POLITECNICO DI TORINO
Master of Science Program

Master Degree in Urban and Territorial Planning - Urban Planning for the Global Urban Agenda

Final Dissertation

Data Preparation for the Construction of Territorial Databases in Mobility Planning

Digitizing pedestrian infrastructure in the City of Turin

BALDIN FILIPPO - 342691

Supervisor:
Professor Emere Arco

Academic year
2025/2026

To my family,
who have always believed in me and supported me every step of the way.

To my grandparents,
guarding over me from up there.

To my friends,
for their unconditional support and for bringing smiles to my face even in the most difficult moments.

To all the people who have contributed,
in ways large and small, to this important milestone.

Thank you from the bottom of my heart.

Abstract

The increasing availability of geospatial data has created new opportunities for urban planning, but it has also highlighted the challenges associated with managing heterogeneous datasets originating from different sources, formats, and collection methods. This thesis investigates the process of preparing pedestrian infrastructure data for integration into the municipal cartographic database of the City of Turin, with particular attention to the requirements of the INSPIRE framework for spatial data interoperability.

The work addresses a practical gap between existing cartographic layers and the requirements of network-based pedestrian analysis, where geometric completeness alone is not sufficient without semantic consistency and topological continuity.

The research focuses on the case study of Circoscrizione 1, the central district of Turin, where pedestrian-related datasets were collected and analyzed, according to the INSPIRE Transport Networks data model; the methodology involved the selection of the most appropriate INSPIRE layer, the harmonization of local data structures, and a validation process aimed at assessing the consistency, completeness, and geometric accuracy of the resulting dataset. The workflow combines DBT/INSPIRE road and infrastructure layers, CAD-derived pedestrian crossings, 2023 orthophotos and manually documented correction layers.

The workflow made it possible to identify, classify and correct a set of geometric and semantic inconsistencies, while documenting cases requiring future validation by municipal offices and CSI Piemonte

Future developments may build upon the standardized dataset produced in this research to create a detailed pedestrian graph for the City of Turin: such a graph could support advanced applications in accessibility analysis, walkability assessment, route optimization, and sustainable mobility planning, contributing to more informed and data-driven urban planning practices. The dataset should therefore be interpreted as a graph-ready infrastructure layer rather than as a complete routable network.

Keywords: Spatial Data Infrastructure (SDI); INSPIRE; Pedestrian Infrastructure; Geospatial Databases; Urban Mobility.

Table of contents

1. Introduction

1.1 Context and relevance of cartographic databases in urban planning

The increasing complexity of contemporary cities and the rapid growth of urban populations have made the management of spatial information a central element of modern urban planning. Planners must deal with large amounts of heterogeneous data related to land use, infrastructure, environmental conditions, and socio-economic dynamics, which require systematic organization and analysis (Sinjari & Kosovrasti, 2015).

In pedestrian mobility planning, this issue is particularly relevant because walkable spaces are often represented across multiple layers rather than as a single continuous network.

In this context, cartographic databases play a fundamental role because they allow spatial data to be stored, structured, and integrated within Geographic Information Systems (GIS), enabling the visualization and analysis of spatial relationships and urban processes (Boeing, 2019): by combining digital maps with attribute data, these databases support the identification of patterns, trends, and interactions between urban elements, providing planners with evidence-based information for decision-making (Fugro, 2025), moreover, the availability of large geospatial datasets from satellite imagery, remote sensing technologies, and open governmental data has further increased the importance of well-organized cartographic databases for urban analysis and planning (Biljecki et al., 2021).

As a result, the preparation and management of reliable spatial datasets have become essential for supporting sustainable urban development, improving the efficiency of planning processes, and enabling more informed policy decisions in increasingly data-driven urban environments (USC GIS Program, 2023).

The object of the broader study in general is the implementation of a research project on the theme “Street lighting and pedestrian movements in the city of Turin”.

This thesis is part of a broader research context in which pedestrian infrastructure data are required as a spatial input for movement analysis and scenario modelling.

The project seeks to develop a comprehensive geospatial framework capable of supporting analyses of pedestrian accessibility, safety, and walking behaviour by integrating heterogeneous spatial datasets within the municipal cartographic database.

This project aims to use an Agent-Based Model (ABM) to analyse:

- the differences between daytime and night-time pedestrian flows;
- the effects of different levels of street lighting on pedestrian movement during evening and night-time hours, and on the perception of safety by vulnerable groups
- the identification of areas where people, after nightfall, might feel uncomfortable when moving on foot.

This model will use the GIS data included in the City’s Cartographic Database and also the geographic data necessary for the development of the model itself, which will be produced in collaboration between the two Parties.

These modelling objectives require a spatial structure able to represent pedestrian choices at segment level; for this reason, the preparation of a graph-ready dataset is a necessary preliminary step.

1.2 Research objectives and question

The objective is twofold. First, the thesis defines a reproducible workflow for preparing pedestrian infrastructure data. Second, it evaluates whether the resulting dataset can support future graph construction and municipal data maintenance.

For the city of Turin, in the field of pedestrian infrastructure, we have the following main data sources: the urban cadastre, local and regional authorities and finally open data sources; specific criteria were used to select the pedestrian features, mainly pedestrian areas, crossings, sidewalks, stairs, traffic islands and connective elements.

Heterogeneous datasets were harmonized by format, attribution structure and compliance with the municipal territorial database model although the majority of the datasets used were already standardized in accordance with normal practice required by the offices of the municipality.

Only datasets directly contributing to the pedestrian infrastructure layer are included in the operational workflow; other sources are discussed as contextual or supporting references.

The preparation and integration of high-quality geospatial data into municipal cartographic databases have significant implications for urban planning processes, particularly in complex urban environments such as Turin: reliable and well-structured mobility and pedestrian infrastructure data enable planners to perform more accurate spatial analyses, supporting evidence-based decision-making in areas such as accessibility, traffic management, and sustainable mobility planning (Longley et al., 2015).

The evaluation focuses on completeness, positional consistency, attribute coherence, geometric validity and potential topological continuity.

The thesis does not aim to fully simulate pedestrian movements, but to prepare and validate the spatial data required for the future construction of a pedestrian graph.

A secondary focal point of this thesis is how heterogeneous data are selected, harmonized and validated for an interoperable dataset prepared for a future pedestrian graph.

The main research question of this thesis is in particular how the mobility data and pedestrian infrastructure data of the area of District (Circoscrizione) 1 are prepared for implementation in the municipal database of the City of Turin by following the INSPIRE directive data structure and specifications.

The research focuses on understanding the procedures, standards, and technical steps required to transform heterogeneous mobility datasets into structured geospatial information that complies with the INSPIRE data model.

In particular, the thesis seeks to answer the following research question:

How are mobility data and pedestrian infrastructure data collected, processed, and structured in order to be integrated into the municipal geospatial database of the City of Turin following the INSPIRE data structure and standards?

The thesis therefore stops at the construction of a validated pedestrian infrastructure dataset and does not claim to deliver a fully routable pedestrian graph.

1.3 Methodology

The methodology adopted in this thesis is structured as an applied geospatial data preparation workflow: the aim is to transform heterogeneous municipal datasets into a coherent and validated pedestrian infrastructure dataset for Circoscrizione 1 of the City of Turin; the workflow is designed to support the future construction of a pedestrian graph and its integration into the municipal geospatial database.

The following figure 1.1 will present a block scheme of the workflow necessary for the preparation of the data, this is to allow for a clear view of the data elaboration needed for an effective result:

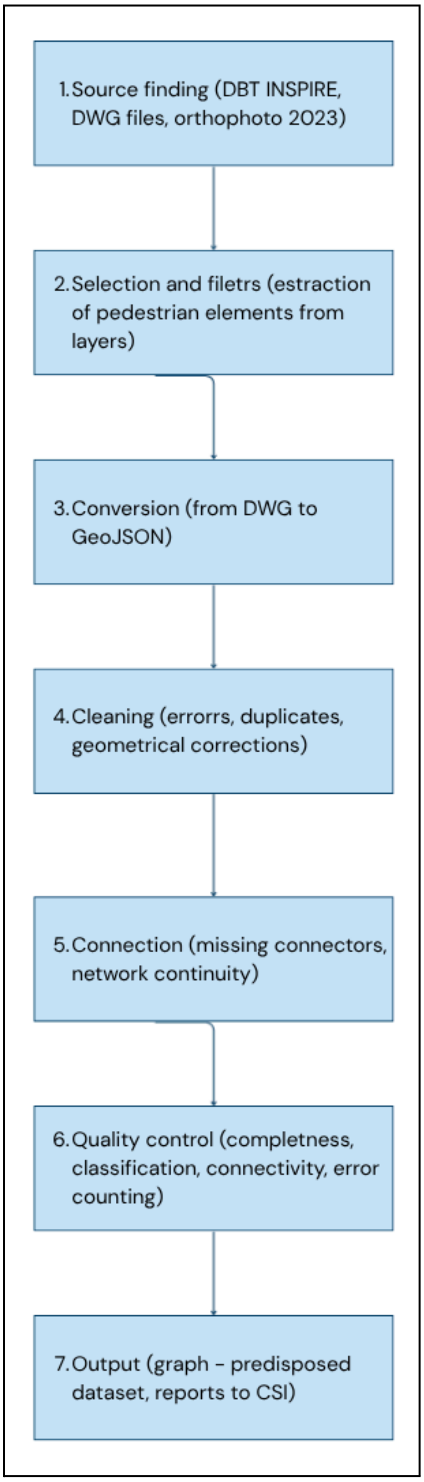

Figure 1.1: Block scheme of the workflow necessary for the data preparation

The data used are:

1. Road network projects drawn on the city cartography by municipal offices using AutoCAD software, the drawing represents all the projected road signs, indicated as geometries, however, these projects were created in a non-standard manner and therefore require data cleansing: in particular, relevant information on pedestrian crossings on the road was extracted.

2. INSPIRE database, organized in a standard manner according to European regulations and customized to meet the needs of the city of Turin, the accuracy of the geographic data, and especially that of the related data, is then analyzed, and any errors are flagged for correction in future versions. (This data is updated quarterly by CSI-Piemonte)

Once analyzed, the data is collected and inserted into specialized professional software such as AutoCAD for cleanup: this process involves eliminating unnecessary or irrelevant data, with the goal of producing an output file containing the specific data; once this phase is completed, the project file is converted for insertion into other professional software such as QGIS for any modifications or corrections.

