Toward Data Integration for Representing Urban Assets: a Case Study of Ferrara, Emilia-Romagna, Italy

L'integrazione di diverse tipologie di dati sorgente per la rappresentazione dell'ambiente urbano: un caso studio a Ferrara

Fabio Planu
PhD(c) IDAUP – Dottorato Internazionale Architettura & Pianificazione Urbana, Università di Ferrara | fabio.planu@unife.it

Gabriele Giau
PhD(c) IDAUP – Dottorato Internazionale Architettura & Pianificazione Urbana, Università di Ferrara | gabriele.giau@unife.it

The need to address the social and environmental challenges of cities has led to the growing international application of digital technologies for data management in smart cities. These efforts have been accompanied by the development of methodological approaches and modeling tools aimed at delivering concrete benefits to urban planning and management. City Information Modelling (CIM) is emerging as a promising approach for the holistic representation of urban environments, facilitating interaction between users and digital models of urban contexts. Today, the source data available for such purposes is often fragmented and heterogeneous. To ensure effective knowledge and analysis, this data must be represented in a critical and interpretative framework. This contribution, through the case study of an urban section in Ferrara, demonstrates some of the results achieved by the research group at the TekneHub Laboratory and the DIAPReM Departmental Research Centre. The study focuses on the creation of an information system that unifies and organizes data from diverse sources.

La necessità di affrontare le sfide sociali e ambientali delle città, ha determinato a livello internazionale la crescente applicazioni di tecnologie digitali per la gestione dei dati nell'ambito delle smart city e lo sviluppo di approcci metodologici e strumenti di modellazione finalizzati ad apportare benefici concreti alla pianificazione e alla gestione urbana. Il City Information Modelling (CIM) si caratterizza come un approccio promettente per la rappresentazione olistica dell'ambiente in cui avviene l'interazione tra utente e modelli digitali di contesti urbani. Il dato sorgente potenzialmente disponibile oggi, frammentato ed eterogeneo, necessita quindi di essere rappresentato ad una dimensione critica ed interpretativa per garantire livelli di conoscenza e analisi. Il presente contributo, attraverso il caso studio di una sezione urbana a Ferrara, intende mostrare alcuni dei risultati ottenuti dal gruppo di ricerca del Laboratorio TekneHub e del Centro di ricerca dipartimentale DIAPReM sulla creazione di un sistema informativo che unifica e organizza le informazione a partire da differenti tipologie di dato sorgente disponibile.

INTRODUCTION

The transition from traditional city management and planning methods to modern systems is increasingly driven by the need for efficient resource use, alongside greater attention to quality of life and urban environments [1]. Challenges related to social, financial, and environmental sustainability compel decision-makers to optimize the management of complex urban systems, which relies heavily on the availability and effective utilization of data. In recent years, there has been a growing awareness among administrations, development agencies, and municipalities about the importance of adopting innovative technologies for urban documentation and analysis. This shift highlights the need for a multidisciplinary and multi-scalar visualization of urban assets, employing a holistic perspective that integrates qualitative and quantitative approaches. Such an approach can deepen understanding and enable monitoring of urban transformations, providing the foundation for more informed management tools [2]. Urban information modeling at various levels — from city-scale models to those of individual buildings — has become a central theme in many smart city projects [3], reflecting advancements in international research across diverse fields.

To initiate meaningful digital transformation, cities must undertake significant efforts to consolidate existing building data from disparate databases. Collecting and structuring data for an entire city or district remains a challenging task, largely due to the lack of integration between databases that were initially created independently for specific purposes without consideration for broader information management goals [4].

In this context, there is growing interest in City Information Modeling (CIM). CIM is widely regarded as a digital representation of a city [5], enabling the identification of optimal strategies to enhance urban environments. It serves as a repository for extensive urban data, encompassing both static models and dynamic objects [6]. By integrating diverse data sources, technologies, and analytical tools, CIM provides a comprehensive framework for visualizing, analyzing, and managing urban environments holistically. This approach aims to empower stakeholders to make informed decisions, optimize resource allocation, and improve urban liveability and sustainability [7].

