FROM 3D SURVEYING DATA TO BIM TO BEM: THE INCUBE DATASET

O. Roman[1,2], E.M. Farella[1], S. Rigon[1], F. Remondino[1], S. Ricciuti[3], D. Viesi[3]

[1] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Email: <oroman><elifarella><srigon><remondino>@fbk.eu

[2] Department of Industrial Engineering (DII), University of Trento, Trento, Italy - oscar.roman@unitn.it

[3] Center for Sustainable Energy, Bruno Kessler Foundation (FBK), Trento, Italy - Email: <sricciuti><dviesi>@fbk.eu

KEY WORDS: surveying, 3D, point cloud, laser scanning, photogrammetry, BIM, BEM, energy efficiency

ABSTRACT:

In recent years, the improvement of sensors and methodologies for 3D reality-based surveying has exponentially enhanced the possibility of creating digital replicas of the real world. LiDAR technologies and photogrammetry are currently standard approaches for collecting 3D geometric information of indoor and outdoor environments at different scales. This information can potentially be part of a broader processing workflow that, starting from 3D surveyed data and through Building Information Models (BIM) generation, leads to more complex analyses of buildings' features and behavior (Figure 1). However, creating BIM models, especially of historic and heritage assets (HBIM), is still resource-intensive and time-consuming due to the manual efforts required for data creation and enrichment. Improve 3D data processing, interoperability, and the automation of the BIM generation process are some of the trending research topics, and benchmark datasets are extremely helpful in evaluating newly developed algorithms and methodologies for these scopes. This paper introduces the InCUBE dataset, resulting from the activities of the recently funded EU InCUBE project, focused on unlocking the EU building renovation through integrated strategies and processes for efficient built-environment management (including the use of innovative renewable energy technologies and digitalization). The set of data collects raw and processed data produced for the Italian demo site in the Santa Chiara district of Trento (Italy). The diversity of the shared data enables multiple possible uses, investigations and developments, and some of them are presented in this contribution.

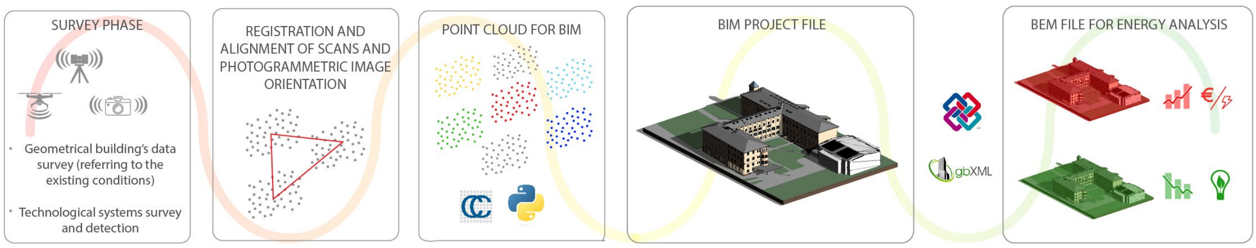


Figure 1. The general workflow for the BIM and the BEM generation, according to the InCUBE project.

1. INTRODUCTION

In recent years, the Architecture, Engineering, and Construction (AEC) sector has experienced a significant surge in the need for Building Information Modelling (BIM) models. The creation of BIM models for new constructions is quite a straightforward process, thanks to the available modern plans and design products. On the other hand, dealing with existing (and also historical) buildings is much more complex, with many bottlenecks in the automatic BIM generation from 3D surveyed data and still many manual processes. Indeed, with existing buildings, to link both geometric and semantic information with the huge variety of shapes, construction techniques and materials (Valero et al., 2022), structural and constructive elements need to be manually modelled based on 3D reality-based surveys. In addition, no standard protocols are defined to guide the modelling phase, which is highly time-consuming and a mostly manual process (Yang et al., 2019, Karasaka et al., 2023).

BIM models are now in high demand for various applications, including building renovations, maintenance, and the planning of structural or energy-efficient retrofits (Durdyev et al., 2021; Cao et al., 2022; Pinti et al., 2022). For energy efficiency of buildings, the employment of BIM models perfectly matches with the European Community directives, such as the Energy Efficiency Directive (EED) [1] or the Energy Performance of Building Directive (EPBD)[2], issued with the goal to build nZEB (near Zero Energy Buildings) buildings. Creating a model in which structural and energetic features can be continuously updated could be a significant step towards the goal of defining nZEB buildings. The BIM-based building energy simulations involve creating Building Energy Models (BEMs), playing a crucial role in reaching the goal of net-zero carbon building emissions by 2050 (Liu et al., 2022). BEM models are promising tools to predict and improve energy efficiency or to plan and manage preventive or predictive maintenance for both buildings and devices. Currently, the lack of interoperability in the BIM-to-BEM transition makes the design process inefficient, error-prone, and poorly automated (Yang. et al., 2022; Tang et al., 2022).

1.1 Paper contributions

Considering all the open issues in the automated generation of BIM models from reality-based 3D data and their importance for many applications, the aim of the paper is two-fold:

- to introduce the InCUBE dataset made available to the R&D community to develop and validate new solutions for increasing automation in 3D data processing, BIM (and

[1] https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive_en

[2] https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energyperformance-buildings-directive_en

Historical Building Information Modelling - HBIM) generation, interoperability between different software and simulation of energy efficiency processes (BEM);

- to test, improve and validate innovative methodologies for increasing automation in the BIM reconstruction process, leveraging the shared InCUBE dataset.

2. RELATED WORKS

In recent years, the AEC sector has increased its interest in the BIM modelling of the as-built structures, aiming to improve the design and management of energy or structural efficiency before on-site operations. It concerns not only existing public and civil buildings but also historical and heritage assets.
Nevertheless, the integration of BIM models in the heritage conservation framework is still a significant challenge due to the complexity of managing a relevant amount of geometric data (Liao et al., 2022) and related semantic information.