Each phase produces an auditable output, including selected layers, converted files, cleaned geometries, connection elements, quality notes and cases marked for future update.

2. Theoretical and Conceptual Framework

2.1 Geospatial data and urban planning

2.1.1 Role of GIS in urban planning processes

A GIS is a system for acquiring, storing, managing, analysing and visualising spatial data. In urban planning, GIS supports the interpretation of spatial relationships and the evaluation of territorial transformations.

Use of GIS in urban planning helps and guides planners for an orderly development of the territory: continued growth of the population in urban centers generates the need for expansion of urban space, for its planning in terms of physical and social infrastructures in the service of the community, based on the principles of sustainable development.

In addition urbanization is accompanied with numerous structural transformations and functional cities like Turin, which should be evaluated in spatial context, to be managed and planned according to the principles of sustainable development: urban planning connects directly with land use and design of the urban environment, including physical and social infrastructure in service of the urban community, constituting a challenge to global levels, therefore, use of GIS in this field is a different approach regarding the space, its development and design, analysis and modeling of various processes occurring in it, as well as interconnections between these processes or developments in space (Xhafa, 2015).

2.1.2 Usage of spatial data for mobility, accessibility, infrastructures and decision support

Spatial data have become an essential resource for understanding how cities function and for supporting planning decisions at different scales: in the field of mobility, geospatial datasets make it possible to analyse movement patterns, identify areas of congestion, evaluate the performance of transport networks, and assess the connectivity between different parts of the city (Longley et al., 2015).

When combined with information on pedestrian infrastructure, public transport, and land use, spatial data can also be used to measure accessibility, helping planners understand how easily residents can reach services, employment opportunities, schools, healthcare facilities, and public spaces (Geurs & van Wee, 2004).

In the management of urban infrastructure, geospatial information provides a detailed representation of roads, sidewalks, utilities, and public assets, facilitating maintenance activities and the identification of areas requiring intervention (Batty, 2013), more broadly, the integration of spatial data within Geographic Information Systems (GIS) supports evidence-based decision-making by allowing planners and public authorities to visualize complex urban phenomena, compare alternative scenarios, and evaluate the potential impacts of planning policies before their implementation (Goodchild, 2007).

As cities increasingly rely on data-driven approaches, the availability of accurate and well-structured spatial information has become a fundamental requirement for effective and sustainable urban development (Longley et al., 2015). In Turin, this is particularly relevant because pedestrian mobility depends on dense interactions between historic street layouts, public transport nodes, pedestrian areas and road crossings

2.2 Type of geographic data

2.2.1 Vector and raster data

In this thesis, vector data are used for pedestrian infrastructure objects, while raster orthophotos are used as reference layers for positional verification.

Vector data provide a way to represent real world features within the GIS environment, while a feature is anything you can see on the landscape. Imagine you are standing on the top of a hill: looking down you can see houses, roads, trees, rivers, and so on, each one of these things would be a feature when we represent them in a GIS Application. Vector features have attributes, which consist of text or numerical information that describe the features: a vector feature has its shape represented using geometry, the geometry is made up of one or more interconnected vertices.

A concrete example would be drawing the contour of a piazza: if it has multiple uses, such as vehicular and pedestrian, these features must be reported to allow for an efficient representation of the piazza and its uses.

Raster data, on the other hand, is used in a GIS application when we want to display information that is continuous across an area and cannot easily be divided into vector features: point, polyline and polygon features work well for representing some features on this landscape, such as trees, roads and building footprints. Other features on a landscape can be more difficult to represent using vector features: an example would be an orthophoto, where we have an innumerable amount of elements that not always can be represented using vector features.

The following figure 2.1 is a prime example of raster data used for elaborations, as orthophotos contain all the necessary data used to check the validity of the used vector data, while figure 2.2 shows a collection of all the orthophotos obtained and joined in order to create a unitary framework that represents the entire city of Turin:

Figure 2.1: Orthophoto of the city of Turin of 2023 with the area of Circoscrizione 1

Figure 2.2: Unitary framework of the orthophotos of 2023 and the various Circoscrizioni

2.2.2 Definition and main characteristics of the treated elements

A “map” and a “cartographic database” are closely related but serve different purposes in the management and representation of geographic information.

A map is a visual representation of spatial features designed to communicate geographic information in a clear and understandable form: it displays selected elements of the real world, such as roads, buildings, land uses, or administrative boundaries, according to a specific scale and purpose (Longley et al., 2015).

A cartographic database, on the other hand, is a structured digital repository that stores, organizes, and manages the geographic data underlying one or more maps: unlike a map, which presents information in a fixed visual format, a cartographic database contains the geometric, topological, and attribute information associated with spatial entities and allows data to be queried, updated, analyzed, and reused for different applications (Shekhar & Xiong, 2008).

Therefore, while a map can be considered the final product used to visualize geographic information, a cartographic database represents the underlying data infrastructure from which multiple maps and spatial analyses can be generated: in modern Geographic Information Systems (GIS), cartographic databases provide the foundation for dynamic mapping, spatial analysis, and decision-making processes in urban planning (Longley et al., 2015).

The municipal database is therefore not only a cartographic product but also an operational infrastructure for maintenance, analysis and data exchange.

3. Sources and data quality

3.1 Institutional Data Sources

3.1.1 Cadastre

In the present workflow, cadastral data are not the main source for pedestrian infrastructure, but they provide a useful reference for parcel boundaries and institutional data integration.

Within the municipal context of Turin, cadastral information can be combined with other spatial datasets—such as mobility networks and pedestrian infrastructure—to create comprehensive geospatial databases used for planning and territorial management, however, the integration of cadastral data into municipal systems often requires processes of harmonization and standardization to ensure compatibility with spatial data infrastructures and European interoperability frameworks, including the INSPIRE directive (Williamson et al., 2010).

Technical Considerations:

- **Reference System**: It is essential to manage the projection. The data is often in ETRF2000 (RDN 2008), in accordance with national law, and it is recommended to use EPSG 6706 or EPSG 25832 (ETRS89) or Gauss-Boaga EPSG 3003 which for example is used by the city of Turin in QGIS.

The geographic reference system of the Italian land registry is historically based on the Cassini-Soldner equivalent cylindrical projection, with the local datum referred to Rome M. Mario. Today, the cadastral cartography is progressively aligned with the national geodetic system ETRF2000 (ETRS89 standard), used in the Cadastral Cartographic Geoportal and in the WMS services of the Agenzia delle Entrate.

- **Data Structure**: Downloaded GML files contain "map" (sheets) and "ple" (parcels), post-processing (field calculator in QGIS) is often required to extract useful information such as the sheet number from the cadastral code string.
- **Geometry**: Downloaded vector data allows you to work on individual elements (parcels) for territorial analysis, selection by attributes (code, sheet, cadastral map), and creation of unique indexes (IDGIS).
- **Land Registry**: In the provinces of Trento and Bolzano, the land registry is managed using the land registry system, which differs from the standard system.

Land registry maps are freely accessible, while detailed data from surveys or plans require authorized access.

3.1.2 Local and regional authorities

In the context of the city of Turin, local and regional authorities represent a primary source of geospatial data used for the development and maintenance of municipal cartographic databases: the Municipality of Turin and the Piedmont Region collect and manage a wide range of spatial datasets related to urban infrastructure, mobility networks, public transportation, land use, and environmental features: these datasets are typically managed within local Geographic Information Systems and are used to support urban planning, infrastructure management, and policy-making processes (Longley et al., 2015).

In recent years, the Municipality of Turin has also made several datasets available through its municipal open data portal, allowing access to spatial information such as transportation networks, pedestrian infrastructure, and administrative boundaries (Comune di Torino, 2023).

At the regional level, the Piedmont Region provides additional geospatial datasets through its spatial data infrastructure, which can complement municipal information and support broader territorial analyses (Regione Piemonte, 2022); however, integrating datasets produced by different administrative bodies may require harmonization processes to address differences in formats, standards, and levels of detail, particularly when adapting the data to European interoperability frameworks such as the INSPIRE directive (European Commission, 2016).

This distinction is important because only official municipal layers can be directly integrated, while thesis-produced corrections require validation before becoming authoritative data.

3.2 Open Data

3.2.1 Open data sources

Open data can support comparison and enrichment, but it cannot replace official municipal datasets when the aim is integration into the city cartographic database.

Open data sources represent an increasingly important resource for the collection of geospatial information used in urban planning and cartographic databases: in the case of the city of Turin, a large number of datasets are made publicly available through the **Municipal City's Geoportal AperTO**, which acts as a central platform where users can search, visualize, and download datasets produced by the municipality and by other local institutions (Comune di Torino, 2023). The portal provides access to a wide range of information related to urban systems, including demographic data, administrative boundaries, and mobility-related datasets that can be used for spatial analysis and research.

In addition, the city offers mobility-related open datasets through the municipal infomobility platform, which includes information such as real-time traffic flows, parking availability, and the geometry of ZTL (Zone Traffico Limitato, *Limited Traffic Zones*) (Città di Torino, 2024). These open data initiatives aim to increase transparency, promote innovation, and encourage the reuse of public information by researchers, developers, and planners, enabling the integration of multiple datasets into municipal geospatial databases used for urban analysis and planning activities (Open Knowledge Foundation, 2023).

Another useful site for general analysis is the DBSN (*National Synthesis DataBase*, Database di Sintesi Nazionale: this site was first made available in 2023, and is derived by regional DBGT, ISTAT and Cadastral data; it is administered by the **IGM** (*Military Geographical Institute*, Istituto Geografico Militare).

Cadastral data also represents a fundamental source for the development of municipal cartographic databases and for supporting urban planning activities: in Italy, cadastral information is managed by the national land registry administered by the Italian Revenue Agency, which provides detailed data on land parcels, property boundaries, and building units (Agenzia delle Entrate, 2022), these datasets constitute an essential spatial reference layer that can be integrated into geographic information systems to support analyses related to land use, property structure, and urban infrastructure (Dale & McLaughlin, 1999).

3.3 Acquisition techniques

3.3.1 Direct surveying (GPS, total stations)

Direct surveying in GIS data acquisition is a primary method of capturing high-accuracy spatial data (points, lines, polygons) directly from the field, often called "primary data acquisition": it involves using advanced electronic equipment to collect precise measurements (X, Y, Z coordinates) on the ground, which are then integrated directly into a Geographic Information System (GIS) for mapping and analysis.

