ARCHITECTURE CIVIL ENGINEERING ENVIRONMENT
The Silesian University of Technology
No. 3/2024

doi: 10.2478/ACEE-2024-0022

EXPLORING MACRO BIM FOR COST PREDICTION: LEVERAGING SIMILARITY ASSESSMENT METHODS AND MARKET DATA

Hanna PACYNO [a*], Michał JUSZCZYK [b]

[a] MSc, Eng. Arch.; Cracow University of Technology, Faculty of Civil Engineering, ul. Warszawska 24, 31-155 Cracow, Poland
ORCID: 0000-0001-9765-0031
*Corresponding author. E-mail address: *h.pacyno@datacomp.com.pl*

[b] DSc, PhD, Eng.; Cracow University of Technology (CUT), Faculty of Civil Engineering, ul. Warszawska 24, 31-155 Cracow, Poland
ORCID: 0000-0002-8353-9823
*E-mail address: *michal.juszczyk@pk.edu.pl*

Received: 22.09.2023; Revised: 01.03.2024; Accepted: 25.04.2024

Abstract
BIM Technology allows for multidirectional analyses using the information models of building facilities. One of the areas where it is used is cost calculations, which fall under the so-called BIM 5D. This article explores the Macro BIM concept, which varies in interpretation according to literature, and its practical application. It discusses using Macro BIM model data alongside market construction data to simulate estimating building construction costs. The aim of the article is to simulate the process of calculating the estimated costs of building construction, according to the proposed course of finding the most similar objects already built to the one that is planned and calculating the unknown sought on the basis of known "reference" data. Creating a database linking market costs with building characteristics would facilitate a classification system for such surveyed buildings. Finding the most similar "reference objects" to the studied "sample" object would allow for quick, preliminary cost calculations based on the cost per unit of volume or area of the designed building. The results obtained in this preliminary study are satisfactory and have provided insight into future directions that require more detailed examination.

Keywords: **Building information modeling (BIM); Macro BIM; Construction cost management; Cost estimates.**

1. INTRODUCTION

The BIM (Building Information Modeling) technology allows participants in a construction project to exchange and process information to achieve specific objectives based on digital models of building structures. Implementing BIM technology transforms the traditional way of understanding and managing many crucial processes during a construction project. A special issue is the use of BIM in the field of cost analyses to which, in accordance with the convention of the so-called "multi-dimensionality", the common term BIM 5D is applied. In other words, a 5D "dimension" corresponds to cost information. The Macro BIM concept is integral to the discourse on cost analysis using BIM. It's worth noting that this term appears in literature, including sources referenced later, and in professional jargon in various formats such as macro BIM, MacroBIM, or macro model. For the sake of consistency, this paper adopts the Macro BIM notation.

The objective of this article is to provide a conceptual analysis, overview, and discussion of the Macro BIM concept. It also aims to highlight its potential in the realm of research and the development of models that aid building cost estimation. Furthermore, the article presents preliminary research results as a case study on cost estimation for educational facilities using the

Macro BIM concept, considering market data and cost indicators.

The aim of the article is to show the possibility of using real data for the purposes of preliminary cost calculations of the planned facility for verification of the initial assumptions of the requirements in relation to financial possibilities (budget). Calculations performed according to unit costs per m3 are carried out according to the selection of "most similar" objects whose construction cost is known. The simulation on a small collection of data shows the investigation (search) for the most similar objects to the object whose construction cost is "unknown" – "searched". The object whose production cost is calculated has features that are determinants of similarity relative to other "reference" objects. Thus the object described by the features and geometry size [m^3] is listed in the computational example as a "sample". It is worth mentioning that the cost of construction referred to the object called "sample", is known, and only for simulation purposes it is defined as "unknown" (the searched value). Thanks to this, after the process of searching for the most similar objects and then calculating, it is possible to compare and confront the calculated value with the actual one.

Cost analyses in the early phase of a construction project, such as those in line with the Macro BIM concept, are designed to support the investment process. Their goal is to assess the feasibility of a proposed project before deciding on its execution. Such analyses also allow for the evaluation of potential changes in the initial concept, ensuring they're considered before incurring significant additional costs. Additionally, they facilitate the selection of the most economically viable option among the compared concepts.

2. MACRO BIM CONCEPT – LITERATURE REVIEW

The "Macro BIM" concept was introduced by Beck Technology in conjunction with the launch of the DProfiler program in 2008. The creators identified a gap in the BIM software available at that time. This program introduced a database featuring 20,000 items, each detailing the costs of individual elements. The data provided, when combined with the dimensions of the rooms, enabled users to estimate the current value of the concept [1].

While the concept originally pertained to cost analysis, the term "Macro BIM" has since taken on various meanings and is used in different contexts and areas. In literature [2, 3], for instance, it is referenced in relation to the evaluation of BIM technology implementation at regional or national levels. Additionally, a variant termed "Macro BIM+" [1] is proposed for cross-sector coordination and simultaneous design purposes, utilizing a schematic model with the most important, characteristic parameters.

