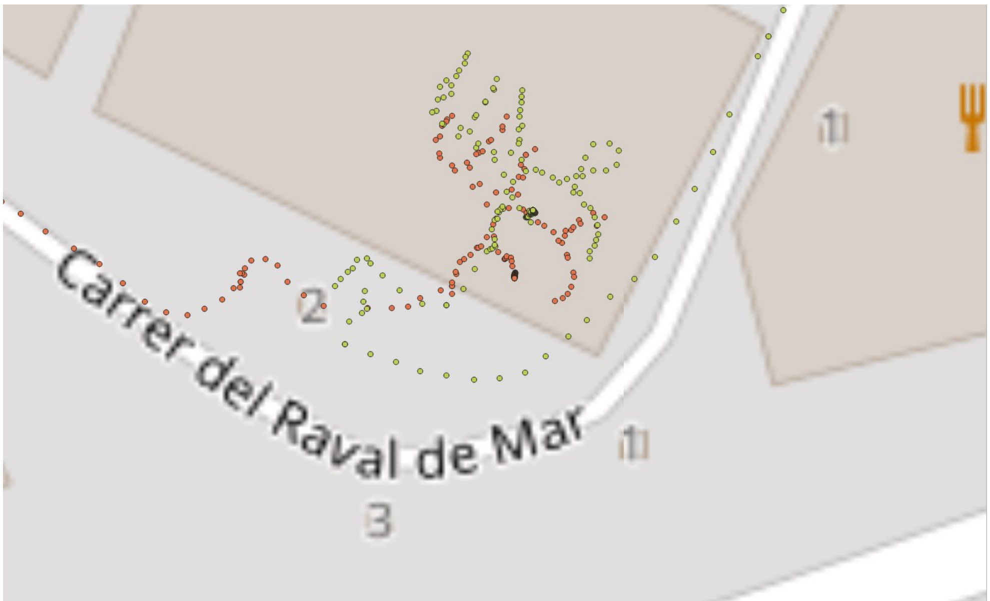


*Figure 45 Last parts of Falconera and Garraf trajectories ending indoor the same building after the new algorithm proposed*

# 3.2. Indoor tracking

Studying the indoor localisation of people and producing a set of information from such data can be done in several ways. The increasing use of IoT in indoor spaces and the improvement in technological field combined with accessible prices are bringing IoT devices into every indoor space, from our homes to big and strategic buildings such as airports, universities, museums or shopping malls. This thesis will concentrate on the last kind of building.

From a technical point of view, indoor spaces are not very friendly when it comes to tracking people. For this reason through the last decade were used many techniques for indoor localization, such as Angle of arrival (AoA), time of flight (ToF), return time of flight (RToF) and received signal strength (RSS). These techniques are based on different technologies: WiFi, radio frequency identification device (RFID), ultra-wideband (UWB,), Bluetooth and others (Zafari, Gkelias, and Leung 2019). The wide use of different technologies is because they satisfy different purposes, even when it comes to what can be seen as the same aim: tracking people indoor. However, each of these technologies has its own difficulties to deal with.

In the previous paragraphs was demonstrated that it is possible to obtain high accuracy trajectories of people in indoor areas. Although, the used technique requires that the user is provided with tools that need to carry through the whole time of the investigation. This approach is perfect for the described situation encountered before. The user can be equipped with such tools and only those users are being monitored. This kind of approach is known as an active system. It means that somehow the user must interact with the installed tracking system, for instance, by carrying a device, such as the previous case, or using an application from their smartphone which requires to be captured from the installed system. It can be the case of using Bluetooth signal (Osaba et al. 2018) installing Beacons in the interested indoor (but also outdoor) space. The main types of active and passive systems used in indoor localisation are described in (Deak, Curran, and Condell 2012).

When the necessity is to cover all the people in a specific indoor environment, it is not possible to rely on active systems. In airports and shopping malls, there are always new people coming to visit the space or passing there only a short amount of time. This implies that it is impossible to provide all the users of a space to carry a specific device or download an application for smartphones which would track them. Even if the application offers a higher grade of security and advantages or discounts in the shops located in that space, it would be challenging to get a representative percentage of people downloading and using it. This inevitably leads to passive systems. Unfortunately, passive systems are less accurate than active ones, but they allow us to have almost the total count and tracking of the people living that space. This is thanks to the fact that near 100% of people own a smartphone from which it is possible to detect a signal.

WiFi localisation is the preferred solution since such service is becoming ubiquitous. It offers two options for tracking purposes. The first one is triangulation but there are not many real applications for such technique because the errors in indoor areas would be too big generating not usable data. Although there are promising research works in literature using multipath triangulation (Soltanaghaei, Kalyanaraman, and Whitehouse 2018) with decimetre-level accuracy it must be said that it works well for small buildings such as offices and is not proven for major buildings such as malls or airports with hundred or thousand people at the same time. Other approaches are going towards machine learning (Salamah et al. 2016) and deep learning (Wang et al. 2017) but today the most used approach is still the Received Signal Strength Indicator (RSSI) fingerprinting (Feng et al. 2012; Keshka et al. 2020; Liu et al. 2012; Wu et al. 2013; Yang, Wu, and Liu 2012; Yiu et al. 2017).

## 3.2.1. WiFi indoor tracking

The approach used in this thesis is similar to the RSSI fingerprinting just mentioned from the literature, but there are still a few issues to solve when it comes to signal strength. The signal is not reliable in many situations. As proved in (Rong-Hou et al. 2008) RSSI is strongly affected by the environment and obstacles. If the tracking would be based only on this aspect, it will cause the signal to describe a not truthful situation. The strongest signal is not always the nearest to the Access Point (AP). Supposing that an AP is placed on the ceiling, even positioning a device, such as a smartphone directly under the AP location it may show a lower signal strength than moving a bit further. Multipath is very common in indoor areas and it makes it hard to get high accuracy in the tracking of people.

The chosen methodology will not be described in detail since the rights on the technological development are property of Wifiproject srl, who financed the work and part of this research thesis. The tracking system is based on the fingerprinting and additionally inspired by AI, training the model to better understand where the devices are located.

The idea behind the RSS fingerprint in indoor environments is to build radio maps and then setting the localization and optimisation problem(Zegeye et al. 2016). Measurements of RSS are then collected and by cross referencing the signal strength against the pre-existing record, the system can assume the device positioning. Fingerprinting consists of two main phases: the offline training phase and the online matching phase

The offline training phase trains the algorithm to learn the RSS at various points. First is defined the indoor floor plan and divided in grids where the users will be placed. The grid must be trained, this has been done by walking around in the area of interest and saving the signal strength for each reference point. The algorithm is where our innovation stands. It will store the data in a database and in the online matching phase, will process the information in order to obtain a good accuracy level while filtering all the devices not attributable to humans. Figure 46 summarises the described methodology. During the online phase the Access points will send the target signal, visitors of the centre will detect the signal through their device and the AP will record the signal strength and upload the fingerprints to the database.

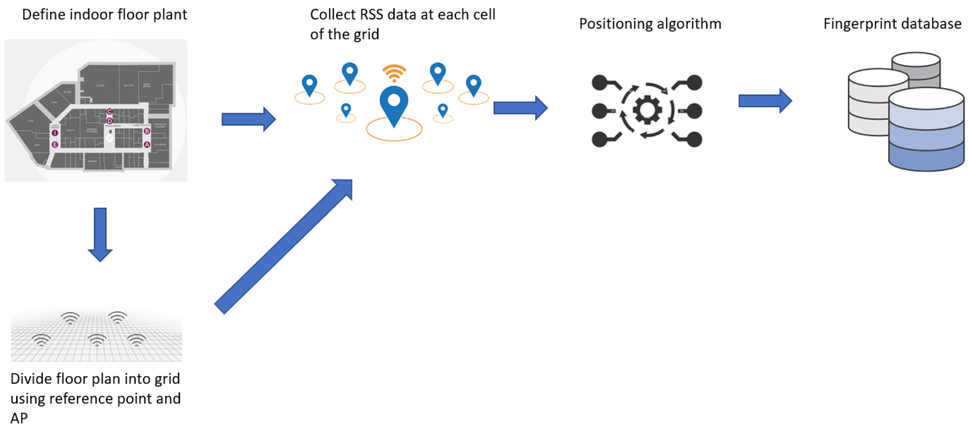



*Figure 46 RSSI workflow*

As in the previous paragraph the aim was not to achieve a high grade of accuracy, but to locate the users of the space and track them through their movement within an error of a few meters. A first test was carried out in a shopping mall by considering only a part of the entire building but representative of the whole. The data collected in this phase will refer to year 2020. Using 8 Cambium cn pilot e600 Access Points it was covered an area of 1600 meter square which was divided into 30 cells. It was possible to increase the number of cells, but for the purposes of this research (marketing and emergency management), this number was considered appropriate. The people detected will be counted within one of these cells. Since the WiFi signal extends beyond the 1600 sqm area concerned inside the shopping center, all the people detected outside this study area will be placed in the extremity cells, filling them with people who in reality are not present in this area. This problem would be solved by covering the entire shopping center with the tracking system. As this was not possible at this stage of the study, these cells will be disregarded.

The system structured with these specific Access Points allows to detect all devices which has WiFi switched on; it is not necessary to have them connected to the WiFi network. The system will detect also those devices connected to their own data network as long as the WiFi is on. The devices on stand-by will be also detected. Since the WiFi system must be able to recognize the devices and through them visitors to the center as they move, the algorithm must be able to read a unique name from each device. This unique name is the MAC address. It must be clarified that this data is not linked to a person's private information. Moreover, modern smartphones can change their MAC address, which is read by the WiFi system, to ensure a greater level of security and privacy. This is the case of randomized MAC address. In order to guarantee the trajectory travelled by the visitor, the MAC address detected must always be always the same otherwise the algorithm cannot recognize the new MAC address if it changes suddenly. However, the randomized MAC address is not representing a problem because it updates every time the device is restarted or that the smartphone's WiFi is turned off and on again. It is very unlikely during a visit to a shopping mall to restart the smartphone or the WiFi, especially when browsing using the own data and not relying on the WiFi. Of course, will be a certain number of people keeping the WiFi off and those will not be detected, but they don't affect much the overall trend because the numbers of those who keep it on are the majority. This tracking system is thought to track the position and permanence of the users in a specific area. It was done by detecting for how long a user would stay in a cell before it moves to a next one. This has made it possible to study not only the trend during the year of visitors but also to carry out analysis by areas, providing data to draw up assessments on the frequentation of each area and identify the major attraction poles at any time of the day, so as to be able to carry out market analysis as well as safety assessments in case of emergency.

For those users who will connect to the WiFi network and will agree to share information for marketing analysis will also be collected personal information such as name, age and sex. From now on, this thesis will refer to the people that decide to share information as "Users" and all the others as "Visitors" which will be tracked but will remain in complete anonymity and any personal information will be missing. The visitors represent the vast majority. Their data are very important because they describe the overall density and movements of the crowd. A first clustering and a partial analysis of the collected data were carried out to prove the system works properly and is used to propose this technology to other shopping malls and strategic buildings. Already some tracking systems have been installed successfully in different shopping malls all over Italy and are collecting data in full respect of the current regulations on privacy.