2.1 Scan-to-BIM

Scan-to-BIM methods can be differently classified, based on the type of modelling (parametric or direct modelling) or the level of automation.
Parametric modelling enables the use of predefined models, which can be adapted following the project requirements and objectives through modifications of pre-set parameters. The project could also be managed through Visual Programming Language (VPL) scripts to improve the modelling phase. The parametric approach can be classified into:

- *Parametric-generative modelling*: this approach seeks innovative strategies to simplify and enhance design solutions. Models can be edited at any stage by adjusting parameter values. This type of modelling is typically managed through VPL interface platforms, which enables interoperability with other BIM software.
- *Parametric object-oriented modelling* is the most common approach in a BIM workflow, where modelling relies on pre-existing libraries and foundational construction elements.

Nevertheless, both modelling strategies are hardly applicable to the HBIM sector, where many challenges are usually faced in the reconstruction phase due to the need to model the complex shapes of a multitude of architectural elements and details (Zhang et al., 2022). In this context, applying traditional automatic or semi-automatic approaches and standardized methods for the "repetitive" modelling of structural elements is not a feasible solution. The direct modelling of the elements is the typical approach in the HBIM field, where the model restitution starts from 3D profiles extracted from point clouds and imported into BIM authoring tools. This manual approach is time-intensive, and the user's expertise significantly influences the final quality of the model (Tran et al., 2019; Liao et al., 2022). As a result, a singular and non-replicable model is created for each case study.
A further categorization is based on the level of automation of the modelling process. In particular:

- *Semi-automatic approaches:* these methodologies make use of scripts (Andriasyan et al., 2020) to address specific process steps, like data pre-processing, mesh creation, or geometries extrapolation. They are integrated as software plug-ins.
- *Automatic approaches:* these methods are rooted in an in-depth manipulation of point clouds, relying on Machine Learning (ML) or Deep Learning (DL) algorithms. These techniques aim to automatically detect opening positions or create watertight room volumes for the definition of primarily walls and floors. These techniques are only partially applicable to the HBIM field.

The automation of the BIM reconstruction framework from reality-based data (Scan-to-BIM) is one of the recent trending topics. Semi-automated approaches, based both on point clouds (Bassier and Vergauwen, 2020; Yang et al., 2023) and raster images (Ning et al., 2021), are currently the most robust in the modelling pipeline. Generally, using point clouds guarantees a more accurate object reconstruction, even if raster image-based methods demand less computational power. Further methods are based on the extraction of lines (Ning et al., 2021; Shi et al., 2018) or planes or meshes. The selection of the suitable approach depends on the reconstruction goal and the dataset dimension.
All methods require input by users, such as indicating wall thickness or the connection types for pairing lines. This prior knowledge, derived from surveyed or different data, needs to be manually added to guarantee a correct BIM reconstruction.
Currently, the ISPRS dataset (Khoshelham et al., 2020) is one of the most popular datasets used for benchmarking and testing Scan-to-BIM methods.

2.2 From BIM to BEM

The transition from BIM (or HBIM) to Building Energy Modelling (BEM) represents a fundamental step within the Architecture, Engineering and Construction (AEC) sector. As a comprehensive digital representation of a structure's geometric and semantic attributes, the BIM can support collaborative design, clash detection, and construction scheduling. By recording the building's physical and operational features, the BIM is the base for the BEM, where complex simulations and analyses come into operation, examining building energy usage and performance, thermal dynamics, and daylight utilization.

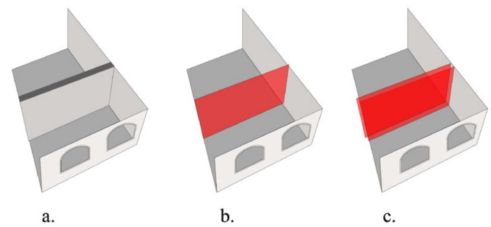


Figure 2. Standard assumptions for the definition of the energy borders. **a.** The indoor space. **b.** The single energy border without gaps. **c.** The double energy borders with a central gap (modified from Yang et al., 2022).

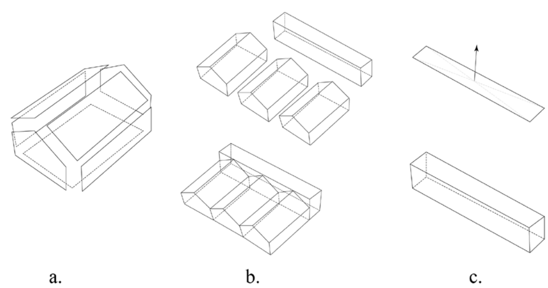


Figure 3. Methodologies to define energy volumes: **a.** Aggregation of surfaces, **b.** Union of volumes, based on primitives, **c.** Extrusion of surfaces to create volumes (modified from Yang et al., 2022).

This product aids engineers in applying solutions for energy-saving, refining HVAC systems, and boosting overall sustainability. Recent works point out different uses of BEM, such as energy prediction and estimation, energy consumption, optimal design, energy evaluation, efficiency, management, or energy optimization (Pezeshki et al., 2019). Furthermore,

https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-175-2023 |

according to the specific application, the model is developed using different standards (Figure 2) and conventions (Figure 3).
Ultimately, the integration between BIM and BEM allows the fusion of design and energy-efficient engineering, leading to structures that balance construction needs and efficiency (Cao et al., 2022). Recent works introduce processes to model both structural elements, such as slabs, roofs and walls (Andriasyan et al., 2020, Yang et al., 2023), and openings, such as windows and doors (De Geyter et al., 2022), needful for conducting energetic simulations in existing buildings. However, obtaining a standard workflow that can be applied on a large scale is challenging due to the wide variety of building typologies in the as-built environment. Since software, data structures, and goals differ depending on the various situations, interoperability could cause significant data loss.