Most commonly, Macro BIM refers to the concept of utilizing low-detail BIM models for rapid cost analysis purposes [4]. Macro BIM can be described as the process of creating a model with embedded (or associated) data for budget planning or feasibility studies [5, 6, 7]. The majority of publications concur that the accuracy level is typically defined as LOD (Level of Development) 100 or 200, though LOD 300 is also deemed acceptable [5, 8, 9, 10]. Some publications [4, 11] stress the importance of merging such models with an integrated price list or a database of aggregated costs, aiming for the utmost automation in the estimation process. In articles focusing on Macro BIM, a pivotal aspect seems to be not just appending information but also forging connections between diverse data sources [12]. The Macro BIM concept is occasionally discussed in a wider context, with sources like [13] envisioning the fusion of preliminary information with algorithms and dependency conditions to facilitate global changes. Several authors highlight the significance of cost indicators, which they believe are essential in the Macro BIM methodology [4, 5, 10]. Some publications underscore the necessity for precise measurements of model components and calculations grounded in geometric data, such as floor space, total area, or volume [4, 10]. There's also an expressed need to employ a classification system [11] or to associate model components with a taxonomy [9].

To support cost evaluations consistent with the Macro BIM approach, various mathematical models have been proposed. One such model, for instance, is based on the CBR (Case Based Reasoning) method combined with the use of fuzzy set theory and BIM technology [4]. Other studies [9, 12, 14] contemplate the incorporation of AI (artificial intelligence) tools, such as ANN (artificial neural networks), to aid in cost analyses. For instance, the authors in [14] detail the application of ANN, varying by network architecture and variable definitions. According to research on the application of ANNs [4, 12, 15], as well as comparisons involving diverse mathematical models [4], it emerges that cost estimates using ANNs in the early stages of construction projects exhibit "low estimation errors", staying within acceptable bounds and thus ensuring a commendable prediction quality.

This validates the ongoing research into the application of ANN models and other AI models concerning cost prediction systems and underscores the importance of targeted research in specific areas. Literature also offers computational examples which illustrate and dissect the application of the Macro BIM concept and the use of ANN, such as a case study analyzing the load-bearing structure of a building floor [14], or studies evaluating the quality of cost predictions based on data from residential building implementations [12].

In Poland, there's an increasing number of projects being realized using BIM technology, encompassing both design and construction phases. While some of these are pilot projects, others serve as sources for analyses based on the collected data [16, 17, 18]. Conferences centered around construction digitization [19, 20] serve as platforms that promote the exchange of experiences with BIM technology among various participants in the investment and construction processes. These gatherings also act as catalysts for change within companies and the broader society. There are also comprehensive studies [5, 21, 22] showcasing "good practices", elucidating the BIM methodology across different stages of the investment and construction process. Public consultations [18] and research [23] are being organized to aid in the potential establishment of universal Polish standards. It's noteworthy that various facets related to BIM technology, such as market-aligned classifications, are being addressed by organizations like buildingSMART Poland and specialized working groups under the Ministry of Development and Technology.

One of the documents issued as "support of the Ministry of Development for the development of an integrated BIM strategy for the construction process in public procurement" [5], serves as an exemplar of good practices. It provides direction for suggested improvements in the construction sector's efficiency and offers proposals for the implementation of the BIM methodology. This is further encapsulated in the "Roadmap for the Implementation of BIM Methodology in Public Procurement" [5]. Within this study, the Macro BIM concept is characterized as a phase in the investment process that ensures economic viability and functions as an integral part of the public procurement process. This approach receives expanded coverage later in this paper. The publication [5] also features a graphic illustrating a sample Macro BIM model (Figure 1) with demonstrated exemplary level of detailing. It is evident that the model is volumetrically organized according to functions, encompassing information on grouped surfaces and volumes for the purpose of cost indicator calculations within the Macro BIM phase.

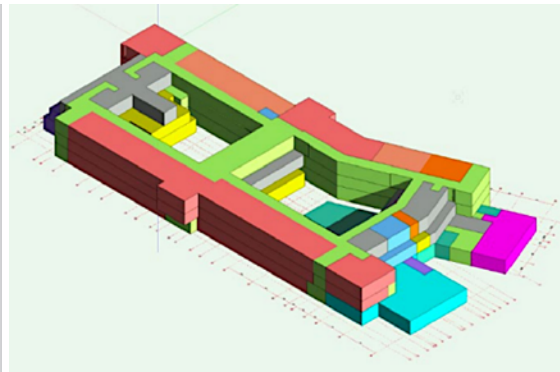

Figure 1.
The exemplary level of Macro BIM model detailing. Graphic source: [5]

Considering the referenced literature, it is possible to identify several key elements that are common to most of them. These include: a low level of detail in the model, the use of information for cost estimation purposes, efforts to associate geometric volume data with cost information, and the application, testing, or pursuit of optimal mathematical models for the stated objectives.