### 3.2.2. Data analysis from indoor tracking

2020 was the year the covid-19 pandemic began. During this year, marked by the effects of the pandemic and restrictions, this study emphasised the importance of managing information by also proving that the installed tracking system would work properly. The data was collected in a business intelligence platform called Tableau, which allowed to process and assess the information from January of that year. The use of Tableau helped to gather and process the data in order to make them readable and accessible to everyone. An example is shown in Figure 47 where can be seen the number of visitors over a period of seven months, starting from January 2020. It must be said that the scope of this thesis is not to produce Human Behaviour Analysis (HBA) but it wants to offer a new approach of collecting data in large and crowded indoor environments and to prove that from these data, it is possible to obtain precious information for safety management and marketing.

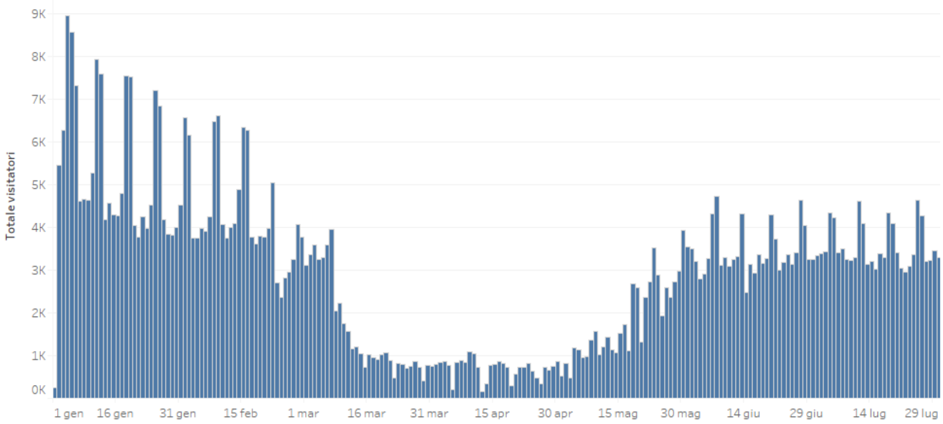



*Figure 47 People affluence in the area subject of study from January to July affected by the pandemic of COVID 19*

Covid-19 containment measures during the pandemic caused the closure of almost all shops within the shopping mall. Considering this, it is immediately clear how those peaks of almost nine thousand people recorded during weekends in the first weeks of 2020 dropped down to the few hundred recorded. The only shops opened were supermarkets, which were necessary to guarantee food supplies.

The collected data were grouped according to different Key Performance Indicators (KPIs), chosen based on the needs of the shopping mall's marketing and manager offices. Generally, the analysed data for each User and Visitor, beyond the tracking itself which is displayed through a heatmap are: daily statistics (Figure 48 and Figure 49), monthly statistics (Figure 50 and Figure 51) and global behaviour analysis (Figure 52 Figure 53 Figure 54 and Figure 55).

The following figures are only a small part of the total graphs produced that can be used for studying human behaviour. To better understand the potential of such system based on WiFi tracking, below are summarised the final graphs that have been produced from the collected information. For a clearer exposition the graphs were divided into four groups:

- Individual user details
- Daily statistics
- Monthly statistics
- Annual or biannual overview

**Individual user details**

If the detected device is associated with a user registered to the WiFi service, the following data are displayed on a daily basis:

- Most frequent cells
- Summary data of the user

- Frequency of visits over the last 15 days
- Maximum and average stay over a period of 15 days
- Preferred time of day and day of the week over a 15-day period
- Visit history over the last 15 days

**Daily statistics**

For each day, the system provides the following statistics:

- Heatmap to highlight the cells with the highest attendance
- Total number of visitors and users
- Distribution of registered users by gender
- Distribution of visitors and users by length of stay in the area
- Distribution of visitors and users by time slot
- Distribution by age group of users

**Monthly statistics**

The following statistics are provided for each month:

*Visitors*

- Monthly trend
- Trend per stay
- Distribution per day of the week
- Distribution per time slot
- Total per permanence

*Users*

- Monthly trend
- Trend per permanence
- Trend by gender
- Trend by age group
- Distribution per gender
- Distribution by age group
- Distribution by day of the week
- Distribution per time slot
- Totals by length of stay
- Totals per gender
- Totals per age group

*Sub-areas (Density and Permanence)*

- Cell map
- Distribution per Area
- Distribution by day of the week
- Distribution per time slot

**Global Statistics**

The following statistics are provided for the period January-July 2020:

*Visitors*

- Global trend
- Distribution per stay
- Distribution per day of the week
- Distribution per time slot
- Distribution per month
- Totals per permanence

*Users*

- Global trend
- Distribution per permanence
- Distribution by gender
- Trend by age group
- Distribution by gender
- Distribution by age
- Distribution by day of the week
- Distribution by time of day
- Distribution by month
- Totals by permanence
- Totals by gender
- Totals by age group

*Sub-areas (Density and Permanence)*

- Cell map
- Distribution by day of the week
- Distribution by time of day
- Distribution by month