2.3 Benchmarks and datasets in the BIM domain

Benchmarks serve as standardized evaluation criteria for objective and comparative analyses of different processing techniques. In the BIM context, by using benchmark data, researchers and practitioners can accurately measure the performance, accuracy, and efficiency of different reconstruction methods, enabling meaningful comparisons and identifying the strengths and weaknesses of each technique. Additionally, benchmarks contribute to advancing the state-of-the-art by fostering competition and innovation among researchers while developing more robust and effective BIM reconstruction methodologies.
Even if the literature reports many examples of BIM reconstruction methods (Rashdi et al., 2022), and many algorithms are shared in repositories, authors point out that, unfortunately, there is a lack of shared benchmarking datasets (Abreu et al., 2023). One of the causes is that authors tend to utilize their own datasets for testing and validating their methods, and this practice does not allow comparisons between different state-of-the-art methodologies.
The ISPRS dataset (Khoshelham et al., 2020) is the most recent benchmark released for testing Scan-to-BIM methods and the reconstruction of indoor environments in the BIM domain. It comprises a six-point cloud dataset representing both Manhattan and non-Manhattan environments. These point clouds refer to offices, residential buildings, and museums, and they are captured using five types of sensors: terrestrial, backpack, and handheld laser scanners.
The presented InCUBE dataset, unlike the ISPRS one, explores the HBIM sector and encourages the development of procedures useful for the parametric representation of a wide and complex variety of architectural elements.

3. THE InCUBE PROJECT

Despite the high necessity for renovations in the built environment, Europe's progress is below expectations. So far, the weighted annual energy renovation rate is about 1%. The renovation is held back by barriers at different points throughout the value chain – from the initial decision to renovate financing and project completion. Recent technological developments and the EU's ambition to place energy efficiency and renewability at the forefront of sustainable development can set off the renovation wave, offering an excellent opportunity to make renovation a win-win for climate neutrality and equitable economic recovery. Pushed by these needs, the EU InCUBE project[3] brings together 23 partners and two affiliated entities from 7 European countries, envisioning to unlock the EU renovation wave through cutting-edge standardized and integrated processes based on 4 key pillars of innovation: advanced digitalization, new technologies and materials for energy efficiency, industrialization of renovation and new market players. The InCUBE solutions will be demonstrated in 3 large-scale demo sites: Zaragoza (ES), Trento (IT) and Groningen (NL). The Italian demo site is the Santa Chiara district (Figure 4), which includes several building blocks close to the city centre. The district comprises a historical and heritage building (called B6 building and of interest in the InCUBE projects for the energy renovation intervention). The structure dates back to 1229, when the Clarisse nuns built the initial monastery. Throughout time, the complex underwent multiple transformations, featuring a total amount of 22,000 m^3, mostly abandoned. An annexed and more recent block, the “Palabocchi” building, will be part of the InCUBE renovation project, and it will host the installation of photovoltaic panels to improve the energy efficiency of the historic building.

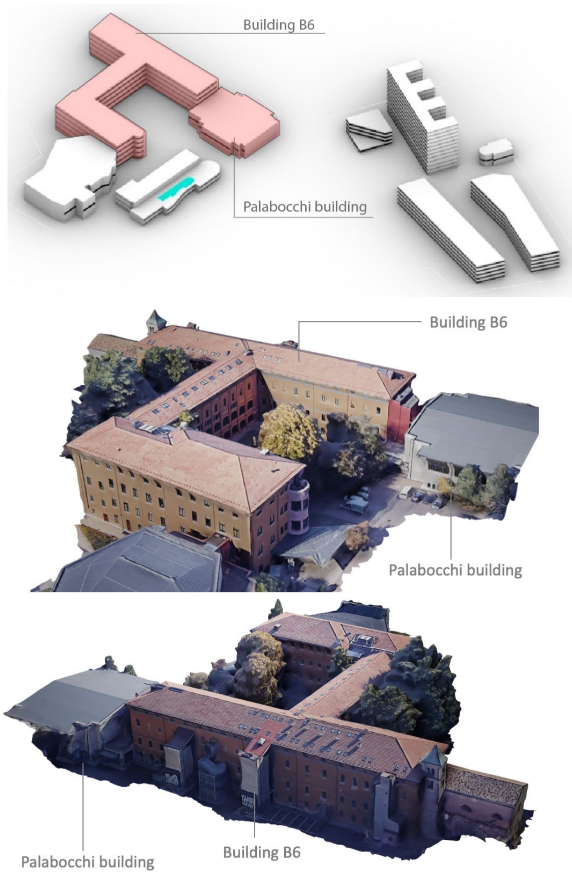


Figure 4. The Santa Chiara district (top) and a view from Google Earth (2023) of the historical buildings (B6) and the nearby Palabocchi structure (center and bottom).