Key Methods of Direct Surveying:

- **Global Navigation Satellite Systems (GNSS/GPS)**: This technology provides high-accuracy 3D positioning, enabling surveyors to capture spatial data for mapping in real-time.
- **Electronic Total Stations (ETS)**: These are commonly used for field surveys to collect precise measurements that are directly converted into digital datasets.
- **Mobile Mapping Systems (MMS)**: These systems use multi-sensor platforms (e.g., vehicle-mounted cameras, lasers, and GPS) to capture high-volume 3D data efficiently.
- **Field Data Recording**: Surveyors enter raw data, measurements, and field sketches directly into mobile devices, ensuring high accuracy, particularly for boundary, as-built, and topographic surveys.

GIS surveying applications:

- **Land surveying and property management**: Land surveys are crucial for establishing property boundaries, and GIS technology enhances this by offering precise measurements and detailed information for land surveys, this allows for efficient property management and supports large-scale projects that require accurate spatial information.
- **Disaster management and emergency response**: Real-time data collection through GIS technology enables quick responses to disasters by identifying specific geographical areas and locations that need immediate attention.
- **Urban planning and development**: GIS work in urban planning ensures efficient utilization of spaces, taking into account various geographical changes and helping in creating sustainable cities.
- **Environmental and natural resources management**: Geographical data, when mapped through GIS, aids in monitoring and managing natural resources, thereby contributing to sustainable development and ecological preservation.

- **Transportation and infrastructure planning**: From scheduling timetables to maintenance schedules, GIS supports transportation and infrastructure projects by providing a detailed view of surrounding geographic features and other locations through topographic surveys.
- **Archeological and cultural heritage surveying**: The past can be mapped in the present. GIS aids archaeologists in capturing data from excavation sites, ensuring the preservation of our cultural heritage.

Advantages and Disadvantages:

- **Advantages**: Direct surveying provides high accuracy (often centimeter-level) and is effective for capturing specific, highly detailed, or new data on the ground.
- **Disadvantages**: It is generally more time-consuming, cost-intensive, and labor-intensive compared to secondary methods like satellite imagery analysis

3.3.2 Remote sensing (satellites, drones)

Remote sensing means acquiring and measuring information about an object or phenomenon via a device that is not in physical or direct contact with what is being studied (Colwell, 1983): to collect remotely sensed data, a platform – an instrument that carries a remote sensing sensor – is deployed.

From the mid 1800's to the early 1900's, various platforms such as balloons, kites, and pigeons carried mounted cameras to collect visual data of the world below, today, aircraft (both manned and unmanned) and satellites collect the majority of remotely sensed data. The sensors typically deployed on these platforms include film and digital cameras, light-detection and ranging (LiDAR) systems, synthetic aperture radar (SAR) systems, and multi-spectral and hyper-spectral scanners; although satellites are considered primary platforms (collecting much of the data used by governmental and research organizations), other platforms such as ground-based solutions are quite prevalent and popular. Different platforms can carry different types of sensors depending upon the needs of the user, and each platform offers advantages and disadvantages in terms of range, cost, stability, frequency, scale, and accuracy.

Key advantages:

- **Remote Monitoring**: Allows for mapping in inaccessible or hazardous areas.
- **High Efficiency & Lower Cost**: Provides systematic, large-scale data collection compared to field surveys.
- **Change Detection**: Detects changes in vegetation, soil moisture, and infrastructure over time

3.3.3 Digitization of analogue maps

Digitization of analogue maps is an important method for transforming historical or paper-based cartographic information into digital geospatial datasets that can be used in geographic information systems: this process typically involves scanning paper maps or aerial photographs, georeferencing them within a coordinate reference system, and converting relevant features—such as roads, parcels, or infrastructure—into vector data that can be stored and analyzed in a GIS environment (GISRSStudy, 2021).

In the case of the city of Turin, the digitization of historical cartographic materials has been an important activity for preserving and reusing spatial information produced in the past: the municipality possesses extensive historical archives, including aerial photogrammetric images covering the city from 1936 to the present, which have been progressively digitized and integrated into GIS platforms to support urban analysis and documentation of urban transformation (Golzio, Pirrello, 2020); these digitization initiatives are often carried out with the technical support of regional institutions such as CSI Piemonte, which manage the full lifecycle of geographic data, including the recovery and digital processing of historical cartographic materials (CSI Piemonte, 2024): through these processes, analogue maps and aerial imagery become accessible digital datasets that can be incorporated into municipal cartographic databases and used for urban planning, historical analysis, and spatial research.

3.4 Quality dimension

3.4.1 Data updating

The correction layer created in the thesis can be interpreted as an issue-tracking layer, distinguishing between already corrected objects and objects to be verified in future update cycles.

In the context of municipal geospatial databases, regular updating processes involve the integration of new data from surveys, administrative records, and real-time sources such as sensors or mobility tracking systems (Goodchild, 2007).

For a case study such as the city of Turin, keeping datasets related to mobility and pedestrian infrastructure updated is particularly important due to continuous changes in traffic regulations, infrastructure improvements, and urban development projects, therefore, implementing systematic updating procedures and version control mechanisms is crucial to ensure that geospatial databases remain accurate, relevant, and aligned with current urban conditions, especially when complying with standardized frameworks such as INSPIRE (European Commission, 2016).

In particular, the city of Turin entrusts the job of updating the city's cartography to the CSI (Consorzio Sistemi Informativi), which executes regular updates every three months, implementing error corrections and new objects placement on the city's official cartography.

3.4.2 Completeness

Completeness is a key dimension of spatial data quality and refers to the degree to which all relevant geographic features, attributes, and relationships are present within a dataset (EFGS, 2023), in the context of geographic information systems, a dataset is considered complete when it accurately represents all the spatial elements necessary to describe a given territory without omissions or missing attributes (Geo Contributor, 2008).

For pedestrian graph preparation, completeness must also be topological: relevant objects must not only exist, but must also be spatially connectable into a continuous network.

For a case study focused on the city of Turin, completeness is particularly important when preparing datasets related to mobility networks and pedestrian infrastructure for integration into municipal cartographic databases: for example, ensuring completeness means verifying that all relevant pedestrian paths, sidewalks, crossings, and associated attributes—such as accessibility or surface type—are represented within the dataset: incomplete spatial data can lead to inaccurate analyses and may negatively affect urban planning decisions or mobility assessments (Spatial Data Support, 2023).

Therefore, during the preparation of geospatial datasets for the municipal database of Turin, completeness checks are necessary to ensure that the data accurately reflect the urban environment and that no significant spatial features are omitted before integration into standardized structures such as those required by the INSPIRE framework.

3.4.3 Methodological formulation of the case study

This research adopts an applied case-study methodology aimed at developing a reproducible workflow for preparing, harmonizing, and validating geospatial data related to pedestrian infrastructure: the case study focuses on Circoscrizione 1 of the City of Turin and uses heterogeneous municipal datasets, including the official territorial database, CAD-based road projects, orthophotos, and manually created correction layers.

The methodology is designed to support the future construction of a pedestrian graph and its integration into the municipal geospatial database.

3.4.4 Difference between “Pedestrian dataset” and “Pedestrian graph”

In the thesis, we often talk about "graph creation," but many operations actually involve preparing pedestrian layers. Two levels are distinguished:

Level	**What it contains**
Level 1 - Structured pedestrian dataset	Physical infrastructure elements: crossings, pedestrian areas, sidewalks, passages, stairs, traffic islands
Level 2 - Pedestrian graph	Topological network of nodes and arches, usable for routing, accessibility and mobility analysis

The thesis builds and validates Level 1 and prepares for Level 2.

Level 1 stores objects as they exist in the cartographic database; Level 2 transforms these objects into a topological model in which movement is represented through nodes, edges and connectivity rules.

The following figure 3.1 demonstrates concretely an example of a level 1 accuracy pedestrian graph applied to the context of the city of Turin:

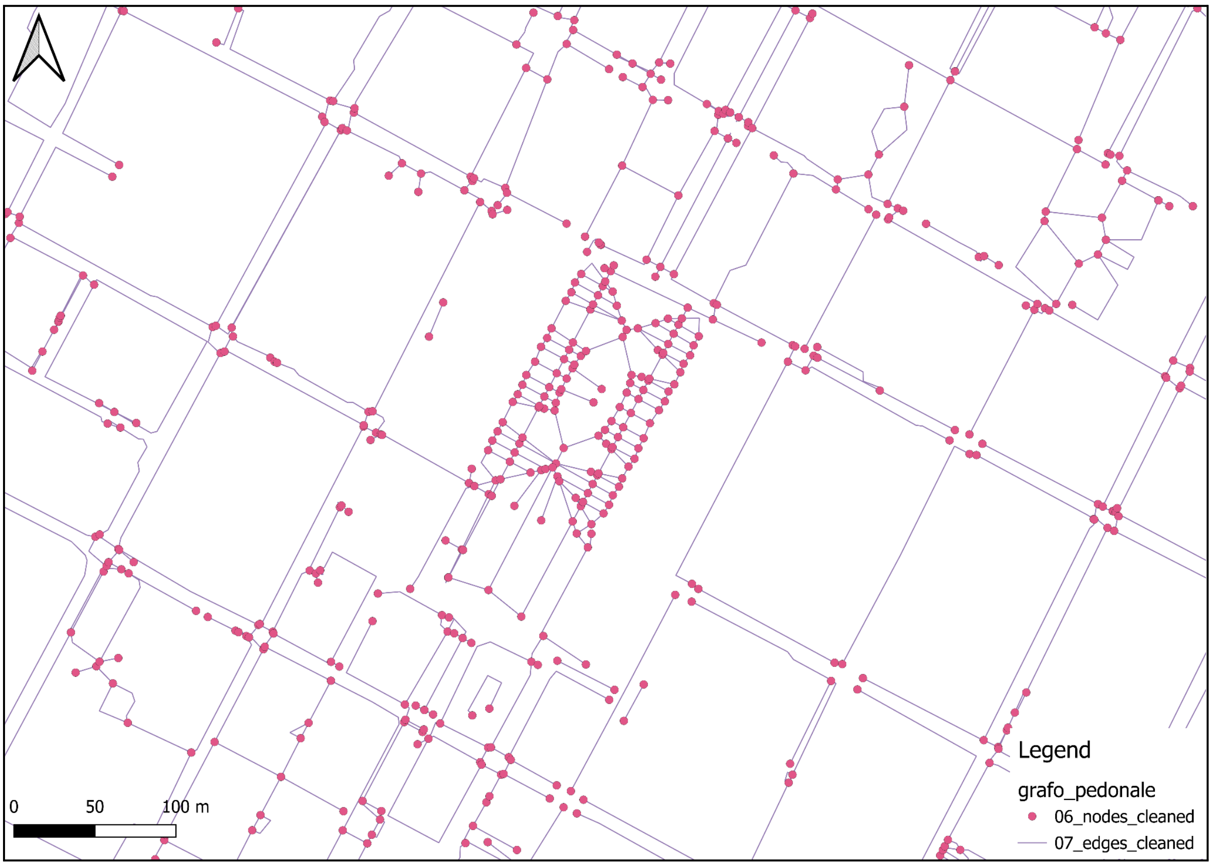

Figure 3.1: pedestrian graph in the area of Piazza Solferino

4. Data Preparation

4.1 Chapter structure: Workflow phases

4.1.1 Case study definition

The case study is limited to Circoscrizione 1 of the City of Turin, selected because of its high pedestrian intensity, dense street network, concentration of public spaces, transport nodes, and institutional functions: the Circoscrizione is used as a pilot area to test a data preparation workflow that could later be extended to the entire municipal territory.