The literature suggests that a viable and cost-effective strategy involves integrating Building Information Modeling (BIM) with Geographic Information Systems (GIS) — commonly referred to as the BIM-GIS approach [8]. For creating As-Built urban models, a promising method is the Scan-to-BIM process, which ensures accurate representations of the built environment. In this process, photogrammetric models play a crucial role by enhancing the morphological accuracy of CIM. These models use photographic images to extract reliable information about the physical elements of the urban landscape.

OBJECTIVE

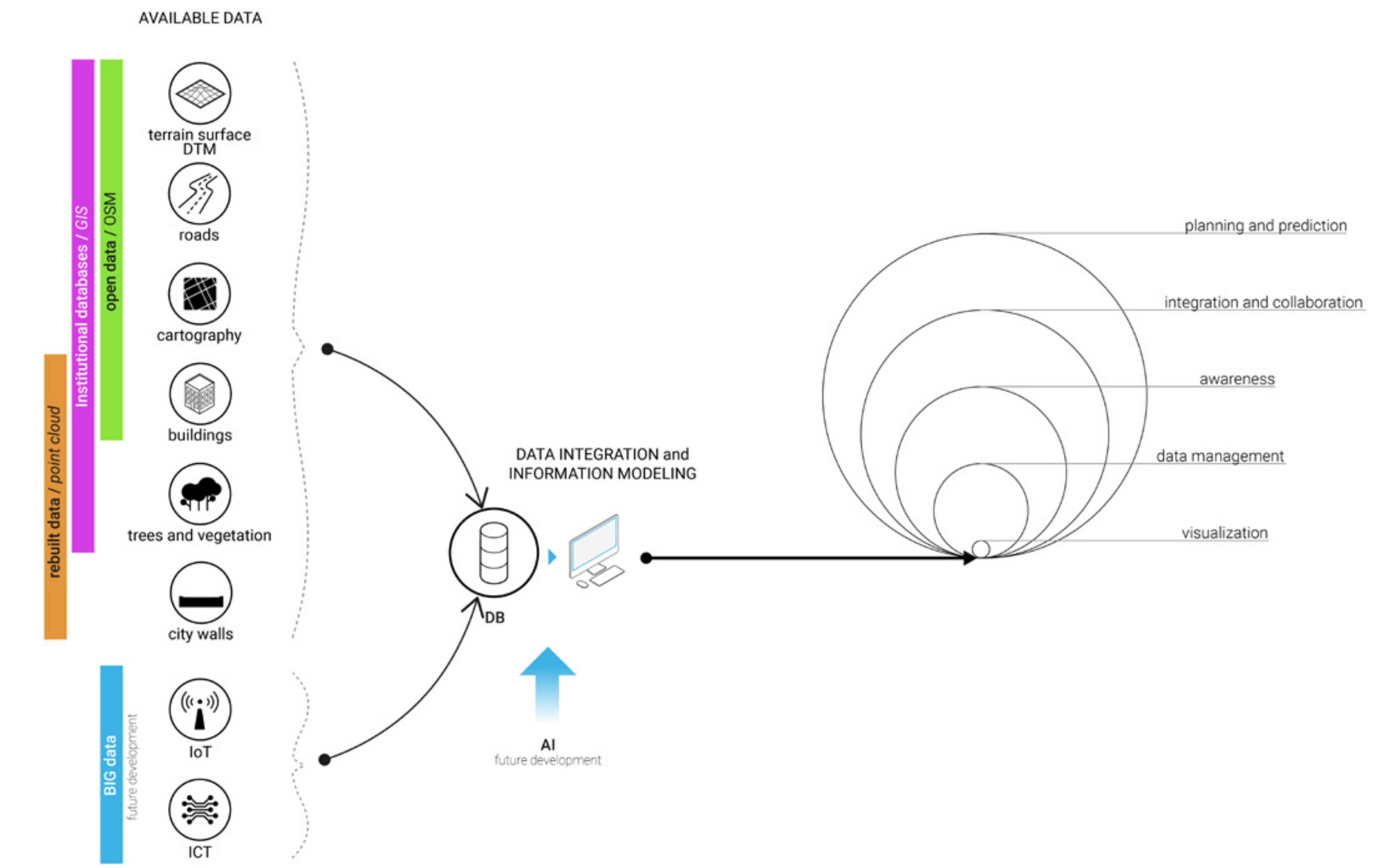
Given the growing importance of City Information Modeling (CIM) in urban planning and management, this contribution focuses on investigating an essential first phase in the generation of city information models. Using an urban section of the municipality of Ferrara as a case study, heterogeneous urban data were collected, organized, and modeled. The various data sources were then compared to support the identification of specific standards for defining descriptive urban models, facilitating an integrated interpretation of urban systems.