4. THE INCUBE DATASET

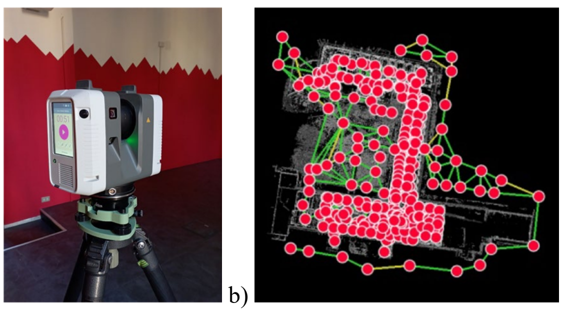
Among the activities of the InCUBE project, a specific task is dedicated to the realization of BIM models of the demo sites to support renovation activities. For the Trento case, a BIM representation of the Santa Chiara historical building (the actual

[3] https://incubeproject.eu/

and the planned renovation state) was produced starting from 3D reality-based point clouds generated from multi-sensor and multi-resolution data. Most of the 3D acquisition campaign was conducted with active technologies. As presented in the following section, a terrestrial laser scanner (TLS) enabled the detailed acquisition of the facades and the indoor spaces, primarily interested in the renovation interventions. A mobile mapping (MM) solution was preferred for digitally acquiring two further areas not accessible in the first campaign. Missing data on the roof were collected, processed, and co-registered, starting from the photogrammetric processing of aerial images (AP). The data collected will be presented in the next paragraph.

4.1 Data collection

A Leica RTC360 TLS, featuring a range from 0.5 m up to 130 m and a speed of up to 2 million points per second (Figure 5), was used for the 3D surveying of the Santa Chiara B6 block. 359 scans were acquired for both indoor and outdoor spaces. Starting from the pre-aligned on-site data, the final registration, performed in Cyclone Register 360, led to some 802 million points, with an original average resolution of 2 mm. Panoramic images acquired from every scan station are also part of the shared datasets.



a) b)

Figure 5. The TLS Leica RTC360 used for the survey (a), and some scan positions after the data registration.

The merged point cloud was then subsampled to improve data management and manipulation in the BIM reconstruction phase, obtaining a point cloud with an average resolution of 5 mm composed of 388 million points (Table 1).

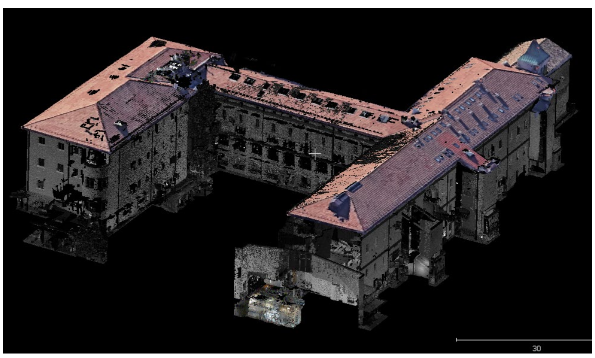


Figure 6. The complete point cloud of the B6 building.

Within a second campaign, two further areas were acquired with the Leica BLK2GO mobile mapping system. It features a nominal range from 0.5 m up to 25 m, including panoramic images. Two scans were acquired for the under-roof of the southern wing (shared in the dataset already aligned and merged), while a further scan was performed on the underground floor of the Palabocchi building (Figure 3). Data from the BLK2GO were processed and co-registered to the RTC360 subsampled point cloud.

The surveying and digitization of the B6 building was completed using available aerial photogrammetric data, which were necessary for the roof part of the structure (Figure 6). The shared benchmark includes all the single scans, the panoramic images of the TLS, and the co-registered subsampled point clouds.

Point cloud	Number of scans	Millions of points	Average resolution
TLS (RTC360): original scans	359	802	2 mm
TLS (RTC360): down sampled and merged scans	1	388	5 mm
BLK2GO	2	186	1 mm

Table 1. Characteristics of the acquired and processed scans for the B6 building.

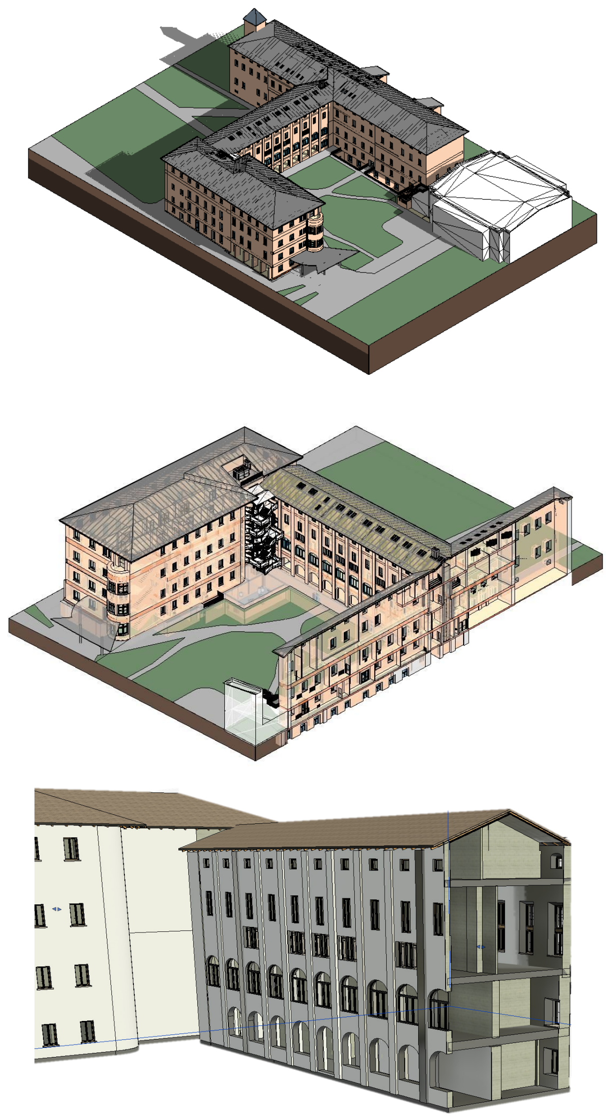


Figure 7. Views of the BIM model of the Santa Chiara B6 and the CAD-based model of Palabocchi buildings.