For the purpose of this research, a definition from the Roadmap [5] was adopted. In this source, the Macro BIM concept is presented within the broader context of the investment process, positioned as one of its preliminary phases. This inclusion has significant rationale, ensuring the protection of the investment, financial security, and preparation for subsequent stages.

3. MACRO BIM AS A PHASE OF A CONSTRUCTION PROJECT

According to the stipulations set out in the Roadmap [5], the Macro BIM phase is envisioned as an added phase of the investment, safeguarding its economic aspects while also assembling teams of specialists from potential contractors to foster collaboration and self-organization. In the Roadmap [5], the MacroBIM phase is integrated into the public procurement procedure. The proposed phase in the investment process "Macro BIM phase" is an additional element of the investment programming stage utilizing BIM technology. As proposed in the referenced document, this supplementary approach should be mandatory for complex investments that

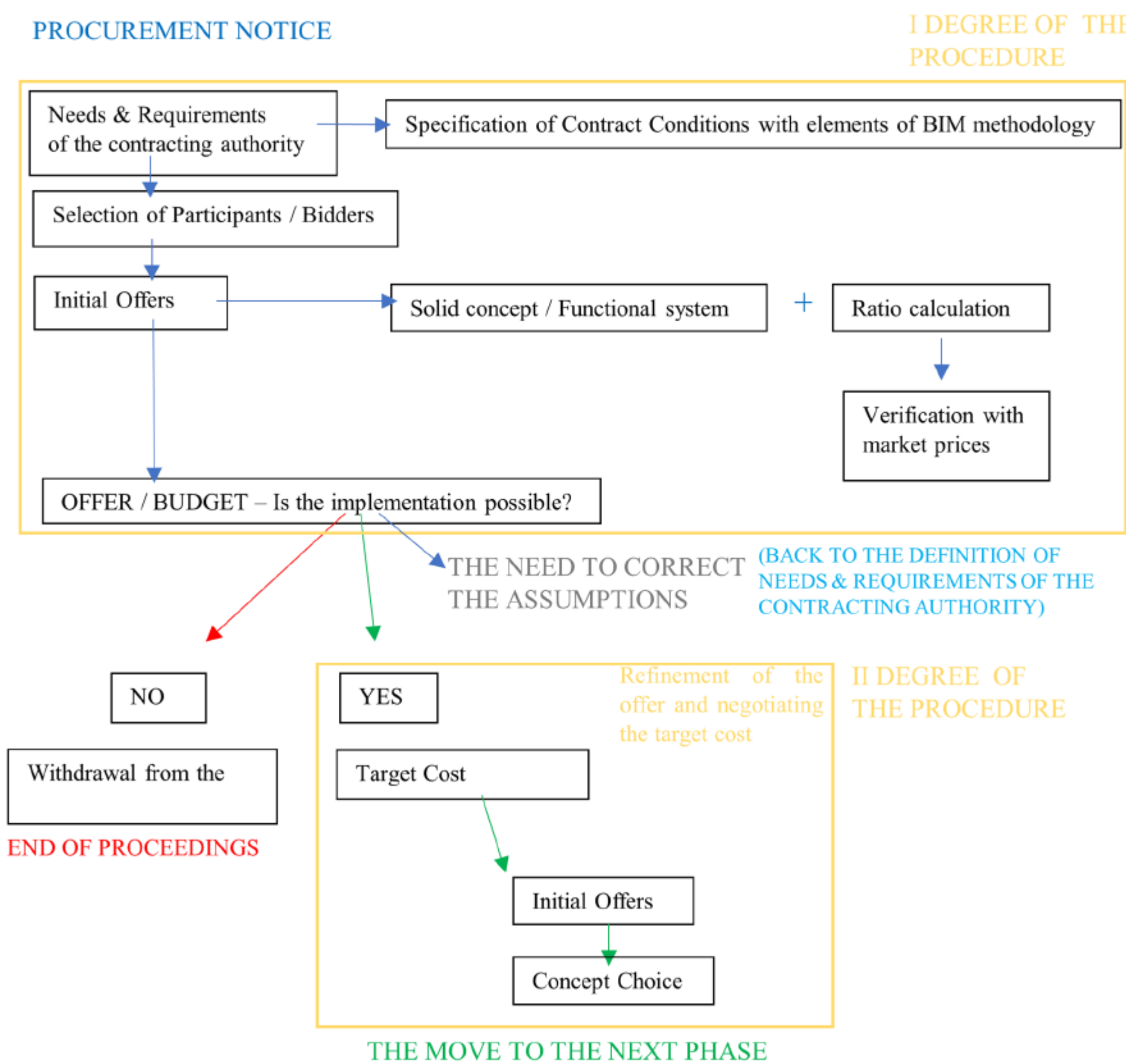

Figure 2.
Macro BIM process scheme. Authors' study based on [5]

carry a potential risk of failure and have budgets exceeding EUR 10 million.