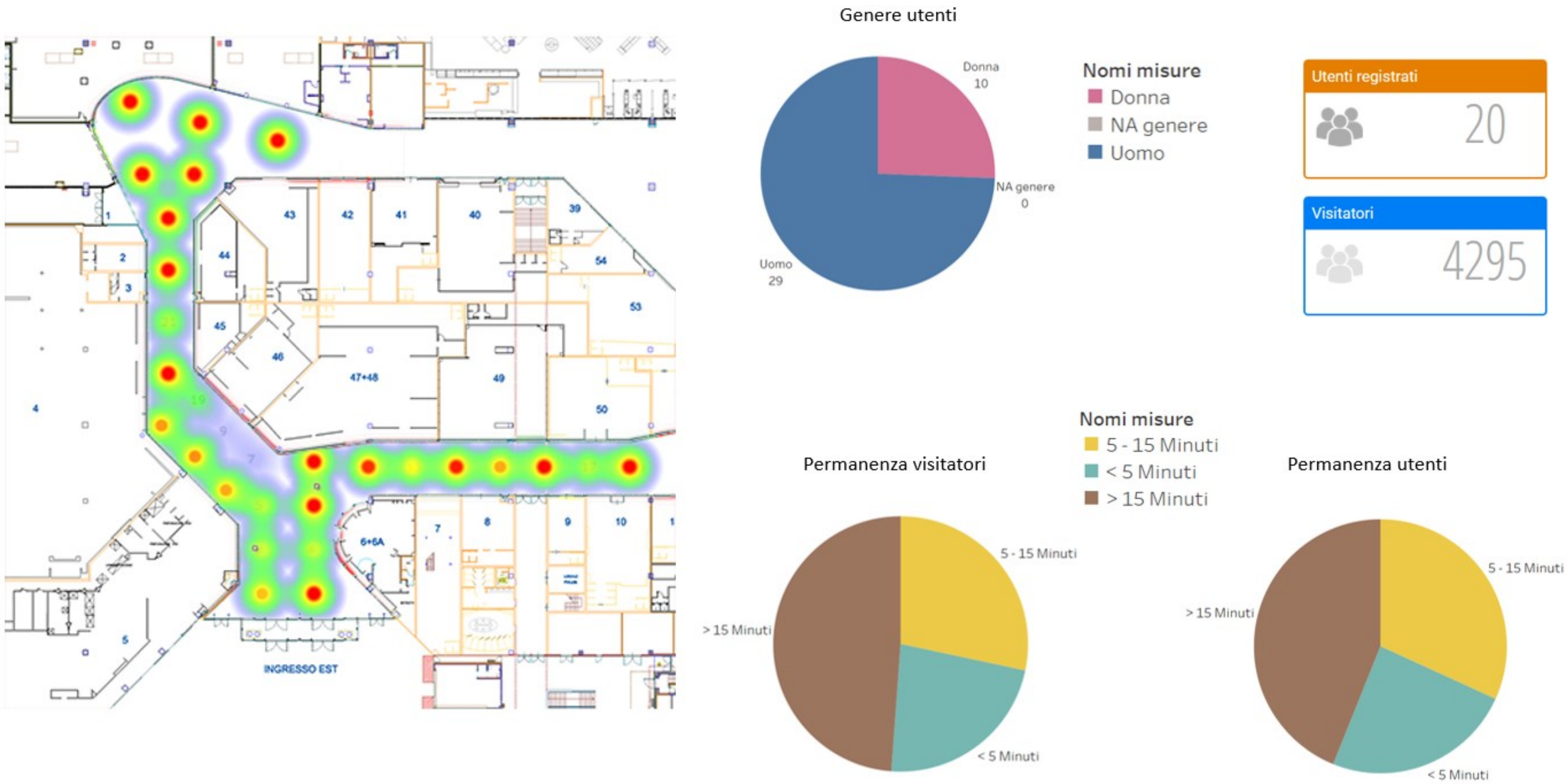



*Figure 48 Daily data distribution of users and visitors*

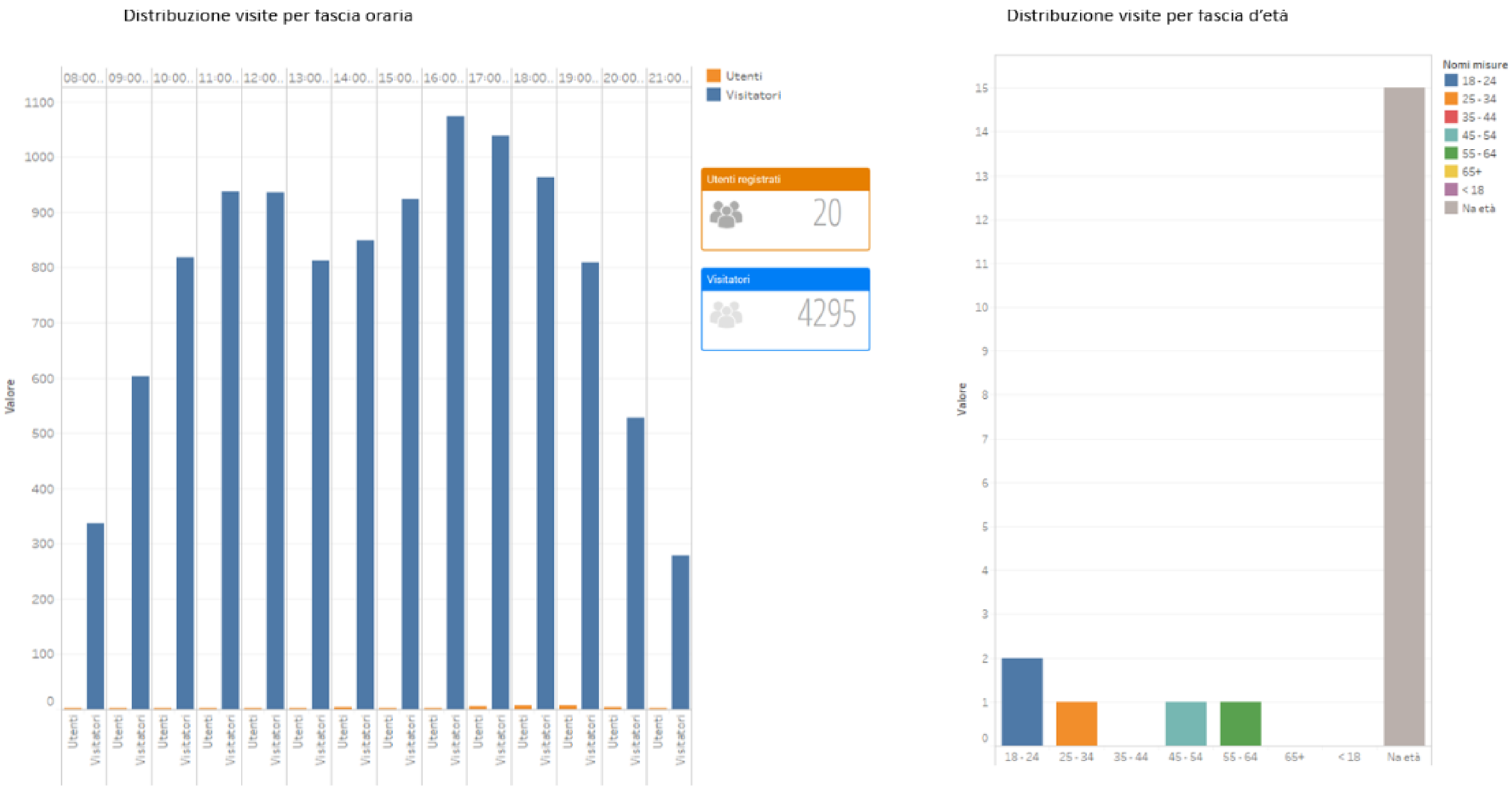



*Figure 49 Daily data distribution per hours on the left and age*

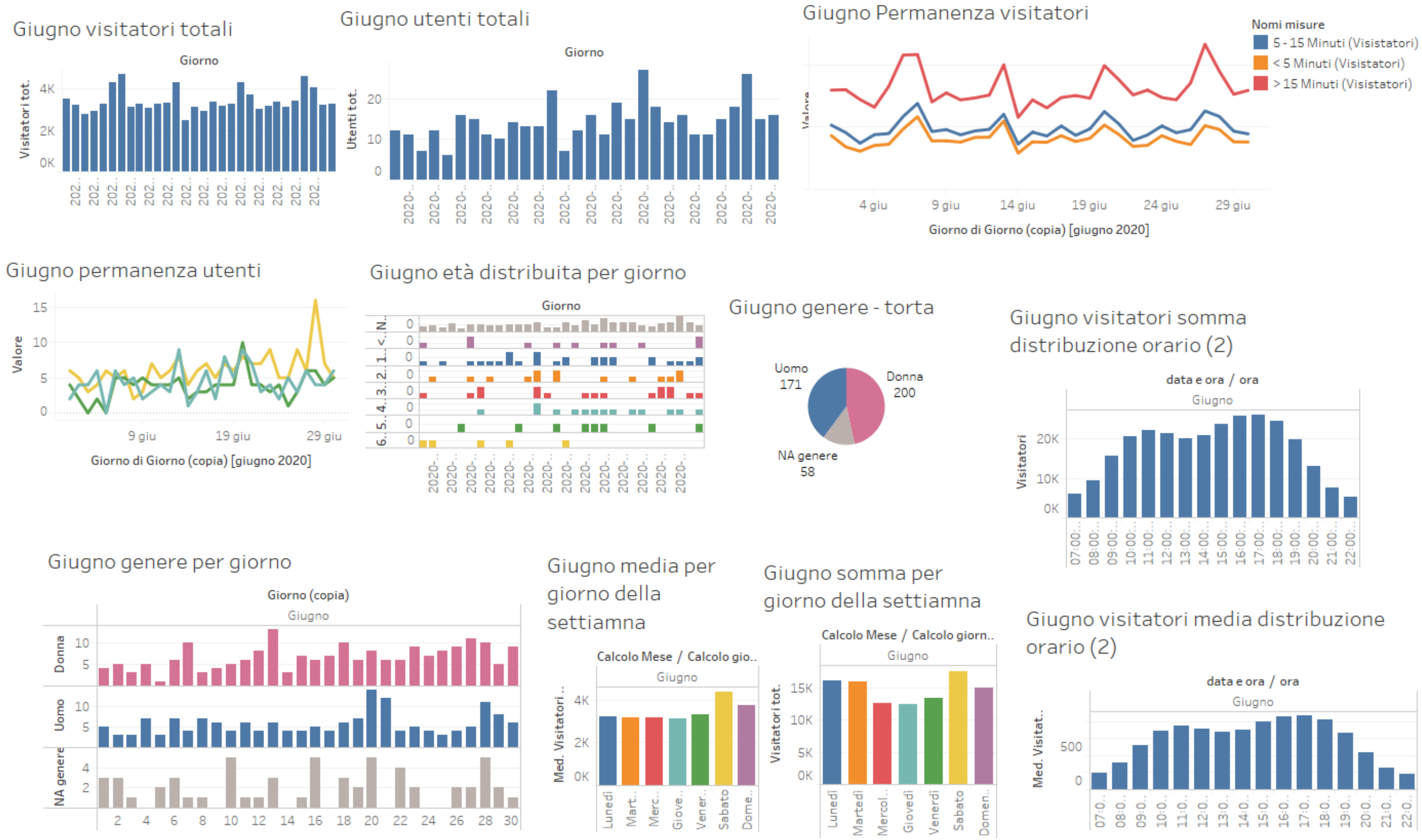



*Figure 50 Monthly statistics Users and Visitors in June 2020*


*Figure 51 Monthly statistics Users and Visitors in July 2020*

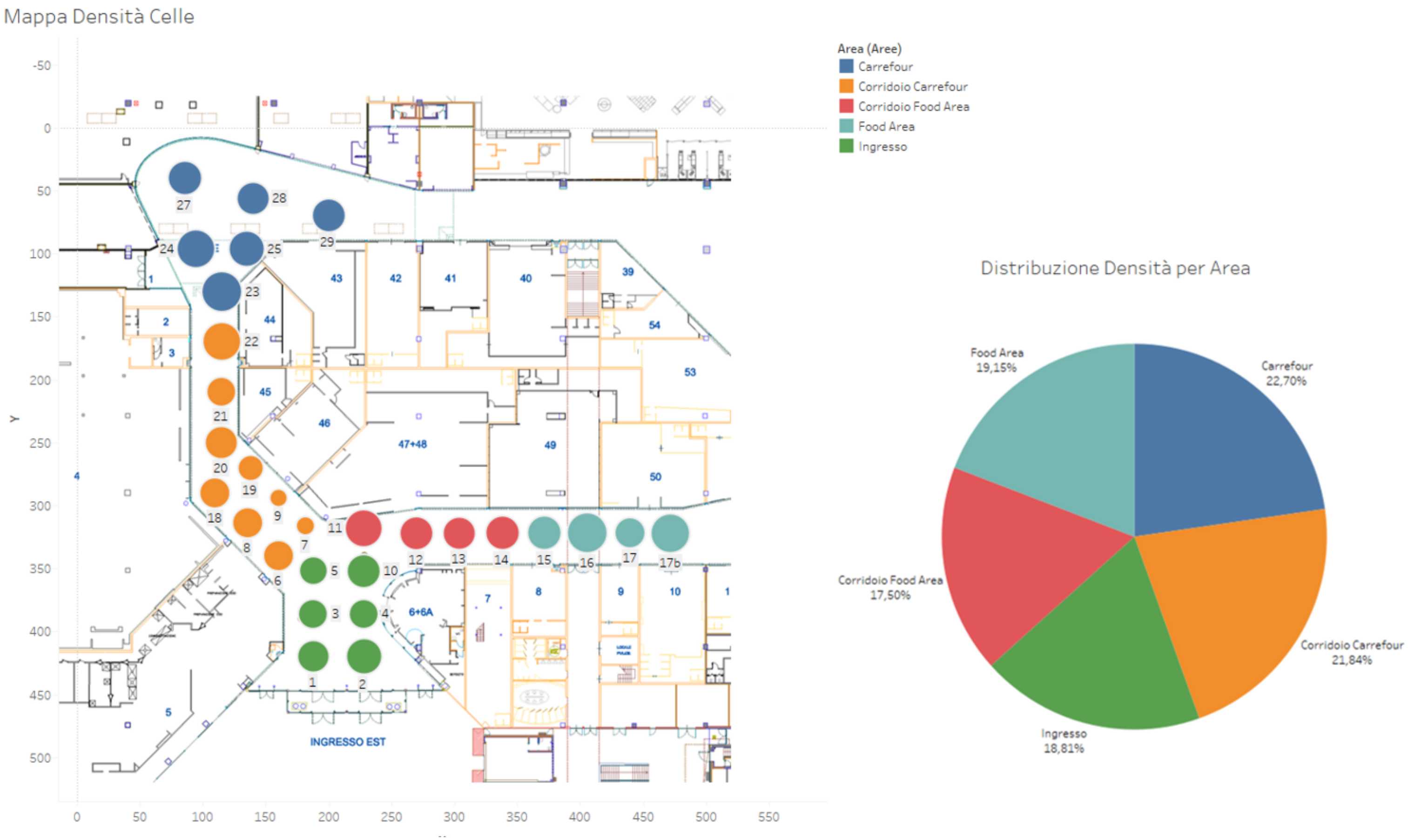



*Figure 52 Global distribution per area*

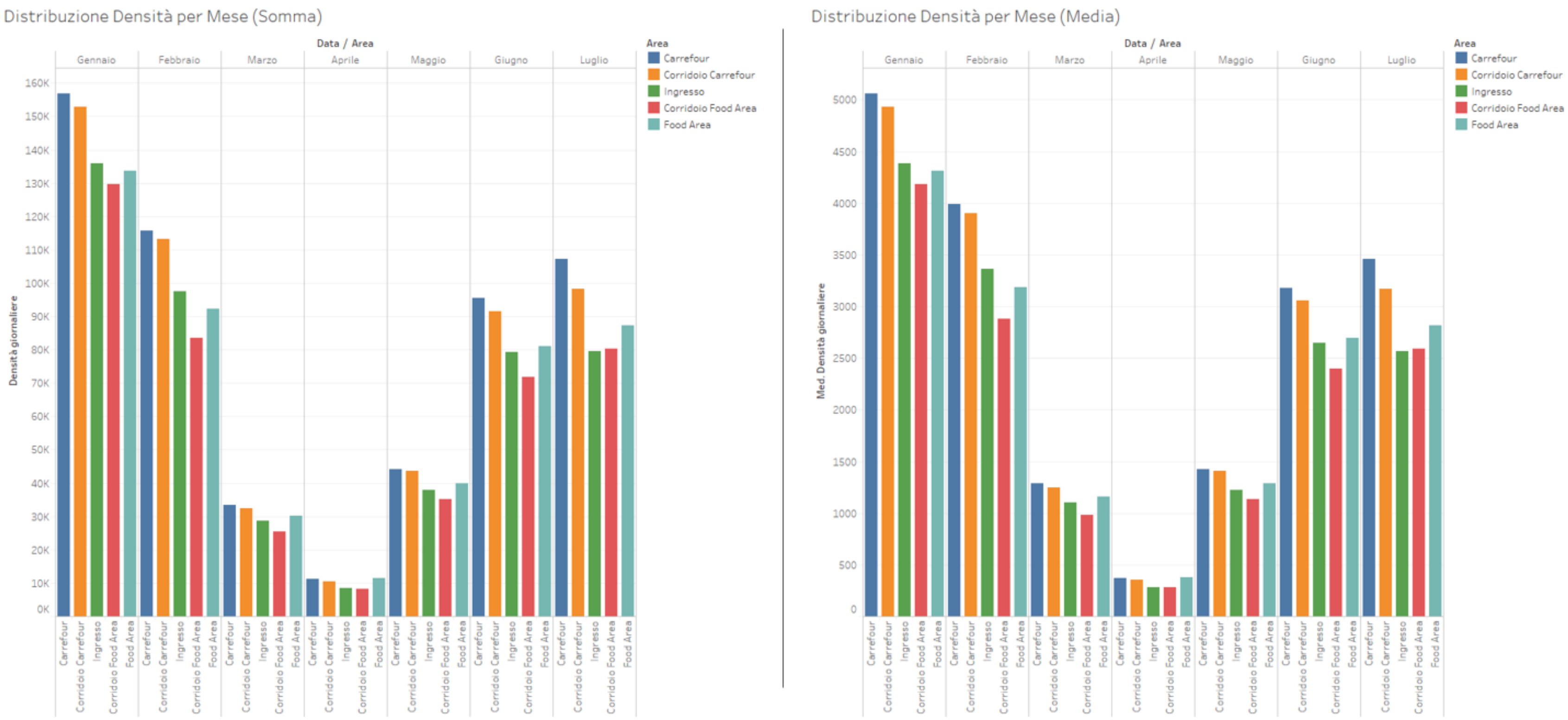



*Figure 53 Global density per month from Janary to July*

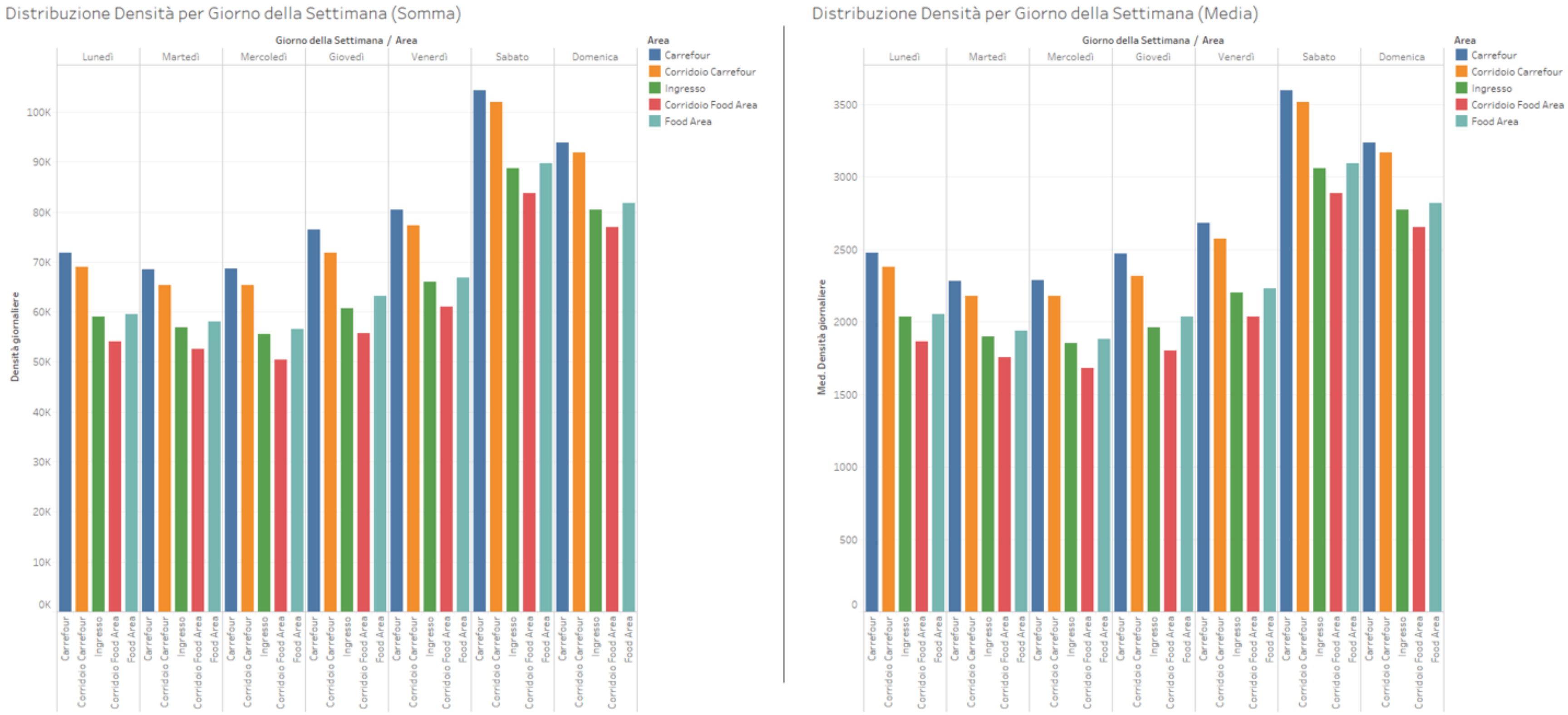



*Figure 54 Global density per day of the week from January to July*

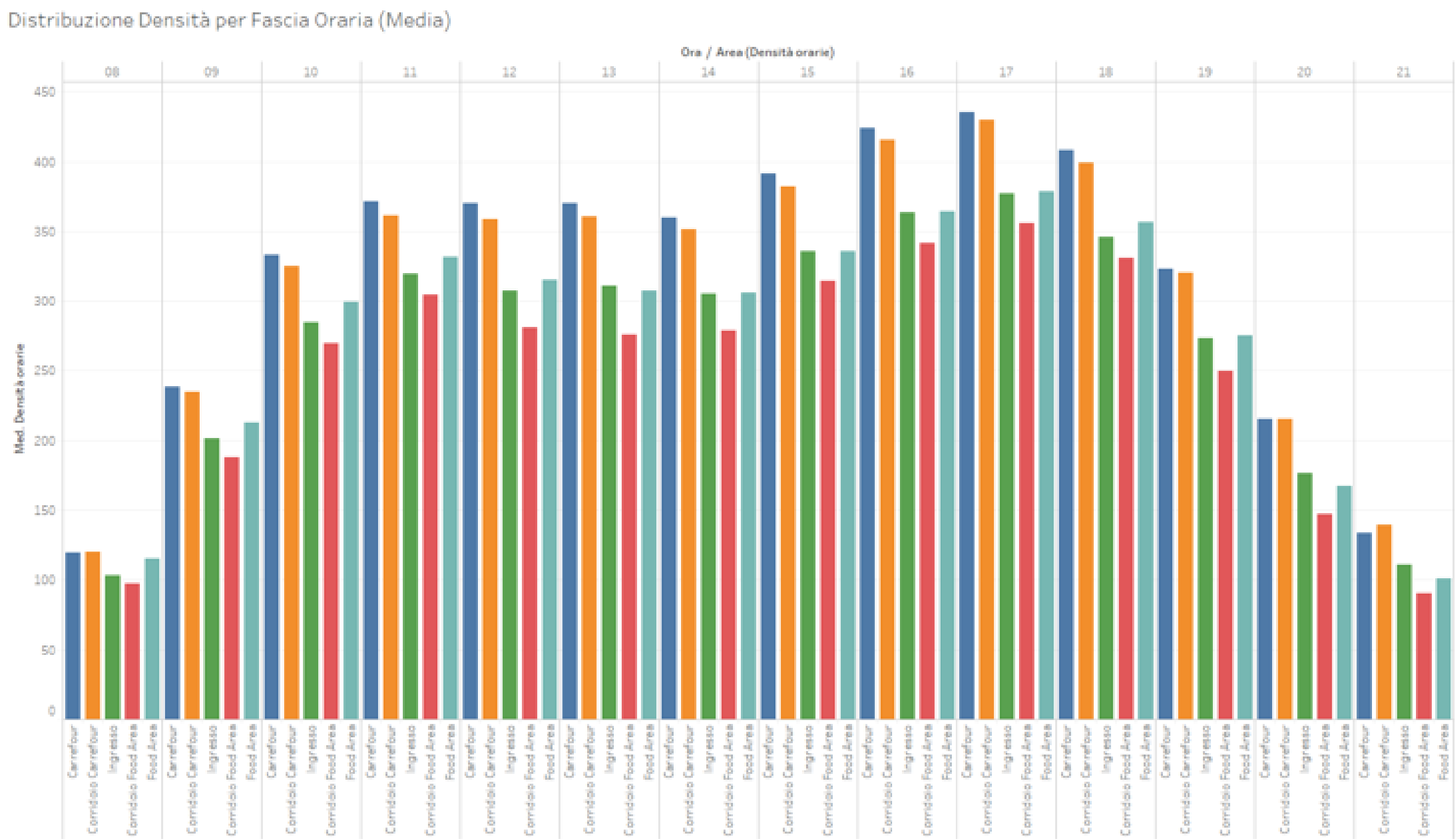



*Figure 55 Global distribution per hours from January to July*

# 4. Discussions

The result obtained on each of the three main aspects of this thesis give an answer to each of the specific research questions asked at the beginning of this thesis. Here those questions are briefly discussed.

- How can geomatics give a contribute to developing more accurate and smart models of digital twins?

As stated, digital twins are already influencing our lives in a positive manner, but there is still a lot of work to be done, especially in AEC (Architecture, Engineering and Construction) field, to reach complete models mostly at the urban scale. Geomatic is already making an important contribution by providing digital models of terrain and buildings. While technology advances, it allows faster and more accurate surveys and is moving towards monitoring what is already digitalized. One of the major aspects that need to be addressed concerns the monitoring of people, which in this thesis is addressed through people tracking outdoor, indoor and transitional areas between the two. So the interest of geomatic researchers is moving towards smart models able to manage various information in a faster way and when it's the case, in real-time. We believe those models will be the digital twins of the future and geomatics is already providing the first tools.

- How can we improve 3D GIS models, how do we represent them and how to manage data when it comes to external information?