4.2 BIM generation

The InCUBE dataset also includes the BIM model of the B6 building (Figure 7), which could serve as ground truth for evaluating modelling algorithms. The model was manually realized starting from the acquired 3D data. Since the Santa Chiara B6 block is a historical building, it features irregular architectural and structural elements and a complex development of floors and indoor spaces. Generally, the rigidity of BIM authoring tools (as the Autodesk Revit used in this case) does not allow to model walls featured by a variation of thickness in *xy* plane, nor out of plumbs in vertical directions. For these reasons, using semi-automated reconstruction techniques in HBIM is a complex processing task.

4.3 Available data and R&D tasks

The diverse data shared in the InCUBE dataset (Table 2) open to several possible investigation tasks and applications.
As an example, the point clouds could offer a consistent amount of data for testing innovative cleaning, registration, segmentation and classification algorithms based on both ML and DL techniques. Similar related experiments could be applied to panoramic images (also useful for SLAM applications).
On the other hand, the BIM model could be employed as ground truth and reference for testing innovative semi-automatic or automatic reconstruction procedures in the BIM domain, based both on frames (raster solution) and on point cloud (points and vectors solutions). The shared data could also be exploited for testing meshing algorithms.
The InCUBE benchmark datasets are available at the following GitHub repository: https://github.com/3DOM-FBK/InCUBE.

#	Element	Docs and Description	Format
1	Single TLS scans	359 separate point clouds	.ptx
2	Panoramic Images	359 panoramic images from TLS camera	.jpeg
3	RTC360 Point Cloud	The aligned and merged point cloud	.las
4	BLK2GO	Two aligned and merged and 1 separate point cloud	.las
5	Roof point cloud	The roof elaborated with photogrammetry	.las
6	BIM model	The BIM model of the Santa Chiara building	.rvt /.IFC

Table 2. The content of the InCUBE dataset.

5. TASKS AND EXPERIMENTS

This section offers an overview of some processing tasks that could be addressed by leveraging the presented InCUBE data. Experiments focus on testing methods for data classification, automatic BIM elements' generation, and energetic analyses, introducing further possible applications related to SLAM processing.

5.1 Point cloud classification

Point cloud classification is an active research field, and many algorithms for automatic data segmentation and classification are nowadays available.
A portion of the InCUBE point cloud (ca 16 million points) is classified using a ML approach based on the Random Forest (RF) algorithm (Grilli and Remondino, 2020). The classes selected for the supervised classification of the building elements of the B6 block are presented in Table 3, while Table 4 summarizes the assessment of the results. After the manual annotation of class elements, geometric features are extracted: *Linearity* [L], *Omnivariance* [O], *Anisotropy* [A], *Planarity* [P], *surface variation* [SV], *Verticality* [V] and *Sphericity* [S]). Therefore, the model is trained, and the inference on the entire point cloud is shown in Figure 8.

#	Classes
1	Walls
2	Windows
3	Doors
4	Floors
5	Furniture
6	Columns

Table 3. Identified classes used for testing the Random Forest (RF) classification algorithm.

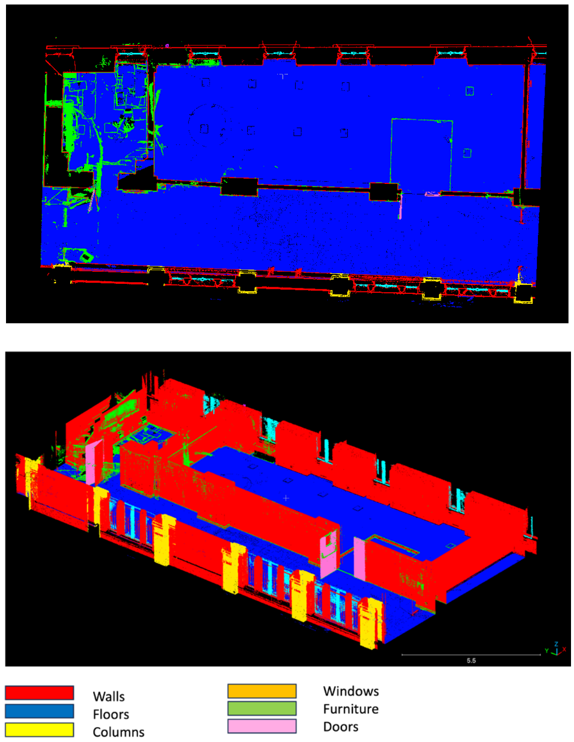


Figure 8. Visual results of the 3D classification on a portion of the second floor plan of the Santa Chiara B6 building (top: plan view; bottom: perspective view).

	Recall	Overall Accuracy	Precision	F1
Central wing 2nd floor	0.93	0.92	0.89	0.90

Table 4. Metrics from the classification with the RF algorithm of the 2nd floor of the central wing of the B6 building.

5.2 Towards automatic BIM generation - methodology

In the case of regular structures and basic shapes, semi-automatic BIM reconstruction approaches have demonstrated their effectiveness (Croce et al., 2023). Nevertheless, many limitations are still evident when handling complex as-built shapes due to the constraints and rigidities of BIM tools. The Santa Chiara case study presents irregular structural components commonly observed in historical built heritage, requiring manual modelling to correctly define geometries.

This section presents a methodology and some experiments to increase automation in reconstructing main structural elements, starting from 3D reality-based data. In the implemented workflow, some pre-processing operations, such as a down sampling of the point clouds or a noise and outliers' removal, are typically performed to improve the processing pipeline outputs.