Its dense morphology makes discontinuities, mixed-use areas and classification errors more visible, making the district a demanding test bed for the workflow.

The methodology used in this thesis revolves around the sourcing of the data necessary from the relevant offices within the municipality and from the Territorial DataBase of the city, updated by the CSI (Consorzio Sistemi Informativi, Consortium of Informative Systems) after indication by the municipality of Turin.

This workflow system is mainly oriented towards replicability: what was done to Circoscrizione 1 can be replicated in any and all circoscrizioni as well as in other cities altogether.

4.1.2 Inventory and classification of data sources

The following table describes the main information in regard to the inventory and classification of the used data sources: as we can deduce the data has a variable update window, a constant Coordinate Reference System, different data file format and common body of origin:

Source	**Body**	**Format**	**CRS**	**Update**	**Use in the thesis**
Municipal territorial database	Municipality/ CSI	SHP/DBT	EPSG 3003	Trimestral	Official data basis
Crossings DWG files	Municipal offices	DWG	EPSG 3003	Variable	Extrapolation of crossing data
Orthophotos 2023	Municipality/ Regional authority	Raster	EPSG 3003	2023	Geometric verification
Corrective layer	Thesis elaboration	SHP/GeoJSON	Coherent	Manual	Missing connections

Table 1: Description of the main informations regarding inventory and classification of data sources used

4.1.3 Definition of destination data model

Before cleaning the data, the following data structure is to be intended as the final product:

Field	Meaning
id	Elemental univocal meaning
feature_type	crossing, sidewalk, pedestrian_area, stairs, traffic_island, connector
source_layer	Original layer
source_format	GeoJSON
crs	Reference system
geometry_status	valid, corrected, uncertain
connection_role	main_path, crossing, connector, barrier, access
quality_note	Notes on errors or uncertainties
update_required	yes / no

Table 2: description of the final product's fields

As is represented in the table, every field has a meaning and a reference, complying with the expected requirements from a vector dataset. Furthermore, the nine fields will contain the necessary data format (complete 32 bit or text) to efficiently communicate the importance of the relevant required data.

4.1.4 Selection and exclusion criterias

Selection is based on both semantic attributes and expected function in pedestrian continuity; not every pedestrian-accessible feature is necessarily retained if it does not contribute to network construction.

The following table will explain not only what layers were used, but also why and their selection criteria; this results in an efficient description of the workflow which includes replicable criterias for a future development:

Element	**Origin layer**	**Selection criteria**	**Motivation**
Pedestrian areas	010101_AREA_DI_CIRCOLAZIONE_VEICOLARE	All pedestrian elements	Walkable paths
Promiscuous vehicular areas	010102_AREA_DI_CIRCOLAZIONE_PEDONALE	Mixed pedestrian transit	Pedestrian accessible
Crossings	Municipal DWG	Conversion and verification on Otrophotos	Connections between road sides
Connecting cycle paths	010103_AREA_DI_CIRCOLAZIONE_CICLABILE	Classification = 1	only if needed for continuity
Traffic islands	020205_MANUFATTO_DI_INFRASTRUTTURA_DI_TRASPORTO	Manual selection / field classification	Intermediate crossing elements
Stairs	020206_AREA_ATTREZZATA_DEL_SUOLO	Type = stairs	Vertical or pedestrian connections

Table 3: used layers and motivation for use in the project

4.1.5 Operative workflow for data preparation

The workflow should be read as a reproducible protocol: each phase produces a layer, a set of attributes, or a validation decision that can be checked by another technician.

The following phases (A through H) will describe a replicable procedure for an operative workflow applicable to other circoscrizioni of the city.

Phase A - Data collection

- Official layer acquisition
- DWG acquisition
- Orthophoto 2023 acquisition
- Establishment of Circoscrizione 1 boundary

Phase B - Data format conversion
- Conversion from DWG to GeoJSON/SHP
- Loss of attributes verification
- Post-conversion geometry control
- Single format save (GeoPackage)

Phase C - CRS harmonization to EPSG 3003
- Project CRS definition
- Verification of original layer CRS
- Eventual transformation of CRS
- Orthophoto and DBT alignment check

Phase D - Geometry cleaning
- Duplicate elimination
- Correction of non-valid geometries
- Snap between near objects
- Check for gaps or overlaps
- Elimination of non-relevant geometries

Phase E - Attributive normalization
- Field name standardization
- Text values normalization
- Filling of missing fields
- Creation of classification fields
- Distinction between original attributes and additional attributes

Phase F - Layer integration
- Fusion of selected elements
- Management of overlaps and conflicts
- Creation of connective elements
- Connection of crossings, pedestrian areas and sidewalks

Phase G - Validation
- Comparison with orthophotos
- Manual verification of critical cases
- Topological and attributive check
- Error classification

Phase H - FInal output
- FInal pedestrian dataset
- Errors/corrections layer
- Metadata scheme
- Scheme for graph transformation

4.2 Destination data model

The project begins with an analysis of the City of Turin's territorial database, compiled according to the INSPIRE standard, and the road plans of the Traffic Services and Construction Site Coordination offices, created using AutoCAD software. From these plans, pedestrian crossings were extracted and then verified using the 2023 orthophotos.

The destination model is organized around feature classes, source lineage and graph-readiness attributes.

Pedestrian crossings (zebra crossings)

The pedestrian crossings in DWG format were converted to "geojson" format and divided into folders for each Circoscrizione named "AttraversamentiPedonali_CircN"

Ex:
AreePdonali_x_Liverpool\AttraversamentiPedonali\Circoscrizione1\AttraversamentiPedonali_Circ1.geojson

Territorial DB layers used (Inspire)

From the layer "01_Viabilità_Mobilità_E_Trasporti",
subsection "0101_Strade"

A) "010102_Area_Di_Circolazione_Pedonale"

This layer was taken in its entirety, because it provides a comprehensive depiction of spaces specifically dedicated to pedestrian movement, such as pedestrian streets, squares, and other traffic-free public areas. It was not filtered and as such all the elements are represented are pertinent to the workflow.

B) 010101_AREA_DI_CIRCOLAZIONE_VEICOLARE

The file has the following data structure:

Field name	**Data type**	**Description**	**Notes**
Surface	Complete (32 bit)	Numerical area	
Bottom	Text (string)	Material used for paving the area	
Site	Text (string)	Level with respect to the road	
Transit	Text (string)	Traffic type of the area	Exclusive, Not Applicable, Mixed Cycle Path, Mixed Pedestrian Path
Position	Text (string)	Location in relation to the road	
Level	Text (string)	Elevation In respect to the roadway	
Zone	Text (string)	Type of area	

Codice_cen	Complete (32 bit)	Territorial Census Information Code	
Codice_ar	Complete (32 bit)	Territorial Information Code of the area	

Table 4: data structure of the shapefile's attribute table

The following were extracted from this layer:

Areas with pedestrian transit (pedestrian areas) (all records with the value "*promiscuo pedonale*" were extracted from the *Transito* field).

The following file name is **AreaDiCircolazioneVeicolare_PromiscuoPedonale.geojson**

C) The file "**010103_AREA_DI_CIRCOLAZIONE_CICLABILE.shp**" was used to connect the various pedestrian route elements. On this layer, a specially created field (named Classification) was assigned the value "1" to select the routes that create connections.

This attribute then allows the extraction of only those records containing the value "1" and highlighted by the filter (filtered layer name "010103_AREA_DI_CIRCOLAZIONE_CICLABILE copia").

From the layer "02_IMMOBILI_ED_ANTROPIZZAZIONI",

subsection "0202_Manufatti"

D) 020205_MANUFATTO_DI_INFRASTRUTTURA_DI_TRASPORTO

From this layer, the records in the "**Tipo**" field with the attribute values "**passaggio pedonale aereo**," "**scala mobile**," and "**marciapiede, sagrato, piazza**" were used, selected using a "Filter"

(file name **020205_MANUFATTO_DI_INFRASTRUTTURA_DI_TRASPORTO.shp**)

The traffic islands were also extracted from the "type" field from this layer.

A new file "**SpartitrafficoDiConnessione_2.geojson**" was created to connect the various pedestrian path elements. On this layer, a specially created field (named Classification) has been assigned the value "1" to select traffic islands that create connections.

This attribute then allows the extraction of only those records containing the value 1, highlighted by the filter (filtered layer name "**SpartitrafficoDiConnessione_2 copy**")

From layer 02_IMMOBILI_ED_ANTROPIZZAZIONE

Subsection 0202_Manufatti

E) 020206_AREA_ATTREZZATA_DEL_SOLO

Stairways within the city limits, taken from layer code "**020206**" **Area_attrezzata_del_suolo**. The layer is filtered with the Value Type field "scale"

F) ElementiDiConnessione_PercorsiPedonali

A file **ElementiDiConnessione_GrafoPedonale.shp** was created to insert missing or incorrect connections in layers requiring correction (information entered in the note and note1 fields).

The project begins with an analysis of the City of Turin's territorial database, compiled according to the INSPIRE standard, and the road plans of the Traffic Services and Construction Site Coordination offices, created using AutoCAD software. From these plans, pedestrian crossings were extracted and then verified using the 2023 orthophotos.