According to the convention proposed in the Roadmap [5], after announcing the procedure, participants wishing to enter the competition are assessed based on several criteria, such as: substantive (e.g., team experience with investments using BIM technology, execution of similarly complex facilities), economic, and organizational. A select group of bidders is invited to present their preliminary offers, which should align with the needs and stipulations set by the Employer. The initial proposals provided by the bidders encompass concrete concepts or functional guidelines, prepared in line with the Contract Conditions Specification drafted by the Employer, as well as estimated costs derived from these, presented as index calculations. The necessary concepts or functional guidelines for the Macro BIM phase possess specific features. They're characterized as low-detail LOD (Level of Development) models at 100 or 200, with allowance up to LOD 300. Estimated cost calculations are based on benchmark values converted to per square meter (m^2) of gross/net function, per cubic meter (m^3) volume, and unit computations. Per the Roadmap, these calculations ought to be drawn from generalized data for grouped functions, with a suggestion to segment them by floor, without detailing individual partitions, the dimensions of openings, etc. Index costs are provided within a minimum-maximum spectrum, and bidders are expected to verify these by comparing with market prices for analogous investments.

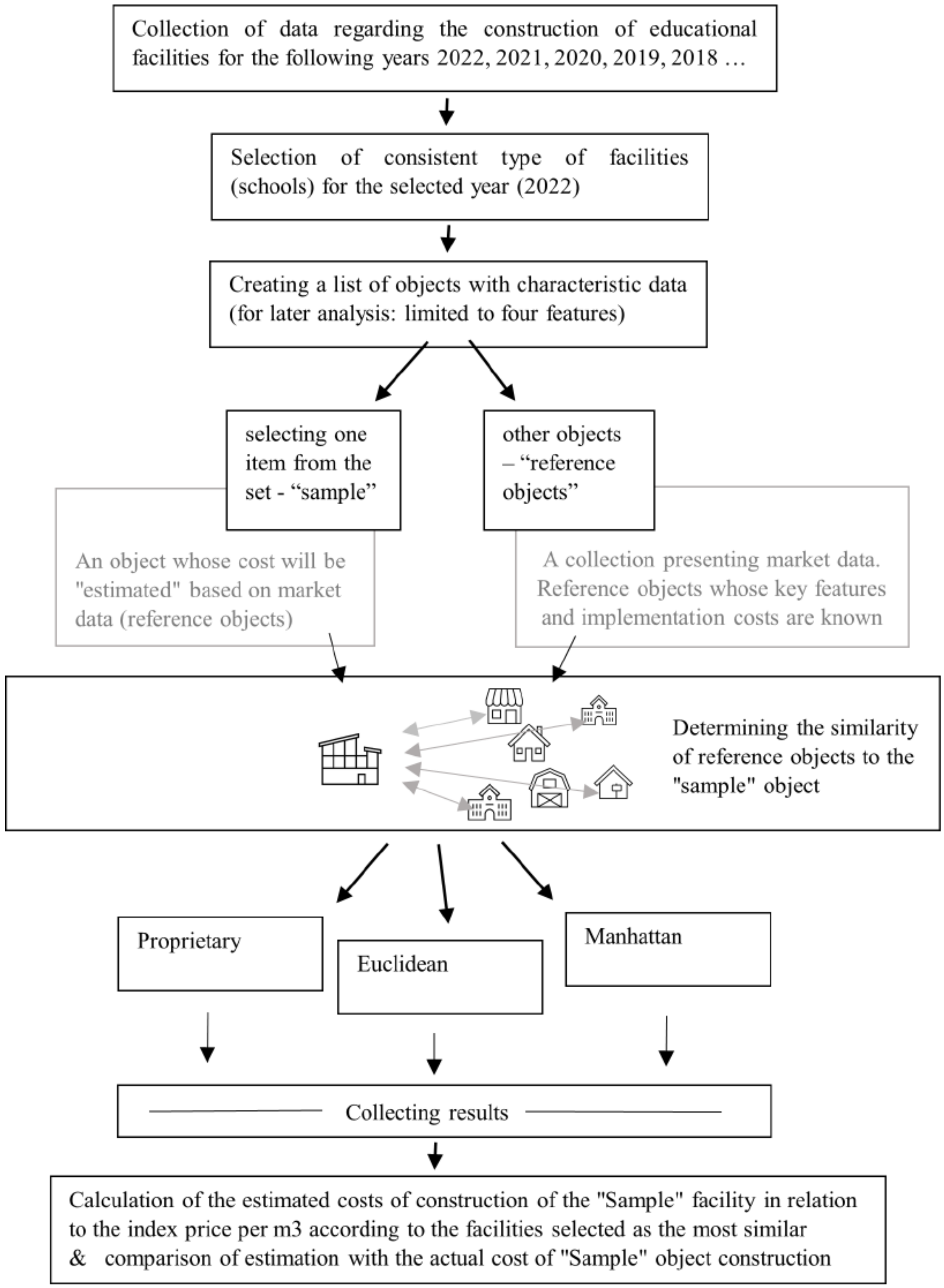

Figure 3.
The scheme presents the objectives and following steps of case study

Based on an analysis of the content provided by the contractors, the objective is to determine whether the investment is financially feasible. If the answer is negative and the estimated costs significantly surpass the economic capacity of the contracting authority, the contract will be withdrawn, and the process terminated. A positive response will lead to the next step, which involves negotiating the target cost (the second stage of the procedure). Upon completing the first stage of the procedure and evaluating the initial offers, it may also become evident that adjustments to the initial assumptions are necessary. At this juncture, the process returns to its initial point, involving the definition of the Specification of Contract Conditions and subsequent steps. Following the negotiation phase, if the offer is approved, the Target Cost is established. This agreed-upon cost acts as the boundary barrier for the project and aligns with an approach in which the starting point is an agreed final cost, rather than calculating costs for the resulting project.