The workflow can be summarized as:

1. *Point cloud segmentation:* points with similar properties are grouped into multiple homogeneous regions;
2. *Point cloud classification:* a class is associated with each segmented region (Table 3);
3. *Supervised-based clustering:* employ supervised-based algorithms to cluster the points, with options for both bottom-up and bottom-down approaches (Sarker, 2021);
4. *Geometry estimation and extraction* algorithm, through the *RANSAC algorithm,* in order to define planes from clusters;
5. *Lines extraction* from planes and clusters;
6. *Filtering and pairing overlapping segment lines* (Figure 9): implement a filtering method to identify and pair overlapping line segments. This process considers various thresholds, including angles (α) and normal values ($\vec{v}$);
7. *Extraction of matrices* which contain three-dimensional coordinates of lines (in *xy* plane), complete with *z* coordinate, which is defined as the average value of z coordinate points;
8. *Customizable Application Programming Interface* (API) to enable semi-automated retrieval of Building Information Model (BIM) objects using the extracted matrices.

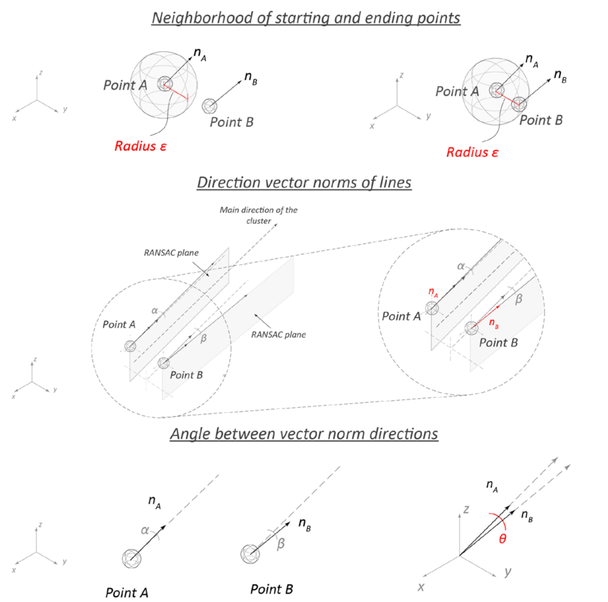


Figure 9. The filters used to process segment lines detected.

The presented reconstruction method takes into consideration, in addition to the three-dimensional points coordinates (*x, y* and *z*), the normal values (n_x, n_y, n_z), which are helpful in defining the main fitting planes, and finally enable the extraction of segment lines (Figure 10).

Lines are subsequently processed and merged through a filtering operation. The implemented filters are set to evaluate:

- the number of points within a defined neighbourhood for each *starting* and *ending_points*,
- the line direction vectors $\vec{v}$ of lines, the relative angle of rotation *(α)* between the detected lines, in order to pair lines featured by similar parameters.

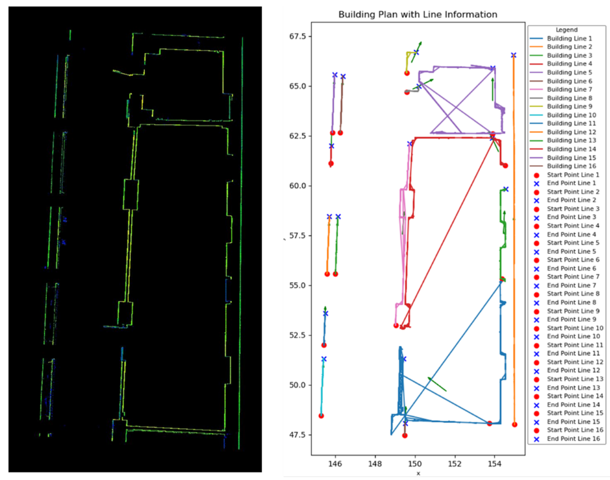
The method also takes as ancillary input prior knowledge information, such as wall thicknesses or connection typologies.

Preset parameters for handling the pairing lines phase are:

- $\Delta\theta$, which is the angle variation (based on the cosine) between the vector direction of two different lines;
- $(\Delta_w, \Delta_f, \Delta_r)$, which are the distances between vector lines.

In the filtering phase, through the angle threshold (th_α), and the thickness thresholds (th_w, th_f, th_r), related to walls, floors and roofs, it's possible to merge or delete lines which lays outside these limits.

Finally, the reconstruction phase is performed using an Add-in developed using C# and Revit API, which requires, as input, the ASCII file (*.txt* or *.csv* format file), where filtered lines are stored in an array, encoded in the following mathematical shape $\vec{v_i} = \{x_{0i}, y_{0i}, z_{0i}, x_{0f}, y_{0f}, z_{0f}\}$.



(a) (b)

Figure 10. (a) Irregular walls of the floor plan of the building. (b) Detected segments featured by starting and ending points. Lines must be filtered in the subsequent phase.

The C# code offers the possibility to automatically generate walls from the array (Figure 11). However, for HBIM environments, many users' inputs are required to manage inaccuracies between elements, such as wall thicknesses or connection typologies.

Finally, doors and windows families have been modelled and provided as a Revit library. For these architectural elements, the semi-automatic approach developed by Roman et al. (2023), is applied. The script, developed within Dynamo for Revit, returns BIM elements from a classified point cloud. Results on two different opening models are presented in Figure 12.

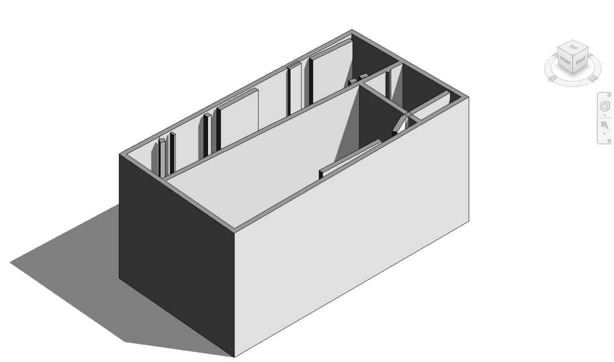


Figure 11. The automatic reconstruction of walls through the developed Revit API.