CASE STUDY AREA

The selected case study focuses on an urban section of the Municipality of Ferrara, located in the Emilia-Romagna region of Italy. This area, situated along the Po di Volano river to the south of the city, was chosen for its unique characteristics. It intersects various zones as defined in the Municipal Strategic Plan (PSC): historical fabric, consolidated fabric, and to be redeveloped fabric. The urban section is distinguished by several key features: the river embankment; residential areas along Via Argine Ducale and Via del Mulinetto; the historic center along Via Piangipane and Via Ripagrande, which houses the Italian National Museum of Judaism and the Shoah (MEIS); and a section of the historic city walls. Additionally, the area between the walls and the Burana Canal is currently undergoing urban redevelopment, aimed at enhancing its connection to the historic center.

01.

Identification of the case study area and urban section.



02.

This diagram illustrates the methodology applied in developing the urban digital model for Ferrara, including the available data sources and potential future enhancements.

03.

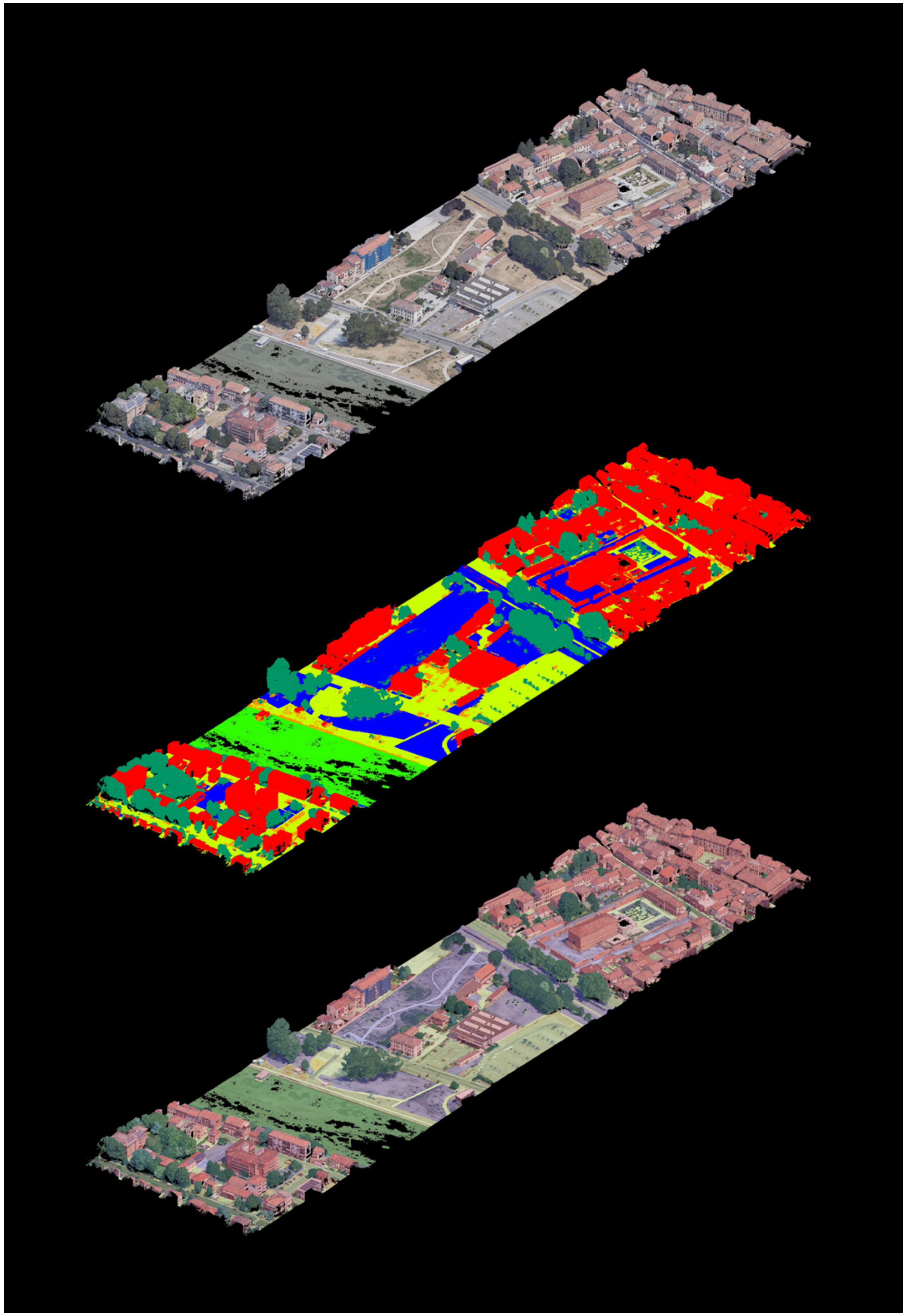
Point cloud derived from photogrammetric survey simulation and classification: buildings (red), roads (yellow), low vegetation (blue), high vegetation (dark green), rivers (light green), street furniture (orange), and vehicles (light blue).