New geomatic techniques are leading toward fast surveying. The use of mobile laser scanning (di Stefano, Chiappini, et al. 2021) and photogrammetry already produces strong and reliable data. The improvement of the 3D models comes with new tools and management. GIS tools have limited ontology and is very difficult to add new information in this environment. Therefore, the complexity of our world is not always easy to be described through only one software or information system. For this reason, many technicians must rely also on written files whenever they can't integrate information into the information system. In order to have better 3D models we need better information and management systems that can allow the input of several kinds of information. CityGML represents a good solution in data storage since it provides a larger ontology schema that allows several urban elements. The main reason for using such a standard is mainly due to the fact that before importing information it is possible to manage the CityGML ontology providing enriched 3D models. In the process of importing external data the proposed work in chapter 2.5 tries to reach the aim by building a workflow which can include a specific information through a semi-automatic process. The reason is not a fully automatic process lies in the fact that the data or at least the link to the external data must be "prepared" in XML in order to fulfil the CityGML structure requirements.

- Can GIS communicate with BIM in order to work with a variable scale of detail and what has been done in terms of interoperability?

Interoperability between GIS and BIM has already been addressed in literature by many researchers. Lately, Autodesk and ESRI are proposing an integration of BIM and GIS and this is already very promising as it is going beyond the limits of these two information models and systems. Unfortunately, there is nothing in open source that allows such interoperability and the integration of external data from other environments is still an issue. In fact, the idea we are moving towards should be of a model that allows interoperability and management of information from several sources. The proposed graph database is going in this direction but it has its own limits. Such solution is not as fast to produce or to manage as the one proposed by the two companies leader in BIM and GIS.