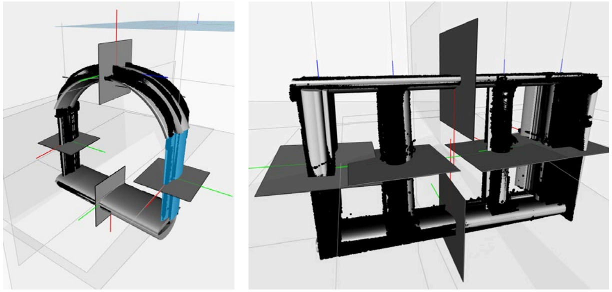


Figure 12. Examples of semi-automatic reconstruction of architectural elements within Dynamo.

5.3 Building Energy Model (BEM)

BIM models are more and more of interest for energy analyses and simulations. The interoperability between BIM and BEM models currently has many limitations, which often result in the loss of critical information and impact the accurate generation of the energy model. Improving interoperability and developing innovative methodologies for the accurate and effective use of BIM data for energy simulations will be the focus of the following InCUBE project activities.

Since in the Santa Chiara district will be established an energy community in the next few years, all the involved buildings were considered in the first energetic analyses conducted for a rough estimation of the energy demand of the district (Figure 13). For these preliminary investigations, a simulation of cooling and heating demand was performed on CAD-based models using the available climate database for the year 2024 and the UMI (Urban Modelling Interface) software[4].

The UMI model is an advanced design environment developed by the Sustainable Design Lab at the Massachusetts Institute of Technology (MIT) and implemented on the Rhinoceros platform. It aims to enhance the precision of urban energy modelling, while reducing the times required for simulation and analysis.

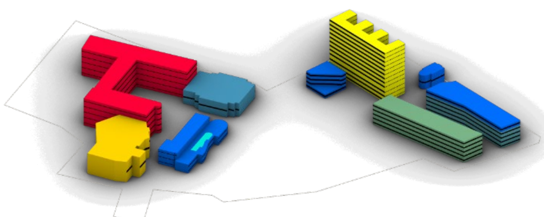


Figure 13. View of the Santa Chiara entire district and all the buildings involved in the analyses within UMI for the establishment of the future energy community (different colors for each CAD building block - B6 structure in red).

Figure 14 shows some insights into the annual energy requirements of the B6 block. This historically protected structure is subject to specific limitations on efficiency upgrades, and it has the most significant impact on the overall energy demands of the district. More robust energetic analyses and simulations on the district will be conducted in the next phase of the project, relying on the vast information contained in the produced BIM model.

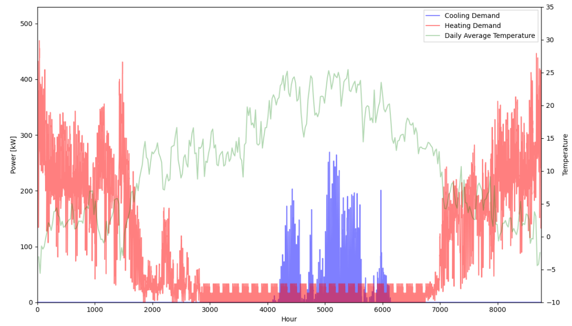


Figure 14. Profiles of simulated energy demand for the Santa Chiara district.

6. CONCLUSIONS

The paper presented a new dataset realized within the activities of the EU InCUBE project. The dataset comprises distinct TLS scans, panoramic images acquired in every scan station, fully registered and merged point cloud data, and the BIM model of the Santa Chiara district in Trento, Italy (Table 2). The considered building is a historic structure, which poses many difficulties in automated procedures for, e.g., BIM generation. The InCUBE dataset could serve as a valuable resource for research and development in several processing tasks: point cloud cleaning, registration or classification, automated BIM generation, etc.

The paper also presented some possible investigation tasks leveraging the shared data. The point cloud classification was successfully performed by leveraging available ML approaches, with the correct identification and classification of structural elements (such as columns, floors and walls). The reconstruction phase, indeed, especially within the BIM environment, is still challenging. An experimental method is introduced for the semi-automatic reconstruction of structural elements, although some bottlenecks and limitations still characterize the phase of extraction and filtering of the segments from the point cloud. A significant computational capacity is also required for data processing, analysis, and the final reconstruction of elements within the digital BIM environment. User interaction is still needed (Yang et al., 2023) in many operations, such as providing inputs to correctly pair lines belonging to the same structural element or adding ancillary information for the not visible parts in the surveyed data, such as the typologies of connection between elements.

Finally, the InCUBE dataset has the potential to serve as a foundation for developing interoperability solutions between Building Information Modelling (BIM) and Building Energy Modelling (BEM), simplifying the stages of monitoring and simulating energy analysis.

ACKNOWLEDGMENTS

The work is supported by the EU InCUBE project (https://incubeproject.eu/) funded by the European Union under the grant agreement no. 101069610.

REFERENCES

Abreu, N., Pinto, A., Matos, A. and Pires, M., 2023. Procedural Point Cloud Modelling in Scan-to-BIM and Scan-vs-BIM

[4] http://web.mit.edu/sustainabledesignlab/projects/umi/index.html

Applications: A Review. *ISPRS International Journal of Geo-Information*, *12*(7), p.260.

Andriasyan, M., Moyano, J., Nieto-Julián, J.E. and Antón, D., 2020. From point cloud data to building information modelling: An automatic parametric workflow for heritage. *Remote Sensing*, *12*(7), p.1094.