4.3 Manual validation, graph and reproducibility

4.3.1 Connecting the dataset to the future pedestrian graph

The following logical passages were applied to the spatial data for its passage to the network, and as we will see all the elements have a corresponding ending element in the graph:

Spatial dataset	**Pedestrian graph**
Crossings	Connective arches between road sides
Sidewalks / pedestrian areas	Walkable arches
Intersections	Nodes
Stairs / passageways	Special arches
Traffic islands	Nodes or micro-arches for crossing
Missing elements	Documented artificial connectors

Table 5: Applied logical passages

Each polygonal feature must be converted into graph edges through a defined simplification or centerline rule, while contact points and crossing endpoints generate graph nodes.

The prepared dataset represents the spatial basis for a future graph structure: in this structure, pedestrian-accessible linear or polygonal elements can be converted into edges, while intersections, crossings, and access points can be represented as nodes.

The present thesis focuses on the preparation and validation of the spatial input data, while the full network modelling phase is identified as a future development.

4.3.2 Improving reproducibility

A fundamental objective of this research was to develop a workflow that could be easily reproduced by other GIS technicians or municipal authorities working on similar datasets. For this reason, the data preparation process was designed as a sequence of standardized and repeatable operations that can be applied to other study areas with minimal modifications.

Because of this the following workflow will explain the process used:

–> *Study area clipping*
–> *Format conversion*
–> *CRS harmonization*
–> *Attribute filtering*
–> *Geometry cleaning*
–> *Integration of pedestrian elements*
–> *Creation of connection elements*
–> *Quality assessment*
–> *Final pedestrian infrastructure dataset*
–> *Preparation for future pedestrian graph*

By documenting each stage of the process and employing widely available open-source software such as QGIS, the proposed workflow can be readily replicated, adapted, and extended to other districts of Turin or to different municipalities facing similar challenges in the preparation of pedestrian infrastructure data. This reproducible methodology contributes to the standardization of municipal geospatial data preparation and facilitates future updates, maintenance, and interoperability within Spatial Data Infrastructures (SDIs).

Reproducibility also requires recording software version, CRS, snap tolerance, field names, filter expressions and manual editing criteria.

For precise indications as to the process and layers used see point 4.2 “Destinational data model”.

4.3.3 Instruments and operative choices

In the workflow, QGIS was used for filtering, format conversion, CRS verification, geometry validation, snapping, visual comparison with orthophotos and final layer export.

One of the main strengths of QGIS is its compatibility with numerous geospatial data formats, largely made possible through the GDAL/OGR library: among the most commonly supported vector formats are GeoPackage, Esri Shapefile, GeoJSON, and KML, while raster data can be managed through formats such as GeoTIFF.

QGIS also provides direct integration with spatial databases, including PostgreSQL/PostGIS and SpatiaLite, as well as with Open Geospatial Consortium (OGC) web services such as Web Map Service (WMS) and Web Feature Service (WFS).

In the context of this research, GeoJSON was the primary format used for data preparation and processing due to its simplicity, flexibility, and ease of integration with GIS workflows, in addition, WMS services were employed to access and visualize the base cartographic layers of the **Municipal Topographic Database (DBT)**, providing a reliable reference for data validation and spatial analysis.

5. Case study: Circoscrizione 1

5.1 Study area

The principal study area for this dissertation is the central district of the city, Circoscrizione 1.

Circoscrizione 1 represents the central administrative district of the city of Turin and constitutes the historical, institutional, and economic core of the municipality: the Circoscrizione includes the historic center and adjacent neighborhoods, particularly Centro and Crocetta, which together form one of the most densely built and functionally diverse areas of the city. This area hosts a large concentration of administrative institutions, cultural landmarks, commercial activities, and services, making it a key hub for urban mobility and daily commuting patterns.

From an urban planning perspective, Circoscrizione 1 is characterized by a compact urban morphology, a dense street network, and a high level of pedestrian activity.

The historic urban fabric, largely shaped between the seventeenth and nineteenth centuries, features a regular grid pattern with arcaded streets and large public squares such as Piazza Castello and Piazza San Carlo, which function as important nodes for pedestrian circulation and urban life; the presence of major transport corridors, public transit lines, and significant pedestrian flows makes the Circoscrizione particularly relevant for studies related to mobility and pedestrian infrastructure.

In addition, the area hosts several key cultural and institutional landmarks, which attract both residents and tourists and further contribute to high levels of pedestrian movement: the combination of historic urban structure, intense mobility flows, and diverse land uses makes Circoscrizione 1 a particularly suitable case study for analyzing the preparation and integration of mobility and pedestrian infrastructure data within municipal geospatial databases.

The coexistence of arcades, squares, mixed traffic streets and public transport nodes makes the area particularly suitable for testing pedestrian network continuity.

It's important to understand the pedestrian routes for the project with the University of Liverpool: specifically, it's important to identify the characteristics of the pedestrian routes and their traffic, this analysis helps improve pedestrian mobility in Turin.

5.2 Layer selection

The following table describes not only the sources of the data used, but also the elements obtained by filtering the layers and the files that were produced as a final output of the workflow; this is useful for a number of reasons, the main being that we can obtain an overview of the input files and their resulting output results:

Starting layer	**Applied filter**	**Pedestrian element obtained**	**File produced**
010102_Area_Di_Circolazione_Pedonale	All the records	Pedestrian areas	010102_Area_Di_Circolazione_Pedonale.GeoJSON
010101_AREA_DI_CIRCOLAZIONE_VEHICOLARE	Transit = Mixed pedestrian	Vehicular areas with pedestrian transit	AreaDiCircolazioneVeicolare_PromiscuoPedonale.geojson
010103_AREA_DI_CIRCOLAZIONE_CICLABILE	Classified N = 1	Connectors on cyclable path	010103_AREA_DI_CIRCOLAZIONE_CICLABILE copia
020205_MANUFATTO_DI_INFRASTRUTTURA_DI_TRASPORTO	Type = sidewalk, escalator, aerial passage	Sidewalk, escalator, aerial passage	020205_MANUFATTO_DI_INFRASTRUTTURA_DI_TRASPORTO.shp
020206_AREA_ATTREZZATA_DEL_SOLO	Type = stairs	Stairs	020206_AREA_ATTREZZATA_DEL_SOLO.shp
Traffic island	Classified n = 1	Connecting traffic island	SpartitrafficoDiConnessione_2
(Manually created)	note / note1	Missing or erroneous connections	ElementiDiConnessione_GrafoPedonale.shp

Table 6: Sources selection synthesis

5.3 Elaborations and corrections

Duplicate records—whether identical points or repeat attributes—can inflate counts, skew statistics, and clutter visualizations, in geospatial data, duplicates might arise from merging datasets, human error, or oversampling.

Duplicate detection was performed by comparing coincident geometries and repeated attributes after conversion and clipping to the study area.

Errors may also arise from incomplete datasets or simple mistakes in the data input process, an example is represented in the following figure 5.1:

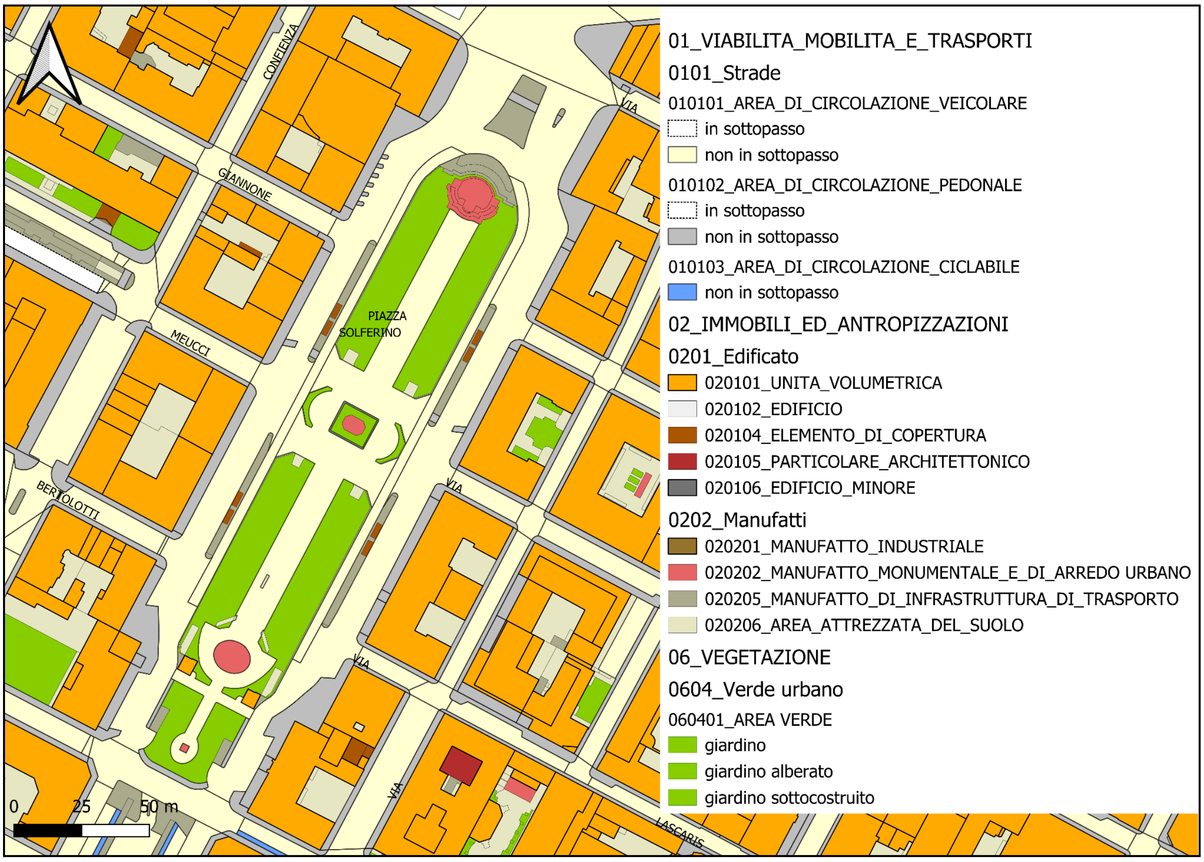

Figure 5.1: Piazza Solferino visualized as part of the INSPIRE database

In the examined case Piazza Solferino, in Turin, was signed as a vehicular circulation area, while in reality it’s reserved for pedestrian traffic: a singular mistake of classification on a central piazza alters the connectivity of a significant portion of the network, which justifies a systematic and non samplicistic quality control before building the pedestrian graph.

The case demonstrates why semantic validation is necessary: a geometrically valid feature can still be unsuitable for pedestrian graph construction if its mobility classification is wrong.