METHODOLOGY

For the definition and description of the urban section of Ferrara identified in the city information model, two different modelling approaches were used, depending on the type of source data: top-down and bottom-up. The top-down approach involved the use of geodata to develop primitive geometries, which were progressively refined with incremental levels of detail. The bottom-up approach relied on survey data, applying reverse modeling techniques to progressively abstract the geometries. The former process assumes the preexistence of geodata that serves as the basis for modeling, with semantic differentiation used to enrich the resulting geometries. The latter, on the other hand, builds the information database downstream by segmenting the geometries obtained through initial instrumental surveys [9]. The choice of such approaches is tied to the objectives; however, the methodology described in this paper aimed to evaluate the use of heterogeneous data sources already available without creating new ones. Consequently, both approaches were employed.

The workflow was implemented in InfraWorks, an Autodesk software designed for urban and infrastructure planning, modeling, and analysis. InfraWorks was chosen for its advanced capabilities in integrating diverse datasets into a unified model. The information was organized into layers indexed according to a typological hierarchy, enabling efficient integration and management of existing data sources.

The initial phase of the modeling involved using Model Builder, a tool within InfraWorks that generates a three-dimensional model of a specified area by collecting datasets from open-



03.

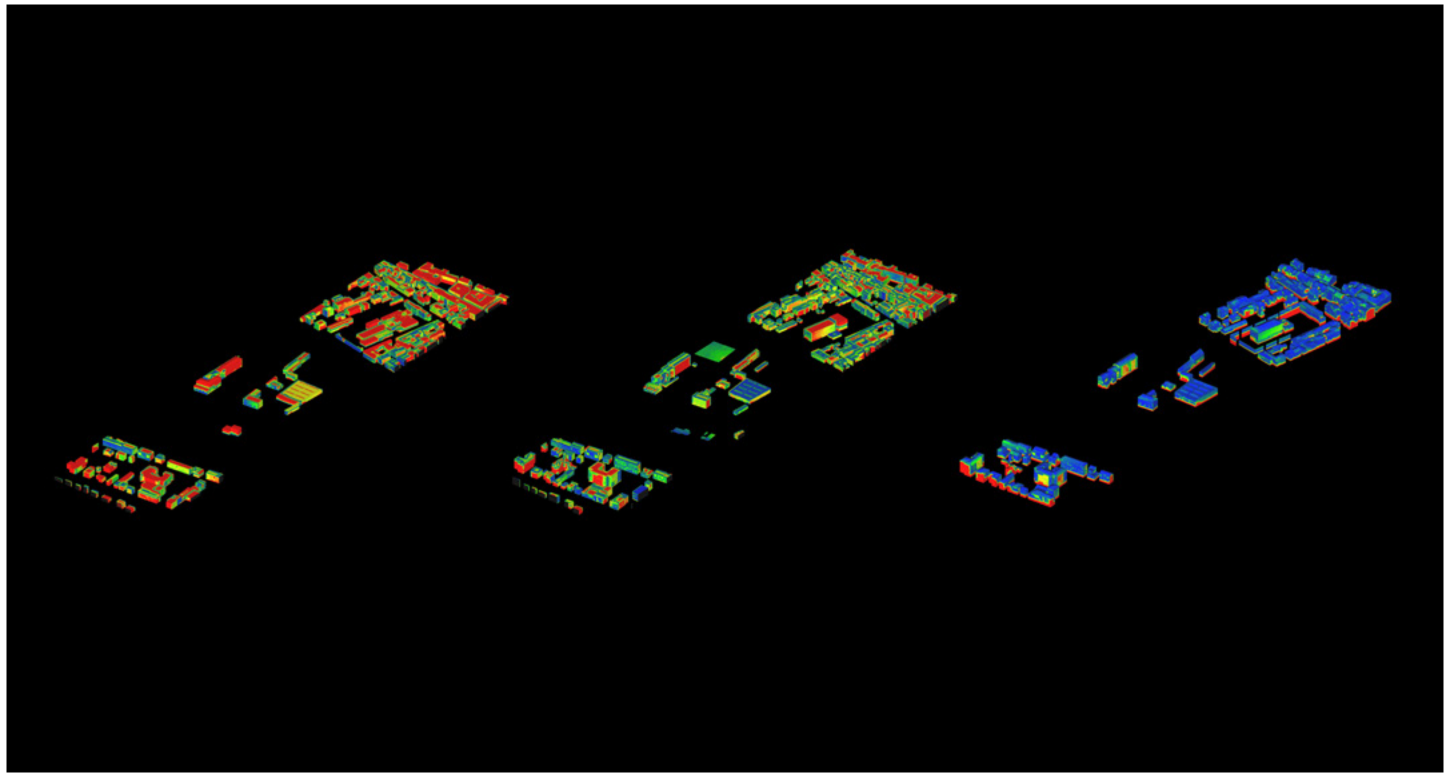
source databases, in this case for the Darsena of Ferrara. Roads, railways, waterways, and buildings data were obtained from OpenStreetMap (OSM), a collaborative open-source project that provides urban infrastructure data under the Open Database License (ODbL).