As per the Roadmap [5], the Macro BIM phase is expected to culminate in presenting the concept to the Contracting Authority, utilizing the Target Cost as the foundation for proceeding to the subsequent phase – the capital phase.

The distinct steps of the Macro BIM phase can be depicted through a diagram (Figure 2), illustrating the points at which particular "deliverables" are generated and provided (requirements, calculations, models). This diagram facilitates the tracking of essential key points.

4. CASE STUDY OF MACRO BIM COST ANALYSIS

The case study was conducted by collecting market data, analyzing it, processing it (data selection and its categorization/interpretation) and dividing the set of items into two parts: "estimated object" and a set of reference objects. The calculation of the estimated implementation cost was assumed as an index calculation per unit m^3 of the facility's cubic capacity, based on the costs of the most similar market facilities (reference objects). The similarity analysis was carried out using three methods to check the consistency of the selection. Selected objects – the most similar to the estimated "sample" object – were used as data for the index calculation in relation to the volume PLN/m^3. The diagram presented in Figure 3 shows the next steps of the presented case study.

In order to conduct the study, data pertaining to the construction of volumetric structures for educational purposes were gathered for subsequent years. Based on the raw data, an initial analysis was conducted, encompassing investments in the establishment of primary schools. The data sources comprise publicly available information and documents associated with the execution of public tenders. Particularly noteworthy information considered when formulating the characteristics table included:

- data identifying individual instances (identified by an individual ordinal number e.g., LP13);
- object locations – categorized as Urban or Rural;
- precise foundational measurements: volume and usable area;
- Investment values according to the investor (the sum designated by the contracting authority for the contract / initial estimated contract value);
- contractual price (based on the successful, most advantageous bid);
- classification of implementation size based on area, grouped as: Small – up to 1000 m^2, Medium – 1000 m^2 to 5000 m^2, Large – over 5000 m^2;
- number of floors (categorized into three groups: single-storey, 2-3 storeys, multi-storey);
- additionally, consideration was given to supplementary attributes that significantly influence implementation complexity and cost.

Using variables related to cubic capacity, area, estimated work values, and contract prices, unit values per m^2 and m^3 were computed, along with average values derived from disparities between the investor's appraisals and the successful bids. The table below (Table 1) presents the source data and the computed price per m^3 based on the Contractual price.

In addition to the gathered data, the accessible design documentation also played a crucial role. This enabled the generation of object volumes in alignment with the Macro BIM concept for selected items. The objective of Macro BIM modeling is to streamline and generalize spatial models, with the highest level of detail encompassing the volumetric aspects segmented into grouped functions by floor (Figure 4). In the provided example, gross cubic volumes were established for individual floors (without further division into functions)(Figure 5).

For the purpose of research, one specific object was selected as a "sample". This sample represents a hypothetical object for which a "valuation" is sought. In alignment with the Macro BIM Concept, the model would serve as a representation of the proposed volume for the designed object. This model would incorporate information about its distinctive attributes in accordance with the requirements outlined in the Specification of Contract Conditions of the contracting party.

In accordance with the Macro BIM concept, the bidder submitting the tender documents is required to furnish the Macro BIM model along with the corresponding cost proposal.

Using the provided foundational values and supplementary characteristics, the contracting authority could autonomously compute estimated costs for

Table 1.
List of primary school facilities. Selected data and cost values [PLN] per unit [m^3]

ID	Value of works according to the investor / initial estimated value [PLN]	Contract price [PLN]	Building volume [m^3]	Investment value / volume [m^3]	Contractual price per volume (price per m^2)	Object size according to area [m^2] (S/M/L)	Number of storeys (O/F/M)	Location	Location Type (Rural, Urban)
LP3	43638488.13	46358700.00	37536.54	1162.56	1235.03	L (6 166.40 m^2)	F - 2	Środa Śląska	U
LP13	9310000.00	10091905.01	7244.94	1285.03	1392.96	M (1269.67 m^2)	F - 2	Dopiewo	R
LP53	48474598.56	50899568.95	38422.20	1261.63	1324.74	L (7110.20 m^2)	F - 3	Warszawa	U
LP103	60121311.66	68861788.62	45515.50	1320.90	1512.93	L (11957.30 m^2)	F - 3	Warszawa	U
LP107	11685000.00	13499681.03	25739.00	453.98	524.48	M (4360.24 m^2)	F - 2	Skokowa	R
LP113	72281756.60	74057723.57	49916.50	1448.05	1448.05	M (11737.00 m^2)	F - 2	Warszawa	U
LP122	2212520.91	2691987.73	1229.00	1800.26	2190.39	S	O - 1	Czarnów	R
LP123	56910569.11	62448078.09	-	-	-	L	M - 3	Toruń	U