- Which aspect of life can be improved by tracking people outdoor/indoor?

At the first question, it was pointed out that tracking people is reaching a great interest in geomatic field. This is happening also because geomatics have the tools to manage information in space. Knowing people's position in real-time it can help decision making and it may also save lives when it comes to workers in dangerous fields. For example, monitoring in real-time firefighters moving from outdoors in new indoor areas may give an important contribution to their safety.

- Which are the limits of outdoor and indoor positioning and tracking?

Although tracking outdoors is very easy to achieve thanks to GNSS devices, in indoor areas is needed an existing infrastructure for indoor tracking. This is a big issue for those who need to be tracked from outdoor into new indoor areas. Indoor is not only buildings, but also caves, tunnels etc. And even where an indoor tracking system exists, it would not connect to the previous outdoor tracking system based on GNSS. So far, seems impossible to have seamless tracking from these two different environments, except for the proposed solution shown in this thesis. Providing the user with a helmet where is mounted a GNSS device and a VIO system will allow a seamless tracking in all kinds of environments, providing data in real-time.

- What kind of information can be collected from people indoor tracking?

While moving only into indoor areas, the possibilities are many. Using "active" tracking, so providing the user with a device or an application for their smartphone is possible to track them precisely using their smartphone's Bluetooth or WiFi. In "passive" tracking is still possible to find the location of the user, this time with much less accuracy. The advantage of passive tracking it is that is not needed to ask the users to interact or carry any new device but all the people in an indoor area can be monitored at the same time. The collected information in this case would be about their location and time. This will allow to plot graphs such as number of people, density per area, busiest time of the day/week/month/year etc as can be seen in chapter 3.2.2. of this thesis.

- How can we use these information to improve people's life?

Knowing the general behaviour of crowds in a specific indoor space allows to make assessments in terms of risks and helps decision making. Another aspect is linked to the marketing analysis of the space correlated to people. Knowing how people move through the space and which are the most frequented areas can open to the opportunity of reassessing the entire indoor space. By moving points of interest is possible to spread people density, causing time saving or pointing more accurate advertisement.

# 5. Concluding Remarks

The aim of this thesis was to study how geomatics influence and how it can improve the overall management of information in a “data-hungry” world. New technologies are opening new possibilities for a better life. Safety has been the main focus since the beginning of this thesis idea. In literature, there is a lot of effort into producing smart models able to process, organise and manage geomatic information regarding the environment, the infrastructures and the people. Even when safety is not the main focus of these research works, they always give a contribute to improving such aspect, as managing information in the fastest possible way leads to better and faster decision making. Those models that can manage information in real-time and give an immediate response are the so-called Digital Twin models. So far, governments and private companies have already invested millions in the research and realisation of these models, but there is still much work to do. The more data are required to be processed and managed, the more difficult the realisation it becomes. This thesis is giving a contribute not only by providing digital models of the terrain, buildings and all what can be already surveyed with the newest geomatic techniques but also by approaching new kind of data to those models. As seen in the previous chapters of this thesis, this aspect is still very ambitious. First, because those software working with geographic data are not very prone to integrating new kinds of information which is not already expected in their data structure. Second, data exchange among different software is not allowed when they have different purposes, for instance from GIS to BIM. As deeply discussed, this aspect brought a big interest in the research community and a lot has been done especially in the last 15 years. Chapter 2 of this thesis tries to go beyond what has already been produced in current literature by managing the data structure of GIS models for new data integration and tries to reach a new integration of information between information modelling and information systems through graph databases in order to close the gap between different scale of representation such as GIS and BIM offers. Two main works were discussed, and the results show that it is possible to obtain good results in terms of data enrichment and interoperability, although today it is still time-consuming and requires knowledge in many fields to make it possible. The processes for both, data interoperability and data enrichment, need to be automated and standardized in order to make them accessible to all stakeholders.

Lately, geomatic research has set his eyes also in the people living those environments surveyed so far. To have a better understanding of the reality is fundamental to integrate those data in such digital environments and to be able to make assessments from the collected info. This means that the data must be reliable from outdoor, where the use of GNSS already provides a very good result in terms of localization into indoor areas. The latter aspect has a lot of issues to deal with. Before going into that, the tracking from outdoor to indoor needed to be provided with a solution. In literature, this aspect is not deeply investigated and only a few research activities were carried out. In collaboration with CTTC, an algorithm for seamless tracking outdoor-indoor was studied and improved. The scope was to track emergency workers, such as firefighters and civil protection staff, while operating in rescue or dangerous activities such as damaged buildings or villages from earthquakes. The system developed is based on GNSS when moving outdoors and Visual Intertial Odometry when moving in indoor spaces. To make the tracking from GNSS to VIO without interruption while the operator was moving, it was necessary to study the behaviour of the GNSS in the space and how obstacles will interfere with it. Many tests were conducted and the use of Mann-Kendall non-parametric test was integrated into the algorithm in order to offer an accurate linking of the two provided trajectories. This can be applied in real-time and offers a valuable contribute to the security of those people working in hazardous situations.

Moving indoors, the question for data is different due to the different condition of work and people living those space. The focus was on large building where thousands of people ever day spend time. As already mentioned, indoor tracking is prone to big errors when it comes to passive tracking, which means tracking people location and trajectory without asking them to directly interact with the installed system: without using application on their smartphones nor carrying any small tracking device. WiFi tracking was the solution to this specific question. A new algorithm was developed based on literature research WiFi fingerprinting and inspired by artificial intelligence the model was trained to locate people on the area of interest. This led to a big amount of data collected which were analysed to prove first the correct working of the WiFi system installed and then study some aspect of human behaviour related to marketing KPIs and safety of the crowd in such big indoor areas. Deeper HBA are not scope of this thesis since the aim was to prove that data can be effectively used for such analysis and the integration of them in smart digital models will help improving not only sales form the shopping mall point of view, but especially manage information in case of emergency or particular scenarios. In this specific case, the data are interested by the pandemic restriction which gave even more interest to those specific data opening the opportunity to a higher knowledge on human behaviour.

As a summary this thesis represents a workflow of collecting data from surveys, managing tools and information to enrich the produced models, studying weaknesses and techniques to solve those issues encountered in the world of digital modelling and management and finally prove that the made developments are opening the road to smart, in real-time digital twin models.

# References

Agugiaro, G. 2016. “FIRST STEPS TOWARDS AN INTEGRATED CITYGML-BASED 3D MODEL OF VIENNA.” doi: 10.5194/isprsannals-III-4-139-2016.

Akinci, Burcu, Hassan Karimi, Anu Pradhan, Cheng Chien Wu, and Greg Fichtl. 2008. “CAD and GIS Interoperability through Semantic Web Services.” *Electronic Journal of Information Technology in Construction* 13:39–55. doi: 10.1201/9781420068061-13.

Almeida, Andreia, Luisa M. S. Goncalves, Ana P. Falcao, and Sonia Ildefonso. 2016. *3D-GIS Heritage City Model: Case Study of the Historical City of Leiria*.

Angelats, E., P. F. Espín-López, J. A. Navarro, and M. E. Parés. 2021. “Performance Analysis of the Iopes Seamless Indoor-Outdoor Positioning Approach.” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 43(B4-2021):229–35. doi: 10.5194/ISPRS-ARCHIVES-XLIII-B4-2021-229-2021.

Angelats, E., and J. A. Navarro. 2017. “Towards a Fast, Low-Cost Indoor Mapping and Positioning System for Civil Protection and Emergency Teams.” Pp. 9–15 in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. Vol. 42. International Society for Photogrammetry and Remote Sensing.

Angelats, E., J. A. Navarro, and E. Parés. 2020. “Towards Seamless Indoor-Outdoor Positioning: The Iopes Project Approach.” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 43(B4):313–20. doi: 10.5194/ISPRS-ARCHIVES-XLIII-B4-2020-313-2020.

Angelats, Eduard, and José A. Navarro. 2018. “A Concept for Fast Indoor Mapping and Positioning in Post-Disaster Scenarios.” *GISTAM 2018 - Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management* 2018-March:274–81. doi: 10.5220/0006780102740281.

Angles, Renzo, and Claudio Gutierrez. 2008. “Survey of Graph Database Models.” *ACM Comput. Surv.* 40(1). doi: 10.1145/1322432.1322433.

Anumba, C. J., J. Pan, R. R. A. Issa, and I. Mutis. 2008. “Collaborative Project Information Management in a Semantic Web Environment.” *Engineering, Construction and Architectural Management*.

Artese, S., and V. Achilli. 2019. “A GIS TOOL FOR THE MANAGEMENT OF SEISMIC EMERGENCIES IN HISTORICAL CENTERS: HOW TO CHOOSE THE OPTIMAL ROUTES FOR CIVIL PROTECTION INTERVENTIONS.” doi: 10.5194/isprs-archives-XLII-2-W11-99-2019.

Baggio, Carlo, Alberto Bernardini, Riccardo Colozza, Artur v Pinto, and Fabio Taucer. 2007. *Field Manual for Post-Earthquake Damage and Safety Assessment and Short Term Countermeasures (AeDES) Translation from Italian: Maria ROTA and Agostino GORETTI*.

Banerjee, Sourav, Chinmay Chakraborty, and Debashis Das. 2020. “An Approach towards GIS Application in Smart City Urban Planning.” Pp. 71–110 in *Internet of Things and Secure Smart Environments*. Chapman and Hall/CRC.

Beetz, Jakob. 2009. “Facilitating Distributed Collaboration in the AEC/FM Sector Using Semantic Web Technologies.” *Eindhoven: Technische Universiteit Eindhoven*.

Bishr, Yaser. 1998. “Overcoming the Semantic and Other Barriers to Gis Interoperability.” *International Journal of Geographical Information Science* 12(4):299–314. doi: 10.1080/136588198241806.

Butler, Howard, Martin Daly, Allan Doyle, Sean Gillies, Stefan Hagen, and Tim Schaub. 2016. *The Geojson Format*.

Cardinali, V., M. T. Cristofaro, M. Ferrini, R. Nudo, B. Paoletti, and M. Tanganelli. 2020. “AN INTERDISCIPLINARY APPROACH FOR THE SEISMIC VULNERABILITY ASSESSMENT OF HISTORICAL CENTRES IN MASONRY BUILDING AGGREGATES: APPLICATION TO THE CITY OF SCARPERIA, ITALY.” doi: 10.5194/isprs-archives-XLIV-M-1-2020-667-2020.

Cheng, Tao, and Jochen Teizer. 2013. “Real-Time Resource Location Data Collection and Visualization Technology for Construction Safety and Activity Monitoring Applications.” *Automation in Construction* 34:3–15. doi: 10.1016/j.autcon.2012.10.017.

Commission, European, Joint Research Centre, K. Poljanšek, A. Casajus Valles, M. Marín Ferrer, A. de Jager, F. Dottori, L. Galbusera, B. García Puerta, G. Giannopoulos, S. Girgin, M. Hernandez Ceballos, G. Iurlaro, V. Karlos, E. Krausmann, M. Larcher, A. Lequarre, M.

Theocharidou, M. Montero Prieto, G. Naumann, A. Necci, P. Salamon, M. Sangiorgi, M. Sousa, C. Trueba Alonso, G. Tsionis, J. Vogt, and M. Wood. 2019. *Recommendations for National Risk Assessment for Disaster Risk Management in EU : Approaches for Identifying, Analysing and Evaluating Risks : Version 0*. Publications Office.