To increase the level of detail in the model, additional datasets were sourced from institutional repositories such as the geoportal of the Emilia-Romagna Region and the Municipality of Ferrara. These included the 5x5 Digital Terrain Model (DTM), derived from the Regional Technical Map, and shapefiles of buildings, roads, and tree surveys for the Darsena area. These datasets were configured to align with the study's goal of connecting existing databases to enable integrated visualization of urban systems.

A photogrammetric survey was simulated to supplement the available data. While urban-scale LIDAR or digital photogrammetry is commonly used to capture large-scale morphology [10], budget constraints prevented a full-scale survey. Instead, aerial images from Google Earth were used in a photomodeling process. Six known GPS points identified in the images enabled georeferencing and metric calibration of the model. As outlined in the literature, the resulting point cloud can be enhanced with semantic information by exploiting artificial intelligence algorithms for segmentation and classification. [11] [12]. These operations were identified as potential processes to be developed for bottom-up modeling approach optimization.
The methodology, following the steps described in literature [13], included defining classes, calculating features, manually mapping portions of the dataset, training and testing a Random Forest algorithm, and applying it to predict classifications for the remaining dataset. Misclassified points were corrected manually where necessary, particularly for small objects such as cars. The final point cloud was classified into categories including buildings, roads, low and high vegetation, rivers, street furniture, and vehicles (Fig. 03.). The classified point cloud was then integrated into InfraWorks as an additional descriptive data layer.