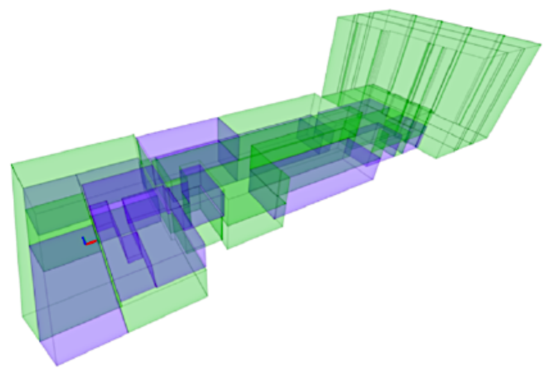

Figure 4.
An example of a Macro BIM model divided into grouped functions

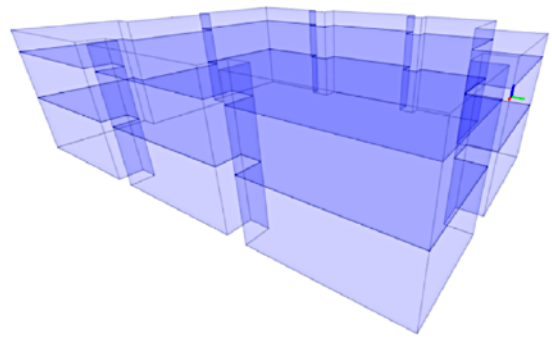

Figure 5.
An example of a Macro BIM model divided by floors

negotiation purposes or to compare various proposal variations, regardless of the furnished "offers".

In the provided example, the "sample" object represents an actual investment with known incurred costs. This allows for a comparison between the final projections of costs and the actual expenses when determining the estimated valuation. The "sample" model in this illustration pertains to a school facility identified in the table by the unique ID number: LP53.

Next to the "sample" object, the example also incorporates data related to "reference objects". These reference objects serve as sources of cost-related information. The collection comprises data concerning objects with an educational purpose (similar to that of the sample), offering characteristic details, including cost-related information.

4.1. Searching for the most similar reference objects to the "Sample" object

Below is an analysis of the similarity of reference models to the "estimated" object – the Sample model. The analysis is conducted using three different methods. The proprietary method of analyzing a set of objects (Table 1) by comparing distinguished features, which was possible to apply due to the small dataset – limited number of features and reference objects. And also by two methods that fall under what is known as Cluster Analysis: the Euclidean distance method, and the Manhattan distance method.

For the purpose of the first analysis – the proprietary method, specific variables were identified and contrasted, leading to the identification of two reference objects that resemble the sample. Using the foundational data, the individual volumes of the reference objects were compared, calculating the smallest disparity relative to the sample volume (Table 2). A smaller difference indicated a closer similarity to the sample. Similarly, the surfaces of the reference objects were compared to the sample surface in an analogous manner. Given that there were seven reference objects, with one being excluded due to insignificant parameters, these two attrib-

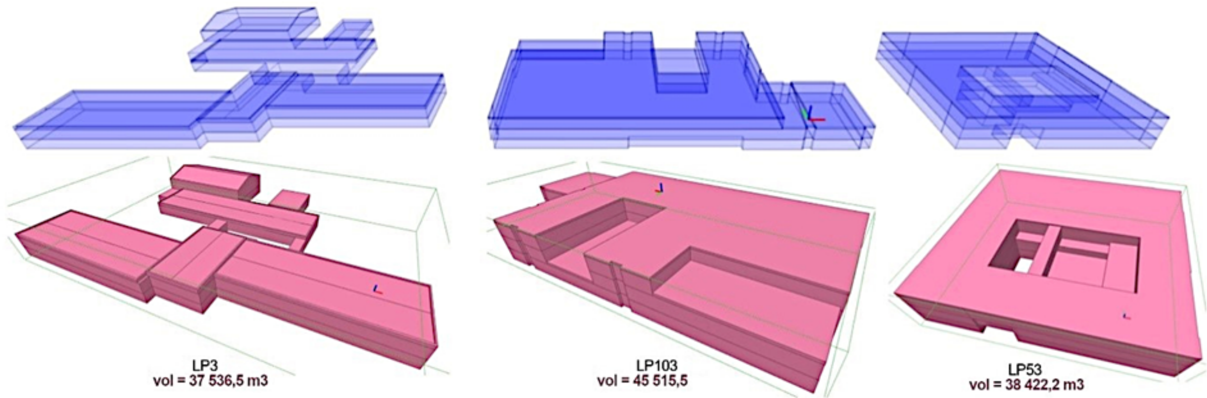

Figure 6.
The two most similar Macro BIM models (LP3 and LP103) and the Macro BIM Sample model (LP53)