Contreras, Diana, Thomas Blaschke, Dirk Tiede, and Marianne Jilge. 2016. "Monitoring Recovery after Earthquakes through the Integration of Remote Sensing, GIS, and Ground Observations: The Case of L'Aquila (Italy)." *Cartography and Geographic Information Science* 43(2):115–33. doi: 10.1080/15230406.2015.1029520.

Deak, Gabriel, Kevin Curran, and Joan Condell. 2012. "A Survey of Active and Passive Indoor Localisation Systems." *Computer Communications* 35(16):1939–54. doi: 10.1016/J.COMCOM.2012.06.004.

Deng, Yichuan, Jack C. P. Cheng, and Chimay Anumba. 2016. "Mapping between BIM and 3D GIS in Different Levels of Detail Using Schema Mediation and Instance Comparison." *Automation in Construction* 67:1–21. doi: 10.1016/J.AUTCON.2016.03.006.

El-Diraby, T. A., C. Lima, and B. Feis. 2005. "Domain Taxonomy for Construction Concepts: Toward a Formal Ontology for Construction Knowledge." *Journal of Computing in Civil Engineering* 19(4):394–406. doi: 10.1061/(ASCE)0887-3801(2005)19:4(394).

Feng, Chen, Wain Sy Anthea Au, Shahrokh Valaee, and Zhenhui Tan. 2012. "Received-Signal-Strength-Based Indoor Positioning Using Compressive Sensing." *IEEE Transactions on Mobile Computing* 11(12):1983–93. doi: 10.1109/TMC.2011.216.

Feriol, Florent, Damien Vivet, and Yoko Watanabe. 2020. "A Review of Environmental Context Detection for Navigation Based on Multiple Sensors." *Sensors (Switzerland)* 20(16):1–30. doi: 10.3390/S20164532.

Fernandes, Diogo, and Jorge Bernardino. 2018. "Graph Databases Comparison: Allegrograph, ArangoDB, Infinitegraph, Neo4J, and OrientDB." *DATA 2018 - Proceedings of the 7th International Conference on Data Science, Technology and Applications* 373–80. doi: 10.5220/0006910203730380.

Fosu, Richelle, Kamal Suprabhas, Zenith Rathore, and Clark Cory. 2015. "Integration of Building Information Modeling (BIM) and Geographic Information Systems (GIS)–a Literature Review and Future Needs." Pp. 27–29 in *Proceedings of the 32nd CIB W78 Conference, Eindhoven, The Netherlands*.

Francini, Mauro, Serena Artese, Sara Gaudio, Annunziata Palermo, and Maria Francesca Viapiana. 2018. “To Support Urban Emergency Planning: A GIS Instrument for the Choice of Optimal Routes Based on Seismic Hazards.” *International Journal of Disaster Risk Reduction* 31:121–34. doi: 10.1016/J.IJDRR.2018.04.020.

Gröger, Gerhard, Thomas H. Kolbe, Claus Nagel, and Karl-Heinz Häfele. 2012. “OGC City Geography Markup Language (CityGML) Encoding Standard.”

Gunawan, F. X. Ferdinandus, and Esther Irawati Setiawan. 2017. “GeoJSON Web Service Based Road Assets Management System for Surabaya City Using Mobile GPS.” *20th International Computer Science and Engineering Conference: Smart Ubiquitos Computing and Knowledge, ICSEC 2016*. doi: 10.1109/ICSEC.2016.7859915.

Hagedorn, Benjamin, Matthias Trapp, Tassilo Glander, and Jürgen Döllner. 2009. “Towards an Indoor Level-of-Detail Model for Route Visualization.” *Proceedings - IEEE International Conference on Mobile Data Management* 692–97. doi: 10.1109/MDM.2009.118.

Hor, A. E. Hadi, G. Sohn, P. Claudio, M. Jadidi, and A. Afnan. 2018. “A Semantic Graph Database for BIM-GIS Integrated Information Model for an Intelligent Urban Mobility Web Application.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4(4):89–96. doi: 10.5194/ISPRS-ANNALS-IV-4-89-2018.

Hor, A. H., A. Jadidi, and G. Sohn. 2016. “BIM-GIS INTEGRATED GEOSPATIAL INFORMATION MODEL USING SEMANTIC WEB and RDF GRAPHS.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3:73–79. doi: 10.5194/ISPRS-ANNALS-III-4-73-2016.

Hor, Abdel-Hadi. 2015. “A Semantic Web Platform for BIM-GIS Integration.” Retrieved September 23, 2022 (https://www.researchgate.net/publication/305775768_A_semantic_Web_platform_for_BIM-GIS_integration?channel=doi&linkId=57a0f71b08aeb1604832b785&showFulltext=true).

Hu, Weiming, Tieniu Tan, Liang Wang, and Steve Maybank. 2004. “A Survey on Visual Surveillance of Object Motion and Behaviors.” *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 34(3):334–52. doi: 10.1109/TSMCC.2004.829274.

Ilic, Jelena Milosevic, Rita Bento, and Serena Cattari. 2020. "3DGIS Representation for Supporting Seismic Mitigation Policies at Urban Scale: The Case Study of Lisbon." *Journal of Cultural Heritage* 45:265–78. doi: 10.1016/J.CULHER.2020.04.001.

Karan, Ebrahim P., and Javier Irizarry. 2015. "Extending BIM Interoperability to Preconstruction Operations Using Geospatial Analyses and Semantic Web Services." *Automation in Construction* 53:1–12. doi: 10.1016/J.AUTCON.2015.02.012.

Kendall, Maurice George. 1948. "Rank Correlation Methods."

Keshka, Ahmed Gamal Abdellatif Ibrhaim, Andrea Masiero, Mostafa Mohamed Ahmed Mostafa, and Antonio Vettore. 2020. "A Comparison of Indoor Positioning Approaches with UWB, IMU, WiFi and Magnetic Fingerprinting." *Communications in Computer and Information Science* 1246:129–41. doi: 10.1007/978-3-030-62800-0_11.

Ketzler, Bernd, Vasilis Naserentin, Fabio Latino, Christopher Zangelidis, Liane Thuvander, and Anders Logg. 2020. "Digital Twins for Cities: A State of the Art Review." *Built Environment* 46(4):547–73. doi: 10.2148/BENV.46.4.547.

Ki-Joune Li. 2022. "IndoorGML.Net."

Kolbe, Thomas H., Gerhard Gröger, and Lutz Plümer. 2005. "CityGML: Interoperable Access to 3D City Models." *Geo-Information for Disaster Management* 883–99. doi: 10.1007/3-540-27468-5_63.

Kray, Christian, Holger Fritze, Thore Fechner, Angela Schwering, Rui Li, and Vanessa Joy Anacta. 2013. "Transitional Spaces: Between Indoor and Outdoor Spaces." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8116 LNCS:14–32. doi: 10.1007/978-3-319-01790-7_2.

de Laat, Ruben, and Léon van Berlo. 2011. "Integration of BIM and GIS: The Development of the CityGML GeoBIM Extension." 211–25. doi: 10.1007/978-3-642-12670-3_13.

Lee, Jiyeong. 2004. "A Spatial Access-Oriented Implementation of a 3-D GIS Topological Data Model for Urban Entities." *GeoInformatica 2004 8:3* 8(3):237–64. doi: 10.1023/B:GEIN.0000034820.93914.D0.

Lenticchia, E., and E. Coïsson. 2017. "The Use of GIS for the Application of the Phenomenological Approach to the Seismic Risk Analysis: The Case of the Italian Fortified Architecture." *International Archives of* *the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42(5W1):39–46. doi: 10.5194/ISPRS-ARCHIVES-XLII-5-W1-39-2017.

Li, Ki-Joune, Giuseppe Conti, Evdokimos I. Konstantinidis, Sisi Zlatanova, and Panagiotis D. Bamidis. 2019. “OGC IndoorGML: A Standard Approach for Indoor Maps.” *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*.

Liu, Hongbo, Yu Gan, Jie Yang, Simon Sidhom, Yan Wang, Yingying Chen, and Fan Ye. 2012. “Push the Limit of WiFi Based Localization for Smartphones.” *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM* 305–16. doi: 10.1145/2348543.2348581.

Ma, Zhiliang, and Yuan Ren. 2017. “Integrated Application of BIM and GIS: An Overview.” *Procedia Engineering* 196:1072–79. doi: 10.1016/J.PROENG.2017.08.064.

Malinverni, E. S., F. Mariano, F. di Stefano, L. Petetta, and F. Onori. 2019. “MOdelling in Hbim to Document Materials Decay by a Thematic Mapping to Manage the Cultural Heritage: The Case of ‘Chiesa Della Pietà’ in Fermo.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42(2/W11):777–84. doi: 10.5194/ISPRS-ARCHIVES-XLII-2-W11-777-2019.

Malinverni, Eva S., Andrea A. Giuliano, and Fabio Mariano. 2018. “3D Information Management System for the Conservation of an Old Deserted Military Site.” *2018 IEEE International Conference on Metrology for Archaeology and Cultural Heritage, MetroArchaeo 2018 - Proceedings* 188–92. doi: 10.1109/METROARCHAEO43810.2018.13618.

Malinverni, Eva Savina, Berardo Naticchia, Jose Luis Lerma Garcia, Alban Gorreja, Joaquin Lopez Uriarte, and Francesco di Stefano. 2020. “A Semantic Graph Database for the Interoperability of 3D GIS Data.” *Applied Geomatics* 14:53–66. doi: 10.1007/s12518-020-00334-3.

Mann, Henry B. 1945. “Nonparametric Tests Against Trend.” *Econometrica* 13(3):245. doi: 10.2307/1907187.

Matrone, F., E. Colucci, V. de Ruvo, A. Lingua, and A. Spanò. 2019. “HBIM IN A SEMANTIC 3D GIS DATABASE.” *The International Archives of the Photogrammetry, Remote Sensing and Spatial*

*Information Sciences* XLII-2-W11(2/W11):857–65. doi: 10.5194/ISPRS-ARCHIVES-XLII-2-W11-857-2019.
Mautz, Rainer. 2009. “Overview of Current Indoor Positioning Systems.” *Geodesy and Cartography* 35(1):18–22. doi: 10.3846/1392-1541.2009.35.18-22.
Moeslund, Thomas B., Adrian Hilton, and Volker Krüger. 2006. “A Survey of Advances in Vision-Based Human Motion Capture and Analysis.” *Computer Vision and Image Understanding* 104(2-3 SPEC. ISS.):90–126. doi: 10.1016/J.CVIU.2006.08.002.
Murphy, Maurice, Eugene Mcgovern, and Sara Pavia. 2009. “Historic Building Information Modelling (HBIM).” *Structural Survey* 27(4):311–27. doi: 10.1108/02630800910985108.
Næsset, Erik, and Jon Glenn Gjevestad. 2008. “Performance of GPS Precise Point Positioning under Conifer Forest Canopies.” *Photogrammetric Engineering and Remote Sensing* 74(5):661–68. doi: 10.14358/PERS.74.5.661.
Nagel, Claus, Alexandra Stadler, and Thomas H. Kolbe. 2009. “Conceptual Requirements for the Automatic Reconstruction of Building Information Models from Uninterpreted 3D Models.” in *Proceedings of the Academic Track of the Geoweb 2009-3D Cityscapes Conference in Vancouver, Canada, 27-31 July 2009*.
Nam, Taewoo, and Theresa A. Pardo. 2011. “Smart City as Urban Innovation: Focusing on Management, Policy, and Context.” *ACM International Conference Proceeding Series* 185–94. doi: 10.1145/2072069.2072100.
Nations Office for Disaster Risk Reduction, United. 2020. *Sendai Framework for Disaster Risk Reduction 2015 - 2030*.
Ortiz, Alberto, Francisco Bonnin-Pascual, Emilio Garcia-Fidalgo, and Joan P. Company. 2015. “Saliency-Driven Visual Inspection of Vessels by Means of a Multirotor.” *Researchgate.Net*.
Osaba, Eneko, Roberto Pierdicca, Eva Savina Malinverni, Anna Khromova, Fernando J. Lvarez, and Alfonso Bahillo. 2018. “A Smartphone-Based System for Outdoor Data Gathering Using a Wireless Beacon Network and GPS Data: From Cyber Spaces to Senseable Spaces.” *ISPRS International Journal of Geo-Information 2018, Vol. 7, Page 190* 7(5):190. doi: 10.3390/IJGI7050190.