To simulate the bottom-up geometric modeling of buildings, the segmented point cloud was also imported into Revit. This approach facilitated the accurate representation of building morphologies, particularly complex roof structures characteristic of Italian historic centers. By isolating the building class within the point cloud, modeling efficiency was improved, and computational demands were reduced. The resulting Building Information Model (BIM) was subsequently integrated as a new layer within InfraWorks. The integration of geodata from the tree survey with the high vegetation class from the point cloud enabled the representation of tree models in InfraWorks. Similarly, combining the point cloud with the DTM allowed the precise modeling of the historic walls, which separates the city center from the Darsena.

For the correct integration of the input data, it was essential to define a common reference system, both for the sources and, consequently, in the modelling software. This aspect, which is often underestimated, must instead be taken into account right from the preliminary stages in order not to incur errors in the coordination of the models.

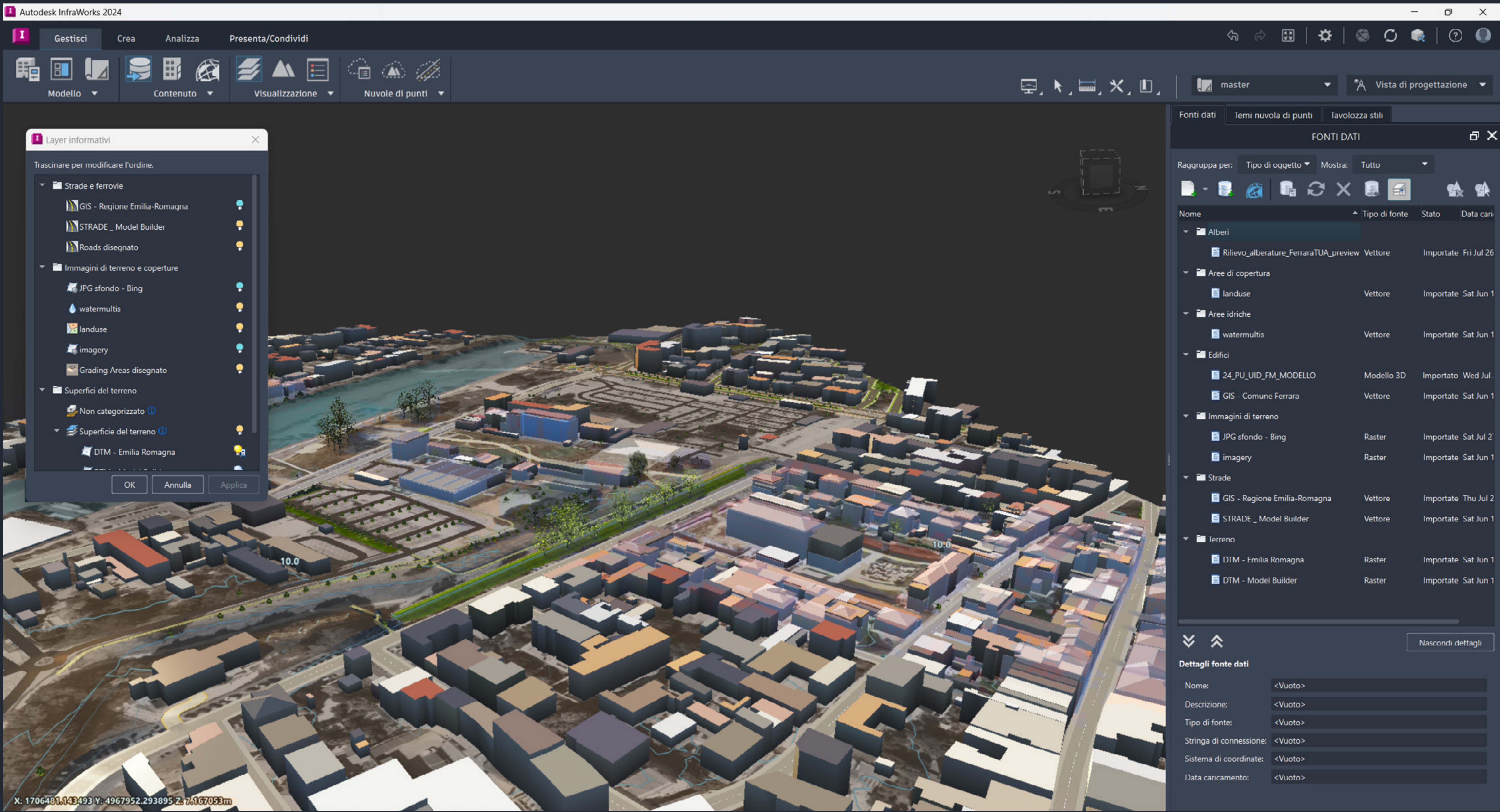


04.