Table 2.
Comparison of geometric values of reference objects compared to the sample (differences)

Sample:									
LP53	38 422.20		L (7110.2 m^2)		F - 3		Warszawa	U	
reference objects (compared):									
ID	Building volume [m^3]	Percentage of diff. of the volume criterion from the sample	Object size according to area [m^2] (S/M/L)	Percentage of diff. of the surface criterion from the sample	Number of storeys (O/F/M)	The diff. in the number of storeys in relation to sample	Location	Location Type (Rural. Urban)	Characteristics of the place
LP3	37536.54	2.31	L (6 166.40 m^2)	13.27	F - 2	diff.: 1	Środa Śląska	U	urban
LP13	7244.94	81.14	M (1269.67 m^2)	82.14	F - 2	diff.: 1	Dopiewo	R	rural
LP103	45515.50	18.46	L (11957.30 m^2)	68.17	F - 3	the same	Warszawa	U	same (the capital)
LP107	25739.00	33.01	M (4360.24 m^2)	38.68	F - 2	diff.: 1	Skokowa	R	rural
LP113	49916.50	29.92	M (11737.00 m^2)	65.07	F - 2	diff.: 1	Warszawa	U	same (the capital)
LP122	1229.00	96.80	S	huge diff.	O - 1	diff.: 2	Czarnów	R	rural
Lp123	-	-	L	-	M - 3	the same	Toruń	U	urban

Table 3.
Sorting by ranking according to the smallest differences for each 4 criteria. Analysis of the best (smallest) results in terms of volume and area values as well as the number of floors and location

	scale 1-6	scale 1-6	scale 1-3	scale 1-3		
ID	Rank position (smallest diff.) by volume criterion	Rank position (smallest diff.) by area criterion	Ranking position of the number of storeys compliance	Location type compatibility ranking	volume and area criterion	The criterion of the number of storeys and location
LP3	1	1	2	2	2	4
LP13	5	5	2	3	10	5
LP103	2	4	1	1	6	2
LP107	4	2	2	3	6	5
LP113	3	3	2	1	6	3
LP122	6	6	3	3	12	6
LP123	-	-	1	2	-	3

utes of volume and area were evaluated through ranking items from 1 to 6 (Table 3).

The next two variables subjected to comparison are the number of floors and the location. Information regarding these variables has been classified into three groups. Concerning the floors, these groups encompassed: single-story, two-/three-story, and multi-story structures. As for the locations, they were categorized as village/city or the classification “same location as the model" (primarily because the sample object is situated in the capital city, a distinctive urban setting). These two subsequent indicators for floors and location were evaluated on a 3-tier scale (Table 3). A smaller value denoted a greater resemblance between the reference object and the sample.

When examining the sums of the volume and surface criteria (on the same scale – ranked from 1 to 6), the object achieving the lowest value (and thus the closest to the sample) was identified as: LP3 (score of 2, the smallest value). Following this, several objects simultaneously attained the next lowest value, summing up to 6: LP103, LP107, and LP113. When evaluating the subsequent two criteria, namely the number of floors and location, the object LP103 achieved the lowest value, trailed by LP113 and LP3. Through a comparison of these four criteria (volume, area, floors, location), two objects that were "the most similar", namely LP3 and LP103, were chosen (Figure 6)(Table 3).

Analysis of the best (smallest) results in terms of volume and area values as well as the number of floors and location indicated the two entries. One of them marked as LP3 that is close to the Sample by terms of volume, similar number of storeys, and the urban location and the other entry marked as LP103 that is similar in terms of volume, the same – capital location, the same number of floors.

To verify and confirm that indeed the two most similar objects have been selected in terms of the few criteria presented and taken into consideration, the data are going to be analyzed using two another, different methods. Both of them fall under topics related to so-called Cluster Analysis, which concerns the issue of data classification and are clustering algorithms [24]. The first method of them is the Euclidean distance method, and the second is the Manhattan distance method. The Euclidean distance method involves measuring the geometric distance in multidimensional space. It's worth noting that "distances are calculated based on raw data, not standardized" [24].

To enable calculations, non-numerical data must be processed. Values of variables related to the number of floors and location, which have been categorized according to their membership in specific groups (floors: single-story building, two or three floors, multi-story building; location: village, city, capital) will be coded as follows:

- Floors: single-story = [1, 0, 0], 2 or 3 floors = [0, 1, 0], multi-story = [0, 0, 1]
- Location: village = [1, 0, 0], city = [0, 1, 0], capital = [0, 0, 1]

Next, for each of the reference objects shown in the table (Table 4), we calculate the Euclidean distance [24] relative to the sample.

$$\text{Distance}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{1}$$

The notation of individual features for the objects in the example is in the form: [volume in m^3, area in m^2, floors according to encoding, location according to encoding]. Notations for individual objects , as well as their resulting Euclidean distance between the referenced objects and the Sample LP53 object, are presented in Table 6.