Pauwels, Pieter, Sijie Zhang, and Yong Cheol Lee. 2017. "Semantic Web Technologies in AEC Industry: A Literature Overview." *Automation in Construction* 73:145–65.

Peachavanish, Ratchata, Hassan A. Karimi, Burcu Akinci, and Frank Boukamp. 2006. "An Ontological Engineering Approach for Integrating CAD and GIS in Support of Infrastructure Management." *Advanced Engineering Informatics* 20(1):71–88. doi: 10.1016/J.AEI.2005.06.001.

Pettitt, A. N. 1979. "A Non-Parametric Approach to the Change-Point Problem." *Applied Statistics* 28(2):126. doi: 10.2307/2346729.

Pollino, Maurizio, Antonio Bruno della Rocca, Grazia Fattoruso, Luigi la Porta, Sergio Lo, Agnese Arolchi Curzio, Valentina James, and Carmine Pascale. 2012. "Open Source GIS Tools to Map Earthquake Damage Scenarios and to Support Emergency."

Rinaudo, F., E. Agosto, and P. Ardissone. 2007. "GIS and Web-GIS, Commercial and Open Source Platforms: General Rules for Cultural Heritage Documentation." Retrieved September 23, 2022 (https://www.scopus.com/record/display.uri?eid=2-s2.0-84974609311&origin=resultslist&sort=cp-f&src=s&st1=GIS+and+Web-GIS%2ccommercial+and+open+source+platforms%3a+general+rules+for+heritage+documentation&sid=12417f7cbe9d47041ff1cf24d95184ad&sot=b&sdt=b&sl=110&s=TITLE-ABS-KEY%28GIS+and+Web-GIS%2c+commercial+and+open+source+platforms%3a+general+rules+for+heritage+documentation%29&relpos=0&citeCnt=11&searchTerm=).

Rong-Hou, Wu, Lee Yang-Han, Tseng Hsien-Wei, Jan Yih-Guang, and Chuang Ming-Hsueh. 2008. "Study of Characteristics of RSSI Signal." *Proceedings of the IEEE International Conference on Industrial Technology*. doi: 10.1109/ICIT.2008.4608603.

Salamah, Ahmed H., Mohamed Tamazin, Maha A. Sharkas, and Mohamed Khedr. 2016. "An Enhanced WiFi Indoor Localization System Based on Machine Learning." *2016 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2016*. doi: 10.1109/IPIN.2016.7743586.

Sebt, M. H., E. Parvaresh Karan, and M. R. Delavar. 2008. "Potential Application of GIS to Layout of Construction Temporary Facilitie." *International Journal of Civil Engineering* 6(4):235–92.

Soltanaghaei, Elahe, Avinash Kalyanaraman, and Kamin Whitehouse. 2018. "Multipath Triangulation: Decimeter-Level WiFi Localization and Orientation with a Single Unaided Receiver." *MobiSys 2018 - Proceedings of the 16th ACM International Conference on Mobile Systems, Applications, and Services* 376–88. doi: 10.1145/3210240.3210347.

Song, Yongze, Xiangyu Wang, Yi Tan, Peng Wu, Monty Sutrisna, Jack C. P. Cheng, and Keith Hampson. 2017. "Trends and Opportunities of BIM-GIS Integration in the Architecture, Engineering and Construction Industry: A Review from a Spatio-Temporal Statistical Perspective." *ISPRS International Journal of Geo-Information* 6(12). doi: 10.3390/IJGI6120397.

di Stefano, Francesco, Stefano Chiappini, Alban Gorreja, Mattia Balestra, and Roberto Pierdicca. 2021. "Mobile 3D Scan LiDAR: A Literature Review." *Geomatics, Natural Hazards and Risk* 12(1):2387–2429.

di Stefano, Francesco, Alban Gorreja, Fabio Piccinini, Roberto Pierdicca, and Eva Savina Malinverni. 2021. "3D GIS FOR A SMART MANAGEMENT SYSTEM APPLIED TO HISTORICAL VILLAGES DAMAGED BY EARTHQUAKE." Universitat Politecnica de Valencia.

Taye, Mohammad Mustafa. 2010. "Understanding Semantic Web and Ontologies: Theory and Applications." *ArXiv Preprint ArXiv:1006.4567*.

Tsilimantou, Elisavet, Ekaterini T. Delegou, Ioannis A. Nikitakos, Charalabos Ioannidis, and Antonia Moropoulou. 2020. "GIS and BIM as Integrated Digital Environments for Modeling and Monitoring of Historic Buildings." *Applied Sciences (Switzerland)* 10(3). doi: 10.3390/APP10031078.

Vacca, Giuseppina, Emanuela Quaquero, Davide Pili, and Mauro Brandolini. 2018. "GIS-HBIM Integration for the Management of Historical Buildings." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42(2):1129–35. doi: 10.5194/ISPRS-ARCHIVES-XLII-2-1129-2018.

Volk, Rebekka, Julian Stengel, and Frank Schultmann. 2014. "Building Information Modeling (BIM) for Existing Buildings - Literature Review and Future Needs." *Automation in Construction* 38:109–27. doi: 10.1016/J.AUTCON.2013.10.023.

Wang, Weiping, Qiang Chang, Qun Li, Zesen Shi, and Wei Chen. 2016. "Indoor-Outdoor Detection Using a Smart Phone Sensor." *Sensors (Switzerland)* 16(10). doi: 10.3390/s16101563.

Wang, Xuyu, Lingjun Gao, Shiwen Mao, and Santosh Pandey. 2017. "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach." *IEEE Transactions on Vehicular Technology* 66(1):763–76. doi: 10.1109/TVT.2016.2545523.

Wikipedia contributors. 2022. "Geographic Information System." Retrieved November 23, 2022 (https://en.wikipedia.org/wiki/Geographic_information_system).

Wikipedia contributors. 2022a. "Graph Database."

Wikipedia contributors. 2022b. "Z Score." Retrieved November 23, 2022 (https://en.wikipedia.org/wiki/Standard_score).

Wu, Chenshu, Zheng Yang, Yunhao Liu, and Wei Xi. 2013. "WILL: Wireless Indoor Localization without Site Survey." *IEEE Transactions on Parallel and Distributed Systems* 24(4):839–48. doi: 10.1109/TPDS.2012.179.

Xie, Peng, and Mark G. Petovello. 2015. "Measuring GNSS Multipath Distributions in Urban Canyon Environments." *IEEE Transactions on Instrumentation and Measurement* 64(2):366–77. doi: 10.1109/TIM.2014.2342452.

Yamamura, Shinji, Liyang Fan, and Yoshiyasu Suzuki. 2017. "Assessment of Urban Energy Performance through Integration of BIM and GIS for Smart City Planning." *Procedia Engineering* 180:1462–72. doi: 10.1016/J.PROENG.2017.04.309.

Yang, L., M. Worboys-Poster presented at COSIT, and undefined 2011. 2011. "Similarities and Differences between Outdoor and Indoor Space from the Perspective of Navigation." *Researchgate.Net*.

Yang, Zheng, Chenshu Wu, and Yunhao Liu. 2012. "Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention." *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM* 269–80. doi: 10.1145/2348543.2348578.

Yiu, Simon, Marzieh Dashti, Holger Claussen, and Fernando Perez-Cruz. 2017. "Wireless RSSI Fingerprinting Localization." *Signal Processing* 131:235–44. doi: 10.1016/J.SIGPRO.2016.07.005.

Yue, Sheng, Paul Pilon, and George Cavadias. 2002. "Power of the Mann–Kendall and Spearman's Rho Tests for Detecting Monotonic Trends in

Hydrological Series.” *Journal of Hydrology* 259(1–4):254–71. doi: 10.1016/S0022-1694(01)00594-7.

Zafari, Faheem, Athanasios Gkelias, and Kin K. Leung. 2019. “A Survey of Indoor Localization Systems and Technologies.” *IEEE Communications Surveys and Tutorials* 21(3):2568–99. doi: 10.1109/COMST.2019.2911558.

Zegeye, Wondimu K., Seifemichael B. Amsalu, Yacob Astatke, and Farzad Moazzami. 2016. “WiFi RSS Fingerprinting Indoor Localization for Mobile Devices.” Pp. 1–6 in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*.

Zhang, Wuming, Jianbo Qi, Peng Wan, Hongtao Wang, Donghui Xie, Xiaoyan Wang, and Guangjian Yan. 2016. “An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation.” *Remote Sensing* 8(6). doi: 10.3390/RS8060501.

Zhang, X., Y. Arayici, S. Wu, C. Abbott, and G. Aouad. 2009. “Integrating Building Information Modelling and Geographic Information Systems for Large-Scale Facilities Asset Management: A Critical Review.” *Proceedings of the 12th International Conference on Civil, Structural and Environmental Engineering Computing*. doi: 10.4203/ccp.91.93.

Zlatanova, S., L. Liu, G. Sithole, J. Zhao, F. Mortari-GISt Report No. 66, and undefined 2014. 2014. “Space Subdivision for Indoor Applications.” *Academia.Edu*.

Zlatanova, Sisi, Liu Liu, and George Sithole. 2013. “A Conceptual Framework of Space Subdivision for Indoor Navigation.” *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, ISA 2013* 37–41. doi: 10.1145/2533810.2533819.

Zlatanova, Sisi, Jinjin Yan, Yijing Wang, Abdoulaye Diakité, Umit Isikdag, George Sithole, and Jack Barton. 2020. “Spaces in Spatial Science and Urban Applications—State of the Art Review.” *ISPRS International Journal of Geo-Information* 9(1). doi: 10.3390/ijgi9010058.