Bassier, M. and Vergauwen, M., 2020. Topology reconstruction of BIM wall objects from point cloud data. *Remote Sensing*, *12*(11), p.1800.

Cao, Y., Kamaruzzaman, S.N. and Aziz, N.M., 2022. Building information modeling (BIM) capabilities in the operation and maintenance phase of green buildings: A systematic review. *Buildings*, *12*(6), p.830.

Croce, V., Caroti, G., Piemonte, A., De Luca, L. and Véron, P., 2023. H-BIM and Artificial Intelligence: Classification of Architectural Heritage for Semi-Automatic Scan-to-BIM Reconstruction. *Sensors*, *23*(5), p.2497.

De Geyter, S., Bassier, M., De Winter, H. and Vergauwen, M., 2022. Review of Window and Door Type Detection Approaches. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, pp.65-72.

Durdyev, S., Dehdasht, G., Mohandes, S.R. and Edwards, D.J., 2021. Review of the building information modelling (BIM) implementation in the context of building energy assessment. *Energies*, *14*(24), p.8487.

Google Earth Pro, 2023. Version 7.3.6.9345.

Grilli, E. and Remondino, F., 2020. Machine learning generalisation across different 3D architectural heritage. *ISPRS International Journal of Geo-Information*, *9*(6), p.379.

Karasaka, L. and Ulutas, N., 2023. Point Cloud-Based Historical Building Information Modeling (H-BIM) in Urban Heritage Documentation Studies. *Sustainability*, *15*(13), p.10726.

Keitaanniemi, A., Virtanen, J.P., Rönnholm, P., Kukko, A., Rantanen, T. and Vaaja, M.T., 2021. The combined use of SLAM laser scanning and TLS for the 3D indoor mapping. *Buildings*, *11*(9), p.386.

Khoshelham, K., Tran, H., Acharya, D., Díaz Vilariño, L., Kang, Z. and Dalyot, S., 2020. The ISPRS benchmark on indoor modelling–Preliminary results. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, pp.207-211.

Liao, L., Zhou, K., Fan, C. and Ma, Y., 2022. Evaluation of Complexity Issues in Building Information Modeling Diffusion Research. *Sustainability*, *14*(5), p.3005.

Liu, Z., Li, P., Wang, F., Osmani, M. and Demian, P., 2022. Building Information Modeling (BIM) driven carbon emission reduction research: A 14-year bibliometric analysis. *International Journal of Environmental Research and Public Health*, *19*(19), p.12820.

Ning, X., Wang, M., Tang, J., Zhang, H., Wang, Y., 2021. Structural Wall Facade Reconstruction of Scanned Scene in Point Clouds. *Advances in Electrical & Computer Engineering*, *21*(4).

Pezeshki, Z., Soleimani, A. and Darabi, A., 2019. Application of BEM and using BIM database for BEM: A review. *Journal of Building Engineering*, *23*, pp.1-17.

Pinti, L., Codinhoto, R. and Bonelli, S., 2022. A review of building information modelling (BIM) for facility management (FM): Implementation in public organisations. *Applied Sciences*, *12*(3), p.1540.

Rashdi, R., Martínez-Sánchez, J., Arias, P. and Qiu, Z., 2022. Scanning technologies to building information modelling: A review. *Infrastructures*, *7*(4), p.49.

Roman, O., Avena, M., Farella, E.M., Remondino, F. and Spanò, A., 2023. A semi-automated approach to model architectural elements in scan-to-bim processes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, pp.1345-1352.

Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3), p.160.

Shi, W., Ahmed, W., Li, N., Fan, W., Xiang, H. and Wang, M., 2018. Semantic geometric modelling of unstructured indoor point cloud. *ISPRS international journal of geo-information*, *8*(1), p.9.

Tang, S., Li, X., Zheng, X., Wu, B., Wang, W. and Zhang, Y., 2022. BIM generation from 3D point clouds by combining 3D deep learning and improved morphological approach. *Automation in Construction*, *141*, p.104422.

Tran, H., Khoshelham, K. and Kealy, A., 2019. Geometric comparison and quality evaluation of 3D models of indoor environments. *ISPRS journal of photogrammetry and remote sensing*, *149*, pp.29-39.

Valero, E., Bosché, F. and Bueno, M., 2022. Laser scanning for BIM. *Journal of Information Technology in Construction (ITcon)*, *27*(23), pp.486-495.

Yang, F., Pan, Y., Zhang, F., Feng, F., Liu, Z., Zhang, J., Liu, Y. and Li, L., 2023. Geometry and Topology Reconstruction of BIM Wall Objects from Photogrammetric Meshes and Laser Point Clouds. *Remote Sensing*, *15*(11), p.2856.

Xiaojuan, N.I.N.G., Man, W.A.N.G., Jing, T.A.N.G., Zhang, H. and Yinghui, W.A.N.G., 2021. Structural Wall Facade Reconstruction of Scanned Scene in Point Clouds. *Advances in Electrical & Computer Engineering*, *21*(4).

Yang, X., Lu, Y.C., Murtiyoso, A., Koehl, M. and Grussenmeyer, P., 2019. HBIM modeling from the surface mesh and its extended capability of knowledge representation. *ISPRS International Journal of Geo-Information*, *8*(7), p.301.

Yang, Y., Pan, Y., Zeng, F., Lin, Z. and Li, C., 2022. A gbXML reconstruction workflow and tool development to improve the geometric interoperability between BIM and BEM. *Buildings*, *12*(2), p.221.

Zhang, K., Teruggi, S., Ding, Y. and Fassi, F., 2022. A Multilevel Multiresolution Machine Learning Classification Approach: A Generalization Test on Chinese Heritage Architecture. *Heritage*, *5*(4), pp.3970-3992.