Figure 5.2: highlighting in brown color of the erroneous qualities of the area (the central part of the Piazza is reserved for pedestrian use but is marked as vehicularly accessible, as is deductible from the legend)

The data is provisionally corrected by hand by the office and highlighted as a mistake to be corrected in the next update window by the CSI.

The following figure 5.3 and figure 5.4 show and area of interest to the south of Piazza Solferino where mistakes were present and corrected manually:

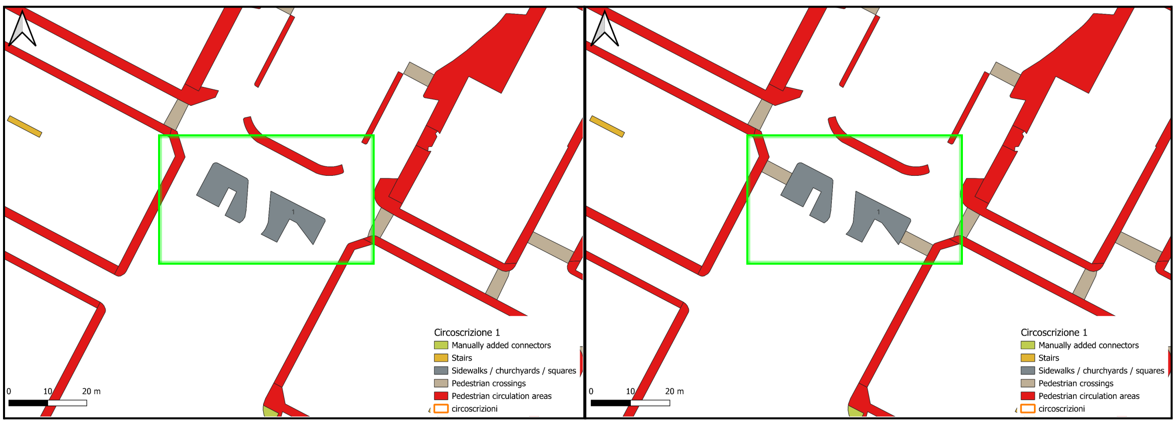

Figure 5.3 (left): before the correction, missing crossings and Figure 5.4 (right): after the correction, crossings marked

5.4 Results

In this chapter, results are reported for the graph-ready dataset, while full graph construction remains a subsequent modelling phase.

This thesis defines a “pedestrian graph” as “A topologically connected network dataset representing pedestrian-accessible infrastructure, composed of nodes and edges describing walkable spaces and their relationships, enabling routing, accessibility analysis, and urban mobility assessment”.

The results of the case study in the city of Turin demonstrate that the preparation and integration of mobility and pedestrian infrastructure data into a municipal cartographic database significantly enhance the capacity for spatial analysis and urban planning.

The following table describes not only the elements present in the project, but also their numerical amounts and the sources of the data, which is useful to obtain an overview over the numerical presence of the different elements:

Element	**Number (Circ. 1)**	**Source / note**
Pedestrian circulation areas	5.628	010102 (extracted)
Vehicular areas with pedestrian circulation	153	010101, filtered for mixed pedestrian use
Pedestrian crossings	2.192	Extracted for GeoJSON and DWG files, along with orthophotos
Sidewalks / churchyards / squares	165	020205, filter by type
Stairs	1.347	020206, filtered as Stairs
Connective traffic islands	4.589	Classification = 1
Manually added connectors	1.022	ElementiDiConnessione_GrafoPedonale
Corrected classification errors	1	Piazza Solferino
Cases marked for CSI	1	Layer for internal use by offices

Table 7: numerical amount of each element present within the circoscrizione

The final dataset for Circoscrizione 1 contains 15,096 selected or created elements. Pedestrian circulation areas represent the largest share of the dataset, followed by connective traffic islands, pedestrian crossings, stairs and manually added connectors. This composition shows that pedestrian movement in the central district cannot be represented only through conventional sidewalks or crossings: several categories of public space contribute to pedestrian continuity, including squares, mixed-use areas and intermediate crossing elements.

The presence of 1,022 manually added connectors is particularly relevant. These connectors account for approximately 6.8% of the total elements and indicate that official layers, although rich and detailed, do not always provide a directly continuous structure for graph construction. Manual connectors should therefore be treated as documented hypotheses of continuity: they improve the usability of the dataset for future network modelling, but they also require explicit quality notes and, where necessary, validation by municipal offices.

The correction of Piazza Solferino illustrates the difference between geometric validity and semantic correctness. The area was represented as a vehicular circulation area, although its actual use includes pedestrian functions. Such errors are critical for graph construction because they can interrupt or distort the representation of walkable space even when the geometry itself is valid. For this reason, semantic validation is as important as geometric correction in the preparation of pedestrian mobility data.

The results therefore show both the potential and the limitations of the existing municipal data. On the one hand, the official database provides a strong basis for pedestrian infrastructure mapping, on the other hand, the future pedestrian graph will require additional topological processing, including node generation, edge creation, snapping rules, treatment of traffic islands, and classification of inferred connectors.

In the following table the mathematical amount of elements and their ratio to the total amount of features are displayed for an effective representation and interpretation of their properties.

Indicator	**Formula**	**Value**	**Interpretation**
Total selected elements	Sum of final features	15.096	It measures the operative size of the dataset
Manual connector ratio	Manual connectors / total elements	6,8%	It indicates how much topological continuity work was necessary.
Crossing share	Crossings / total elements	14,5%	Shows the weight of crossings in pedestrian connectivity.
Traffic island share	Traffic islands / total elements	30,4%	Highlights the role of intermediate crossing elements.

Table 8: Resulting ratio of components to total amount of features

The following figures, 5.5 and 5.6, represent the area of the circoscrizione 1 of the city of Turin and a focal zoom to an area of interest, this is to show both the extent of the circoscrizione and the area of interest present within its boundaries:

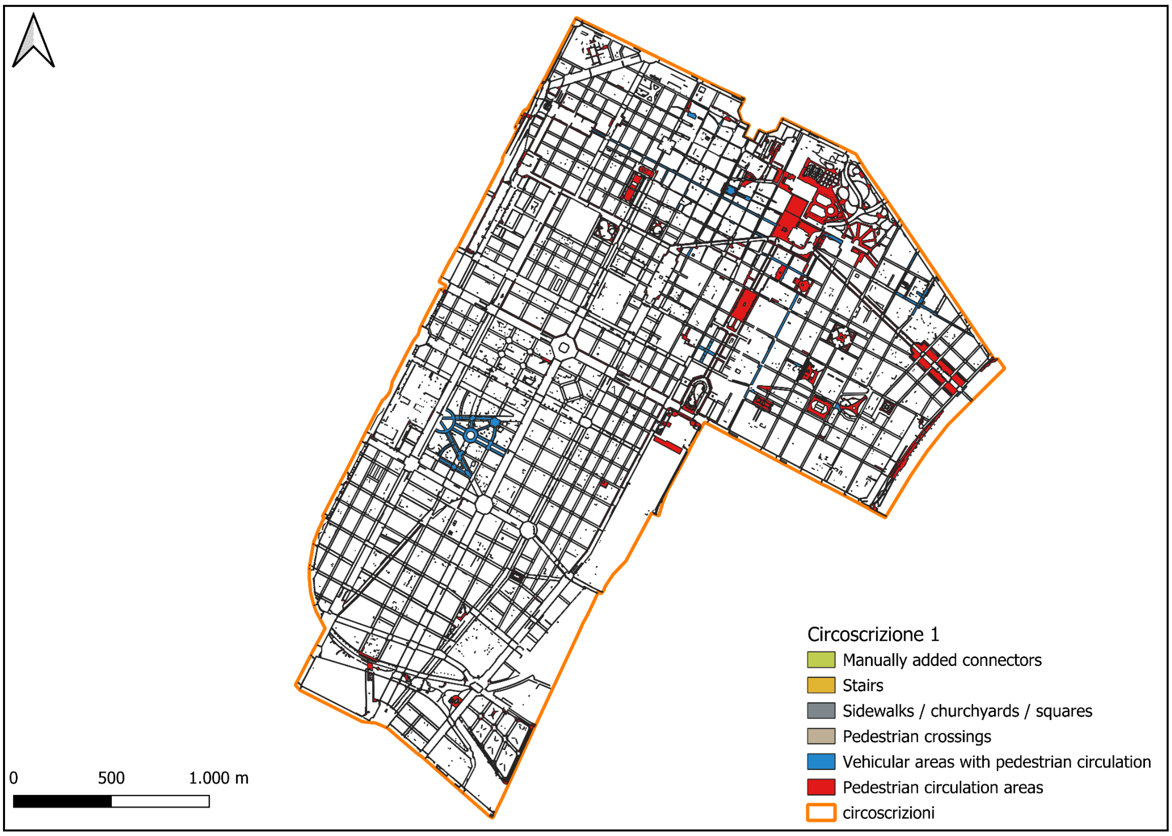

Figure 5.5: The Circoscrizione 1 represented with borders and the aforementioned elements pre - elaboration