# Appendix A.

# Python scripts

## A.1. Python script from CityGML to JSON

```
#-------------------------------------------------------------------------------
# Name:        module1
# Purpose:
#
# Author:      joaquin
#
# Created:     20/05/2019
# Copyright:   (c) joaquin 2019
# Licence:     <your licence>
#-------------------------------------------------------------------------------
import xml.etree.ElementTree as ET
import os
import ogr

#variable for getting a different key for each row/document in arangodb
_global_key = 1;
_global_edge = 1;
def ElementName(element):
    return element[element.find("}")+1:]

def WriteElement(_property, _key):
    file = open("Vertices.json", "a" )
    elemento ="{\"_key\": \""+str(_key)+"\", "+_property+"}\r\n"
    file.writelines(elemento)
    file.close
    pass

def WriteEdge(_from, _to, _predicate):
```

```
        global _global_edge
        file = open("EdgesPrueba.json", "a" )
        elemento = ""
        elemento                                                                                                   =
"{\"_key\":\"edge"+str(_global_edge)+"\",\"_from\":\"Vertices/"                +
str(_from)        +        "\",        \"_to\":\"Vertices/"        +str(_to)        +"\",
\"_predicate\":\""+_predicate+"\"}\r\n"
        file.writelines(elemento)
        file.close()
        _global_edge += 1
        pass
    def hasID(elemento):
        for attribute in elemento.attrib:
            if (attribute.endswith("id")):
                return True
        return False

    def Contar(elemento, key_parent, _id):
        for elem in elemento:
            name = ElementName(elem.tag)
            global _global_key
            key_children = _global_key
            #Check if a gml:id exist. If it exist we use it as the vertex key.
            if len(elem)>0:
                if len(elem.attrib)>0:
                    if hasID(elem):
                        key_children = elem.get('{http://www.opengis.net/gml}id')
                        _id = key_children
                        _global_key-=1

            #if an element is inside we get its attributes and go for the next
elements
            if len(elem)>0:
                #if it is a geometry we patch it with ogr library
                if "lod" in name:
                    elem_child = elem[0]
                    geom = ET.tostring(elem_child)
                    geomJSON = ogr.CreateGeometryFromGML(geom)
```

```
            geomJSON = geomJSON.ExportToJson()
            WriteEdge(key_parent,"geom_"+_id,name)
            WriteElement("\"name\":\""+name+"\","    +    geomJSON[1:-
1],"geom_"+_id)
          elif "Attribute" in name:
            writeValues(key_children, elem)
            WriteEdge(key_parent,key_children,name)
            _global_key+=1
          else:

            WriteElement("\"name\":\""+name+"\"",key_children)
            WriteEdge(key_parent,key_children,name)
            _global_key+=1

            #if there are attributes we got them
            if len(elem.attrib)>0:
              #escribimos todos los atributos enlazados al elemento actual
              hasAttributes(elem, key_children)

            Contar(elem, key_children,_id)
        else:
          WriteEdge(key_parent,key_children,name)
          WriteElement("\"name\":\""+name+"\","+
"\"results\":\""+str(elem.text)+"\"",key_children)
          _global_key+=1
      pass

  def hasAttributes(elemento,key_parent):
      global _global_key
      key_attribute = _global_key

      #take name of the value
      data = str(elemento.attrib)
      data_ok = "\"name\":\"Properties\","
      for letter in data[1:-1]:
        if letter == "'":
          data_ok+= "\""
        else:
```

```
            data_ok+= letter

        #take the value which is save inside the attribute
    ##    elem_child = elemento[0]
    ##    results = str(elem_child.text)

        WriteEdge(key_parent,key_attribute,"Properties")
        WriteElement(data_ok,key_attribute)
        _global_key+=1


        pass

    def writeValues(key_parent, elem):
        #get name of the type attribute
        type_elem= ElementName(elem.tag)

        #get value
        value =elem[0]
        if value.text == None:
            attribute_value = "NULL"
        else:
            attribute_value = value.text

        #get name of the value
        data = str(elem.attrib)
        data_ok = ""
        for letter in data:
            if letter == "'":
                data_ok+= "\""
            else:
                data_ok+= letter

        WriteElement(data_ok[1:-1]+","+"\"Type\":\""+type_elem+"\","+
"\"value\":\""+attribute_value+"\"",key_parent)

    def main():
        global _global_key
```

```
    tree = ET.parse(r"C:\Users\Alban\Desktop\Tesi\Nueva carpeta\indoorgml\citygml_edi_id.gml")
    root = tree.getroot()

    NODO = 0

    root_key =0
    name = ElementName(root.tag)
    WriteElement("\"name\":\""+name+"\"",root_key)

  # Cicle for all main elements
    for elem in root:
      key_parent = _global_key
      name =ElementName(elem.tag)
      WriteElement("\"name\":\""+name+"\"",key_parent)
      WriteEdge(root_key,key_parent,name)
      _global_key +=1
      print('****************** '+ str(NODO) + ' ******************')

      #FUNCTION to get all elements inside
      Contar(elem, key_parent,"")

      NODO +=1

    #repeat the process for roads.
    tree = ET.parse(r"C:\Users\Alban\Desktop\Tesi\Nueva carpeta\indoorgml\citygml_road_id.gml")
    root = tree.getroot()

  # Cicle for all main elements
    for elem in root:
      key_parent = _global_key
      name =ElementName(elem.tag)
      WriteElement("\"name\":\""+name+"\"",key_parent)
      WriteEdge(root_key,key_parent,name)
      _global_key +=1
```

```
        print('******************          '+          str(NODO)          +          '
*******************')

        #FUNCTION to get all elements inside
        Contar(elem, key_parent,"")

        NODO +=1
    pass

if __name__ == '__main__':
    main()
```

## A.2. Python script from IndoorGML to JSON

```
#-------------------------------------------------------------------------------
# Name:        module1
# Purpose:
#
# Author:      joaquin
#
# Created:     20/05/2019
# Copyright:   (c) joaquin 2019
# Licence:     <your licence>
#-------------------------------------------------------------------------------
import xml.etree.ElementTree as ET
import os
import ogr

#variable for getting a different key for each row/document in arangodb
_global_key = 1;
_global_edge = 1;
def ElementName(element):
```

```
        return element[element.find("}")+1:]

    def WriteElement(_property, _key):
        file = open("Vertices_indoor.json", "a" )
        elemento ="{\"_key\": \""+str(_key)+"\", "+_property+"}\r\n"
        file.writelines(elemento)
        file.close
        pass

    def WriteEdge(_from, _to, _predicate):
        global _global_edge
        key_edge = _global_edge
        _global_edge += 1
        file = open("Edges_indoor.json", "a" )
        elemento = ""
        elemento =
"{\"_key\":\"edge"+str(key_edge)+"\",\"_from\":\"Vertices_indoor/" +
str(_from) + "\", \"_to\":\"" +str(_to) +"\",
\"_predicate\":\""+_predicate+"\"}\r\n"
        file.writelines(elemento)
        file.close()
        pass
    def hasID(elemento):
        for attribute in elemento.attrib:
            if (attribute.endswith("}id")):
                return True
        return False

    def hasxlink(elemento):
        for attribute in elemento.attrib:
            if (attribute.endswith("}href")):
                return True
        return False

    ##def Transition(elemento, key_parent):
    ##    ##OJOOO FALTA CREAR EL EJE TRANSITION Y SU NODO,
POSTERIORMENTE CREAMOS LAS CONEXIONES YA ESCRITAS
    ##    for elem in elemento:
```

```
##       name= ElementName(elem.tag)
##       if name == "connects":
##          link = elem.get('{http://www.w3.org/1999/xlink}href')
##          WriteEdge(key_parent,"Vertices_indoor/" + link[1:],name)
##       else:
##          link = elem.get('{http://www.w3.org/1999/xlink}href')
##          WriteEdge(key_parent,"Vertices/" + link[1:],name)
##    pass

def Contar(elemento, key_parent, _id):

   #CASO ESPECIAL EN EL QUE SEA CELLSPACE STATE+DUALITY, O TRANSITION +GEOMETRY

   for elem in elemento:
      global _global_key
      key_children = _global_key
      name= ElementName(elem.tag)

      #caso en el que sea un transition es necesario especficiar cual es al citygml y cual al indoorgml
##       if name == "Transition":
##          Transition(elem, key_parent)
##          continue
      #Check if a gml:id exist. If it exist we use it as the vertex key.
      if len(elem.attrib)>0:
         if hasID(elem):
            key_children = elem.get('{http://www.opengis.net/gml/3.2}id')
            _id = key_children
            _global_key-=1

      #Check if a href exist. If it exist we connect it to the element, for this we don't need to create a new vertex
      if len(elem.attrib)>0:
         if hasxlink(elem):
            link = elem.get('{http://www.w3.org/1999/xlink}href')
            if "#" in link:
               link = link[1:]
```

```
            if name == "duality":
                WriteEdge(key_parent,"Vertices/" + link,name)
            elif name == "geometry":
                WriteEdge(key_parent,"Vertices/geom_" + link,name)
            else:
                WriteEdge(key_parent,"Vertices_indoor/" + link,name)
            continue

        #if an element is inside we get its attributes and go for the next
elements
        if len(elem)>0:
            #if it is a geometry we patch it with ogr library
            if "geometry" in name:
                elem_child = elem[0]
                geom = ET.tostring(elem_child)
                geomJSON = ogr.CreateGeometryFromGML(geom)
                geomJSON = geomJSON.ExportToJson()

WriteEdge(key_parent,"Vertices_indoor/"+str(key_children),name)
                WriteElement("\"name\":\""+name+"\"," + geomJSON[1:-
1],key_children)
                _global_key +=1
            else:

                WriteElement("\"name\":\""+name+"\"",key_children)

WriteEdge(key_parent,"Vertices_indoor/"+str(key_children),name)
                _global_key+=1

                #if there are attributes we got them
##                if len(elem.attrib)>0:
##                    #escribimos todos los atributos enlazados al elemento actual
##                     hasAttributes(elem, key_children)

                Contar(elem, key_children,_id)
        else:
            #registramos el v?rtice del grafo
            results = str(elem.text)
```

```
WriteEdge(key_parent,"Vertices_indoor/"+str(key_children),name)
            WriteElement("\"name\":\""+name+"\","+
"\"results\":\""+results+"\"",key_children)
            _global_key+=1
      pass

   def writeValues(key_parent, elem):
      #get name of the type attribute
      type_elem= ElementName(elem.tag)

      #get value
      value =elem[0]
      if value.text == None:
         attribute_value = "NULL"
      else:
         attribute_value = value.text

      #get name of the value
      data = str(elem.attrib)
      data_ok = ""
      for letter in data:
         if letter == "'":
            data_ok+= "\""
         else:
            data_ok+= letter

      WriteElement(data_ok[1:-1]+","+"\"Type\":\""+type_elem+"\","+
"\"value\":\""+attribute_value+"\"",key_parent)

   def main():

      tree          =          ET.parse(r"C:\Users\Alban\Desktop\Tesi\Nueva
carpeta\indoorgml\prova.gml")
      root = tree.getroot()

      NODO = 0
```

```
    root_key =0
    name = ElementName(root.tag)
    WriteElement("\"name\":\""+name+"\"",root_key)

  # Cicle for all main elements
    for elem in root:
      name =ElementName(elem.tag)

      print('******************           '+           str(NODO)          +          '
*******************')

      if name == "cellspace":
        key_children = elem.get('{http://www.opengis.net/gml/3.2}id')
        WriteEdge(root_key,"Vertices/" + key_children,name)
        NODO +=1
        continue

      global _global_key
      key_parent = _global_key
      WriteElement("\"name\":\""+name+"\"",key_parent)
      WriteEdge(root_key,"Vertices_indoor/"+str(key_parent),name)
      _global_key +=1
      #FUNCTION to get all elements inside
      Contar(elem, key_parent,"")

      NODO +=1

    pass

  if __name__ == '__main__':
    main()
```