Comparison between the segmented point cloud of the buildings and the volumes obtained from OSM data (left), GIS data (middle), and BIM model (right). In red are the major distances, in blue the minor ones. The scan-to-bim process model is the most consistent. The model from GIS data shows good overall consistency, but is deficient in describing roofs. The model from OSM data is the least consistent.

RESULTS

The result of the study is a three-dimensional model composed of a complex and heterogeneous database. All input datasets, categorized according to their type, are organized into separate layers. This organization enables an integrated reading of urban systems, allowing simultaneous visualization of their complexity or selective comparisons, such as examining buildings and greenery or streets and greenery. Moreover, depending on the construction of the datasets, information parameters can be attributed either to the entire dataset or to individual objects. These categorized datasets, represented at varying levels of geometric detail, provide a detailed visualization of the urban system within the Ferrara Darsena. The integration of data from various sources, with the objective of avoiding duplication and relying on existing datasets, revealed certain overlaps. For example, when analyzing buildings, three different sources — GIS data, OSM data, and the BIM model — produced slightly differing volumes. Consequently, it was necessary to evaluate which source was the most reliable for this category to determine the appropriate dataset for use. The evaluation method involved pairwise comparisons between the volumes obtained from each source and the point cloud derived from the photogrammetric survey simulation. Deviations were visualized using a color scale. In this case, the point cloud served as a benchmark, as it was considered the most reliable source of geometric and morphological data, despite being the result of an indirect process. The Google Earth data was regarded as the most up-to-date source. The comparison showed that, from both planimetric and volumetric perspectives, the deviations between the buildings modeled in BIM and those in the Municipality of Ferrara's GIS dataset were comparable. Excluding outliers, the degrees of accuracy were considered acceptable for urban-scale representation (Fig. 05.). Additionally, a specific advantage of buildings modeled through the BIM process was observed: this method supports the geometric modeling of roofs, a morphological component that is fundamental for describing the urban context typical of Italian cities. This capability adds significant value to the representation of the urban environment in Ferrara.



05.

InfraWorks screen showing integration of various data sources. The model combines data derived from both approaches: top-down and bottom-up.

FUTURE DEVELOPMENTS

The significance of CIM in urban planning and management lies in its ability to integrate heterogeneous data without duplication, relying instead on existing datasets to represent urban contexts effectively through information models. The approach presented in this paper contributes to identifying specific standards for defining a descriptive urban model, facilitating an integrated understanding of urban systems.

From a representational and descriptive perspective, the possibility of implementing additional, even customized, information parameters for systems and individual objects within the model, as seen in BIM, is particularly compelling. Moreover, a more detailed classification of the point cloud — potentially through a multilevel and multiscalar procedure [14] — could optimize the transfer of information from the cloud to the model.

The integration of BIM and GIS offers the combined advantages of modeling buildings alongside their surrounding environments. However, urban-scale applications remain underexplored and face challenges such as data interoperability, open standards, and visualization integration. Additionally, the GIS data available to public administrations are often outdated, requiring continuous investment to ensure relevance and accuracy. A critical area for further development involves advancing an open CIM approach, exemplified by exporting models in the CityGML format — an open standard for creating and exchanging 3D virtual city and landscape models. To transition the urban model into a simulation environment like a Digital Twin, a deeper focus on developing a data layer oriented toward Big Data collection is essential. Such a layer is increasingly recognized as a prerequisite for enabling analytics capabilities and integrating IoT functionalities [15]. There is growing interest in experimenting with and sharing best practices that can support the Smart City transition. This transition leverages digital technologies and collected data to guide urban transformations, highlighting the role of CIM in fostering innovative and data-driven urban development strategies.