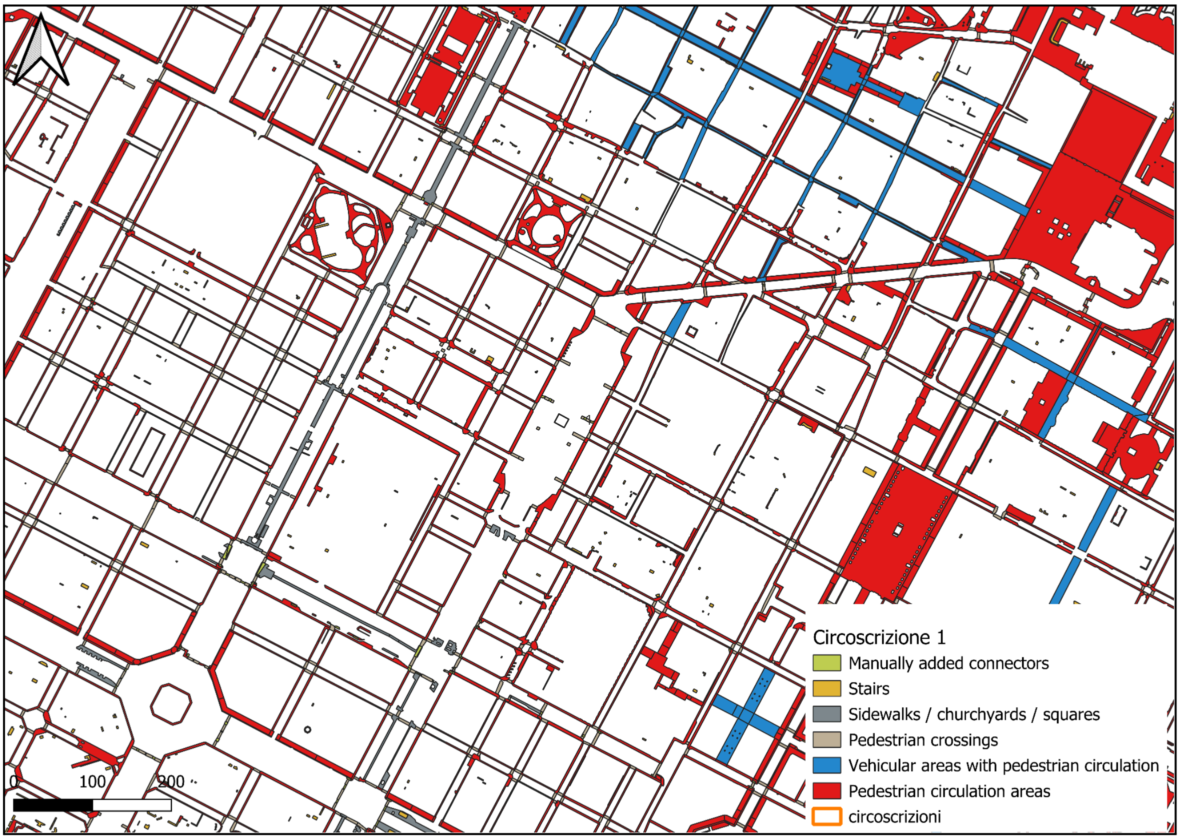

Figure 5.6: the area of Piazza Solferino

Critical points are concentrated along the axis with mixed traffic, marking that the criticality is born from the classification vehicular/pedestrian and not from casual mistakes, another criticality found is that the discontinuities are concentrated around complex intersections and on ordinary pedestrian traffic areas, where the starting data is less coherent; that is the point where the municipal database requires more maintenance.

5.4.1 How the prepared dataset enables the pedestrian graph

The standardized dataset developed through this research represents an essential first step toward the creation of a comprehensive pedestrian graph for the City of Turin: by harmonizing heterogeneous pedestrian infrastructure data according to the INSPIRE Transport Networks schema, the study establishes a consistent and interoperable spatial framework in which sidewalks, pedestrian areas, crossings, and related mobility features are represented using common definitions and attributes.

However, the resulting dataset does not yet constitute a fully functional pedestrian graph: several additional steps are required, including the creation of a topological network structure, the identification and validation of nodes and edges, the explicit modelling of pedestrian crossings and connectivity between infrastructure elements, and the incorporation of routing attributes such as walking distance, travel time, accessibility constraints, and barrier information.

Further work would also be needed to verify network continuity through field surveys and quality-control procedures, ensuring that all pedestrian links are correctly connected and navigable: once these processes are completed, the dataset could serve as the foundation for a detailed pedestrian graph capable of supporting route calculation, accessibility analysis, walkability assessment, and sustainable mobility planning within the city of Turin.

In the following figure 6.1 we can see an initial draft of a pedestrian graph created using the QGis elaborated datasets in the area of Porta Nuova in Turin, we can infer from the image that the numerous pedestrian passages create a network usable by pedestrians to move around the area without the danger of vehicular traffic.

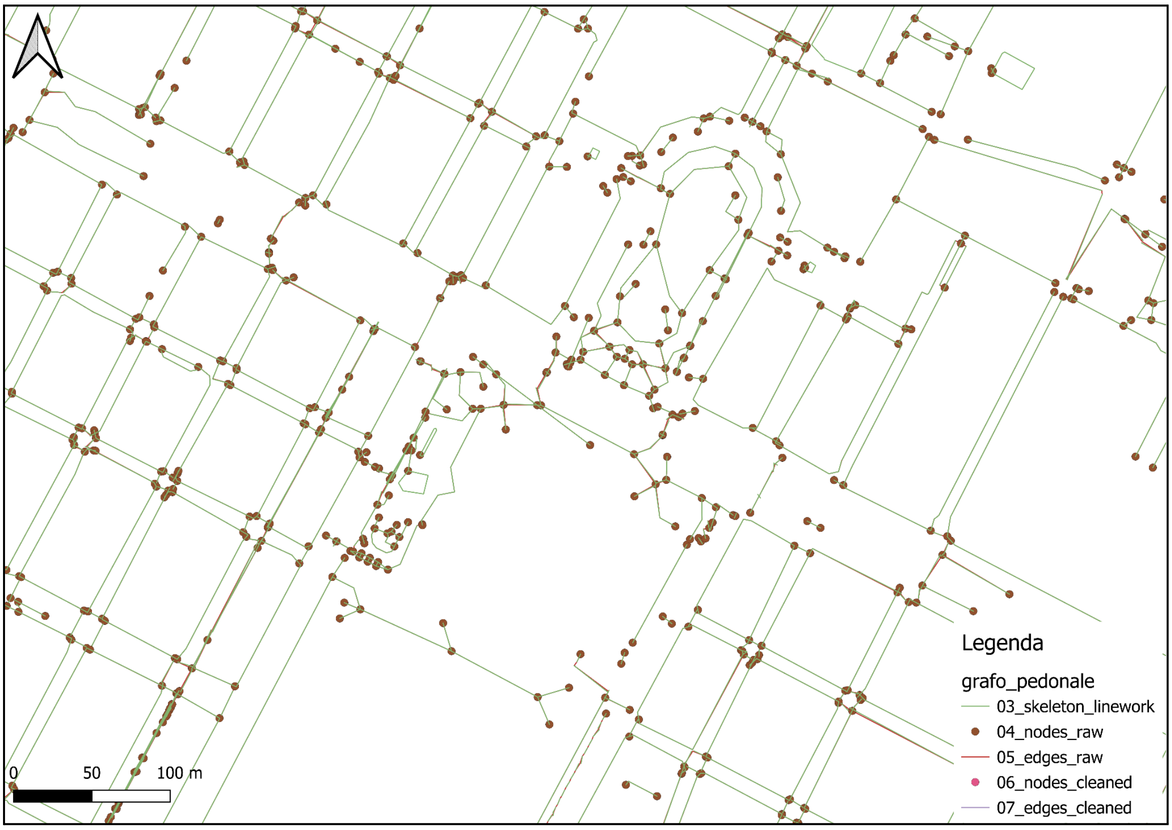

Figure 5.7: Preliminary graph representation in the Porta Nuova area, distinguishing original pedestrian features, inferred connectors and potential graph nodes.

5.4.2 Proposed pedestrian graph model

Although the objective of this research was not the construction of a complete pedestrian graph, the standardized dataset produced through the data preparation workflow provides a solid foundation for its future development. A pedestrian graph requires a topologically connected representation of the pedestrian environment, where infrastructure elements are modelled as interconnected nodes and edges enriched with descriptive attributes. Based on the results of this study, the proposed data model consists of four main components, summarized in the following Tables.

Nodes	Type	Meaning
node_id	Complete (32 bit)	Node's stable identifier
geometry	Complete (32 bit)	Node's coordinates in EPSG 3003
node_type	Text (string)	intersection, crossing_endpoint, access_point, traffic_island, stair_landing, artificial_connector
source_feature_id	Complete (32 bit)	ID of the source object from which the node derives
source_layer	Text (string)	Original or corrective layer
level	Text (string)	Level/elevation if available
quality_status	Text (string)	validated, corrected, inferred, uncertain
update_required	Text (string)	Indicates if a check is needed from offices/CSI

Table 9: Nodes properties table

Edges	Type	Meaning
edge_id	Complete (32 bit)	Edge's stable identifier
from_node	Complete (32 bit)	Original node
to_node_from_node	Complete (32 bit)	Destination node
geometry	Text (string)	Simplified or derivative
edge_type	Text (string)	sidewalk, crossing, pedestrian_area_axis, stairs, bridge, underpass, traffic_island_link, connector
lenght_m	Complete (32 bit)	Metrical length of the edge
bidirectional	Text (string)	Normally true, barring specific constraints
acessibility_status	Text (string)	Accessible, limited, stairs, unknown
surface	Text (string)	Material/floor if available
lighting_id	Text (string)	Future connection to illumination database
source_layer	Text (string)	Source layer
quality_status	Text (string)	Validated, corrected, inferred, uncertain
impedance	Text (string)	Future travel cost: length, time, safety, accessibility

Table 10: Edges properties table

Lineage	**Type**	**Meaning**
source_id	Complete (32 bit)	Original data source or file
feature_id	Complete (32 bit)	Original feature identifier
graph_object_id	Complete (32 bit)	Node or edge produced from transformation
transformation_rule	Text (string)	Example: polygon_to_centerline, crossing_endpoint_generation, snap_connector
operator_note	Text (string)	Manual note of doubt, correction or check
date_processed	Complete (32 bit)	Date of transformation

Table 11: Lineage properties table

Validation_issues	**Type**	**Meaning**
issue_id	Complete (32 bit)	ID of the identified problem
object_id	Text (string)	Feature, node or edge affected
issue_type	Text (string)	missing_crossing, wrong_classification, gap, overlap, invalid_geometry, uncertain_attribute
severity	Text (string)	Low, medium, high
action	Text (string)	corrected, provisional_connector, reported_to_CSI, left_uncertain
evidence	Text (string)	Orthophoto, DBT, DWG, manual inspection
note	Text (string)	Case description

Table 12: Issue validation properties table

5.5 Dataset limitations

The main limitation of the dataset is that it improves graph readiness but does not yet guarantee routability. Some connections are derived from visual interpretation or manual correction, and therefore they must be flagged as inferred elements. In addition, polygonal features such as squares or pedestrian areas still require a transformation rule before they can become graph edges. The dataset also depends on the update cycle of the municipal DBT and on the validation of reported issues by the responsible offices. These limitations do not weaken the usefulness of the workflow; rather, they identify the technical steps required to move from a validated infrastructure dataset to an operational pedestrian graph.

6. Tools and Technologies

6.1 Spatial Data Infrastructures, interoperability and INSPIRE

6.1.1 Spatial Data Infrastructure

A spatial data infrastructure (SDI), also called geospatial data infrastructure, is a data infrastructure implementing a framework of geographic data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way. Another definition is "the technology, policies, standards, human resources, and related activities necessary to acquire, process, distribute, use, maintain, and preserve spatial data" (The White House, Office of Management and Budget, 2002): most commonly, institutions with large repositories of geographic data (especially government agencies) create SDIs to facilitate the sharing of their data with a broader audience.

6.1.2 Interoperability

Interoperability is one of the fundamental principles of a Spatial Data Infrastructure (SDI), as it enables geospatial data produced by different organizations to be shared, integrated, and used together effectively; three dimensions of interoperability are particularly important: semantic, geometric, and organizational interoperability.

“Semantic interoperability” refers to the ability of different datasets to convey the same meaning through common definitions, classifications, and data models: for example, a road segment or pedestrian crossing should be interpreted consistently regardless of the institution that produced the data (European Commission, 2016).

“Geometric interoperability” concerns the spatial consistency of datasets, ensuring that geographic features align correctly when combined, despite originating from different sources or coordinate systems: this requires the use of common reference systems, data specifications, and quality standards to avoid spatial mismatches and inaccuracies (INSPIRE Thematic Working Groups, 2013).

Finally, “Organizational interoperability” relates to the institutional arrangements, policies, and collaborative processes that allow different organizations to exchange and manage geospatial information efficiently: it involves agreements on data sharing, responsibilities, maintenance procedures, and governance structures that support long-term cooperation among stakeholders (Masser, 2005).

Together, these three forms of interoperability form the basis of modern SDIs and are essential for enabling the integration of spatial data across administrative boundaries and sectors, particularly within frameworks such as INSPIRE (European Commission, 2016).

6.1.3 INSPIRE standard and its limitations

INSPIRE is a European Union (EU) Directive that came into force on May 15, 2007 with the aim to create a common spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment.

One of the major goals is to create harmonised spatial data sets that can be used seamlessly in cross-border applications: in order to reach this goal common data specifications have been agreed for 34 themes as well as technical requirements for their publication via network services.

Recent policy discussions have however emphasized the need to simplify and modernize INSPIRE implementation, especially for data reuse and integration with newer data ecosystems.

INSPIRE was adopted in 2007, and many practitioners have criticized:

- its complex data models;
- the high cost of compliance;
- the effort required to transform local datasets into INSPIRE schemas;
- its reliance on XML/GML formats, which are often perceived as cumbersome compared to modern web-based formats such as GeoJSON.

These criticisms are particularly relevant for municipalities, where local data structures are often simpler than the INSPIRE conceptual model.

The European Commission is currently revising and simplifying INSPIRE through the GreenData4All initiative: the objective is to modernize the directive, reduce administrative burdens, align it with newer EU data legislation, and improve access to geospatial data, the emphasis is on modernization rather than replacement.

In 2025–2026, the Commission formally proposed revisions to simplify the directive and align it with contemporary technologies and European data spaces.

The INSPIRE Directive was transposed into Italian law with Legislative Decree No. 32 of January 27, 2010, establishing the National Infrastructure for Spatial Information and Environmental Monitoring in Italy as a hub of the EU infrastructure.

The Ministry of the Environment, Energy Security and the Sea is the competent authority for the implementation of Legislative Decree No. 32/10, and the National Contact Point for the INSPIRE Directive has been established within the Ministry.

6.2 Privacy and accessibility

6.2.1 Privacy and data accessibility issues

Privacy and data accessibility represent significant challenges in the management of geospatial data within municipal cartographic databases: in urban contexts, datasets related to mobility and pedestrian infrastructure may include sensitive information, particularly when derived from GPS traces (such as in this case, which are derived by Phone Carriers), mobile devices, or sensor-based systems, raising concerns about the identification and tracking of individuals (Kitchin, 2014).

The present thesis does not process individual mobility traces; privacy issues are discussed as relevant for future applications that may combine the pedestrian graph with mobility-flow data.

In the case of the city of Turin, ensuring compliance with European data protection regulations, such as the General Data Protection Regulation (GDPR), is essential when collecting, processing, and sharing mobility-related data (European Union, 2016): this often requires the anonymization or aggregation of datasets before their integration into municipal geospatial databases or publication through open data platforms.

However, balancing openness with privacy protection can limit the level of detail or completeness of publicly available datasets, potentially affecting their usability for detailed spatial analysis (Janssen et al., 2012), therefore, in the Turin case study, managing privacy and accessibility involves adopting strategies that ensure both compliance with legal frameworks and the provision of usable, high-quality geospatial data for urban planning purposes.

7. Conclusions

7.1.1 Summary of findings

The case study of the city of Turin highlights the complexity and importance of preparing geospatial data for integration into municipal cartographic databases, particularly when dealing with pedestrian mobility and infrastructure: the research shows that data preparation is a multi-step process involving collection from heterogeneous sources, cleaning, georeferencing, standardization, and alignment with the INSPIRE data structure, one of the main findings is that data quality—especially in terms of completeness, accuracy, and consistency—plays a crucial role in determining the effectiveness of the resulting database for urban planning applications.

The study produced a graph-ready dataset composed of selected official features, CAD-derived crossings and manually documented connectors, revealing both data potential and maintenance needs.

The study also reveals that interoperability challenges arise when integrating datasets from different administrative levels, requiring significant efforts in data harmonization and schema transformation, furthermore, the use of GIS tools proves essential for automating workflows and managing large datasets efficiently.

Overall, the case study demonstrates that well-structured and standardized geospatial databases can significantly enhance urban analysis, supporting more informed and sustainable planning strategies in Turin while also proving essential for more in-depth analysis by external agents such as universities or research entities.

7.1.2 Answers to the research questions

The answer is operational: data are prepared through selection, conversion, CRS harmonization, semantic filtering, geometric cleaning, connection creation and quality assessment.

The research shows that mobility data and pedestrian infrastructure data can be effectively prepared for integration into the municipal geospatial database of the city of Turin through a structured process aligned with the INSPIRE data framework: this process begins with the collection of data from heterogeneous sources, including municipal datasets, regional spatial data infrastructures, and open data platforms; these datasets are then subjected to data cleaning procedures to remove errors, inconsistencies, and duplicates, followed by georeferencing to ensure alignment within a common coordinate reference system.

Data integration then involves merging multiple datasets, resolving discrepancies, and establishing relationships between spatial entities such as road networks and pedestrian paths: throughout this process, particular attention must be paid to data quality dimensions such as accuracy, completeness, and consistency, as well as to compliance with metadata requirements defined by INSPIRE (European Commission, 2016).

Overall, the study demonstrates that the preparation of mobility and pedestrian infrastructure data following the INSPIRE data structure enables the creation of interoperable, reliable, and scalable municipal geospatial databases, which are essential for supporting informed urban planning and decision-making processes in Turin while also providing a base for further research by universities or other research entities.

7.1.3 Implications for urban planning

The preparation and integration of high-quality geospatial data into municipal cartographic databases have significant implications for urban planning processes, particularly in complex urban environments such as Turin: reliable and well-structured mobility and pedestrian infrastructure data enable planners to perform more accurate spatial analyses, supporting evidence-based decision-making in areas such as accessibility, traffic management, and sustainable mobility planning (Longley et al., 2015).

For municipal planning, the main value is not only routing but also the ability to identify discontinuities, prioritize database maintenance and support pedestrian accessibility policies.

In the Turin case study, the alignment of local datasets with standardized frameworks such as INSPIRE enhances interoperability and facilitates data sharing between institutions, improving coordination across different levels of governance.

Furthermore, the availability of updated and integrated geospatial information allows for better identification of critical issues, such as gaps in pedestrian networks or areas with high congestion, enabling targeted urban interventions (Geurs & van Wee, 2004), however, the effectiveness of these applications depends on the quality, completeness, and consistency of the underlying data, highlighting the importance of robust data preparation processes: overall, the development of structured cartographic databases contributes to more efficient, transparent, and sustainable urban planning practices in the city of Turin.

This thesis demonstrates that the construction of a pedestrian graph is not only a network-modelling task, but also a data preparation and data governance problem. In the case of Turin, the availability of detailed municipal layers provides a strong starting point, but the transition toward routing and mobility modelling requires additional harmonization, semantic validation, topological correction and documentation of uncertain elements. The proposed workflow contributes to this transition by producing a graph-ready pedestrian infrastructure dataset and by defining the data model requirements for its future transformation into a complete pedestrian network.

8. Bibliography

8.1 Academic literature

- Sinjari, Sonila & Kosovrasti, Albana. (2015). Geographic Information Systems (GIS) in Urban Planning. European Journal of Interdisciplinary Studies. 1. 85. 10.26417/ejis.v1i1.p85-92.
- P.Dave, Chintan & Joshi, Rahul & Srivastava, S.. (2015). A Survey on Geometric Correction of Satellite Imagery. International Journal of Computer Applications. 116. 24-27. 10.5120/20389-2655.
- Moon, Wooil. (2012). Geographic Information Systems and Science (3rd Edition) by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (Book Review). The Leading Edge (Society of Exploration Geophysicists). 31. 975 - 976.

8.2 Online sources

Vector data in GIS software, GIS
- https://docs.qgis.org/3.44/en/docs/gentle_gis_introduction/vector_data.html

Raster data in GIS software, GIS
- https://docs.qgis.org/3.44/en/docs/gentle_gis_introduction/raster_data.html

Geographic Information Science & Technology Body of Knowledge Visualization and Search
- https://gistbok-topics.ucgis.org/DC-03-026

GIS data acquisition
- https://www.gisci.org/Want-to-be-a-GISP/The-Exam/Exam-Preparation-Info/Study-Guide/Data-Acquisition

Data Types and Spatial Data Models
- https://www.opengeomatics.ca/types-of-data.html#types-of-data-1

What Is Metadata? Definition, Types, and Uses
- https://www.salesforce.com/data/what-is-metadata/#:~:text=What%20is%20the%20main%20purpose,use%20data%20quickly%20and%20effectively.

Standards | Metadata | ISO Geospatial Metadata Standards
- https://www.fgdc.gov/metadata/iso-standards

Coordinate Reference Systems, QGis
- https://docs.qgis.org/3.44/en/docs/gentle_gis_introduction/coordinate_reference_systems.html

Working with Projections in QGIS
- https://medium.com/introduction-to-critical-spatial-media-beyond-the/working-with-projections-in-qgis-f675405d6778

Database di Sintesi Nazionale
- https://www.igmi.org/it/dbsn-database-di-sintesi-nazionale