BIBLIOGRAPHICAL REFERENCES

[1] Pasquinelli, A., Mastrolembo Ventura, S., Guzzetti, F., Ciribbini, A. (2016). City intelligence information modelling. Proceedings of the 3rd International Academic Conference on Places and Technologies, Belgrade, pp. 731 – 737.

[2] Parrinello, S., Picchio, F., De Marco, R. (2018). Urban modelling experiences for the representation of the historical city in Holy Land. In Disegnarecon 11/21, pp. 5.1-5.22.

[3] Del Giudice, M., Osello, A., Fonsati, A., De Luca, A., Musetti, a., Marchi, F., Patti, E., Brundu, F., Acquaviva, A. (2017). District data management, modelling and visualization via interoperability. Proceedings of the 15th IBPSA Conference, San Francisco, CA, USA, Aug. 7-9 2017, pp. 1239-1248.

[4] La Russa, F.M., Genovese, G., Santagati, C. (2021). City Information Modeling as a tool for territorial knowledge and prevention: an application in post-emergency scenarios. In Tommaso Empler, Adriana Caldarone, Alexandra Fusinetti (a cura di). 3DModeling&BIM. Data Modeling & Management for AECO Industry. Web-conference, 14th May 2020, pp. 305-321. Roma: Tipogrfia del Genio Civile.

[5] Li, M., Xue, F., Wu, Y., Yeh, A.G.O. (2022). A room with a view: automatic assessment of window views for high-rise high-density areas using City Information Models and deep transfer learning. Landsc. Urban Plan. 226.

[6] Xu, X., Ding, I., Luo, H., Ma, L. (2014). From building information modeling to city information modeling. J. Inform. Technol. Constr. 19, pp. 292–307.

[7] Thompson, E. M., Greenhalgh, P., Muldoon-Smith, K., Charlton, J., Dolník, M. (2016). Planners in the Future City: Using City Information Modelling to Support Planners as Market Actors. In Urban Planning, 1(1), pp. 79-93.

[8] Zhang, X. (2024). City Information Modeling and Its Applications: A Review. In Cheshmehzangi, A., Batty, A., Allam, Z., Jones, D. City Information Modelling, Springer Singapore, pp. 63-74. DOI: 10.1007/978-981-99-9014-6_4.
Okonta, E.D., Vukovic, V. (2023). A Scientometric Analysis and Visualisation of City Information Modeling (CIM): Trends, Research Areas, and Future Directions.

[9] La Russa, F.M., Santagati, C. (2020). From OpenData to city models: an Antifragile approach for City Information Modeling. In Building Information Modeling, Data & Semantics, 7/20, pp. 83-95.

[10] Mora, O.E., Suleiman, A., Chen, J., Pluta, D., Okubo, M.H., Josenhans, R. (2019). Comparing sUAS Photogrammetrically-Derived Point Clouds with GNSS Measurements and Terrestrial Laser Scanning for Topographic Mapping. Drones, 3(3):64.

[11] Grilli, E., Menna, F., Remondino, F. (2017) A Review of Point Clouds Segmentation and Classification Algorithms. Int. Arch. Photogramm.Remote Sens. Spat. Inf. Sci., 42, 339–344.

[12] Gao, L., Liu, Y., Chen, X., Liu,3D Bench Y., Yan, S., Zhang, M. (2024). CUS3D: A New Comprehensive Urban-Scale Semantic-Segmentation mark Dataset. Remote Sens., 16, 1079. https://doi.org/10.3390/rs16061079

[13] Weinmann, M., Jutzi, B. and Mallet, C., (2017). Geometric features and their relevance for 3D point cloud classification. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, p.157.

[14] Teruggi, S., Grilli, E., Russo, M., Fassi, F., Remondino, F. (2020). A Hierarchical Machine Learning Approach for Multi-Level and Multi-Resolution 3D Point Cloud Classification. Remote Sens., 12, 2598; https://doi:10.3390/rs12162598

[15] Campos Fialho, B., Codinhoto, R., Márcio Minto, F. (2020). BIM and IoT for the AEC Industry: A systematic literature mapping. XXIV International Conference of the Iberoamerican Society of Digital Graphics