*In the table the area value of Object 6. (LP122) was assumed as 500 because its exact value is not known.

*Object 7. (LP123) were rejected

The smallest Euclidean distance is for objects LP3 and LP103, which means the selection of the most similar reference objects is consistent with previous

Table 4.
Basic data for calculations about the sample and reference objects

sample:					
LP53	38 422.20	L (7110.2 m^2)	F - 3	Warszawa	U
reference objects:					
ID	Building volume [m^3]	Object size according to area [m^2] (S/M/L)	Number of storeys (O/F/M)	Location	Location Type (Rural, Urban, Capital)
LP3	37 536.54	L (6 166.40 m^2)	F - 2	Środa Śląska	U
LP13	7244.94	M (1269.67 m^2)	F - 2	Dopiewo	R
LP103	45 515.50	L (11957.3 m^2)	F - 3	Warszawa	C (same as sample)
LP107	25 739	M (4360.24 m^2)	F - 2	Skokowa. gmina Prusice	R
LP113	49 916.50	M (11737 m^2)	F - 2	Warszawa	C (same as sample)
LP122	1 229	S *	O - 1	Czarnów. Gmina Górzyca	R
LP123	-	L	M - 3	Toruń	U

Table 5.
The notation of individual features and the resulting Euclidean distances for the objects (sample LP53 object – reference model indicated by ID)

	ID	Notation of individual features (vol.[m^3], area [m^2], encoded floors. encoded location)	Euclidean distance for the object
	LP53	38 422.2. 7 110.2. 0. 1. 0. 0. 0. 1	(sample object)
1.	LP3	37 536.54. 6 166.40. 0. 1. 0. 0. 1. 0	1 294.27
2.	LP13	7244.94. 1269.67. 0. 1. 0. 1. 0. 0	31 719.60
3.	LP103	45 515.50. 11957.3. 0. 1. 0. 0. 0. 1	8 591.23
4.	LP107	25 739. 4360.24. 0. 1. 0. 1. 0. 0	15 434.16
5.	LP113	49 916.50. 11737. 0. 1. 0. 0. 0. 1	12 390.57
6.	LP122	1 229. 500. 1. 0. 0. 1. 0. 0	37 776.04

calculations. However, it's worth noting that due to the lack of standardization of input data, the form taken for analysis matters. Thus, it can be concluded that the two criteria – location and number of floors – had negligible influence on the result. For this reason, an analysis using the third method, which includes normalization of input data, is also important. It takes to account the normalization of input data, so it can be considered as more objective and less dependent on the units/type of input data.

Another method of Cluster Analysis is the "Manhattan" city block distance. Below (Table 6), selected objects have also been analyzed using this method.

To apply the method, a matrix of criteria values was created for individual positions. Each position represents one of the objects. The first position – the first row – pertains to the "sample", and the following ones to the reference objects.

For the volume and area criteria, actual values were taken into account. For the number of floors, if the object had two or three floors (equal to the value of the sample), it was given a value of 1. If it was a single-story building or had more than three floors, it was given a value of 0. For the Location criterion, the city or village was given a value of 0, and for the capital (as in the sample), it was given a value of 1.

Subsequently, individual values (Table 6) were normalized (Table 7) according to the following formula:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m}(x_{ij})^2}} \quad (2)$$

As a result, individual values were brought closer to each other and thus became "comparable".

The next step was to calculate the distances for individual criteria in multidimensional space according to the Manhattan method.

Table 6.
Values before normalization

	Volume	Area	Storeys	Location
Sample	38 422.20	7110.2	1 (3)	1
LP3	37 536.54	6 166.40	1 (2)	0
LP13	7244.94	1269.67	1 (2)	0
LP103	45 515.50	11957.3	1 (3)	1
LP107	25 739	4360.24	1 (2)	0
LP113	49 916.50	11737	1 (2)	1
LP122	1 229	500	0 (1)	0

Table 7.
Values after normalization

	Volume	Area	Storeys	Location
Sample	0.425208809	0.359952511	0.4082483	0.57735
LP3	0.415407433	0.312172817	0.4082483	0
LP13	0.080177926	0.0642768	0.4082483	0
LP103	0.503708573	0.605336019	0.4082483	0.57735
LP107	0.284847029	0.220736314	0.4082483	0
LP113	0.552413332	0.594183374	0.4082483	0.57735
LP122	0.013601033	0.025312404	0	0

$$\text{Distance}\ (x, y) = \sum_i |x_i - y_i| \quad (3)$$

To differentiate the significance of the criteria and to emphasize those that are subjectively considered most important for finding the most similar objects, weights were calculated based on the importance analysis of individual pairs of criteria. As can be seen from the matrix (Table 9), volume is determined as slightly more important than area. Less important than they are the criteria of location and the number of floors (with the location criterion considered slightly more important than the number of floors).

Table 8.
Determining the importance of criteria – Matrix for weight calculations

	Volume	Area	Storeys	Location
Volume	1	2	3	3
Area	0.5	1	3	3
Storeys	0.333333	0.333333	1	0.5
Location	0.333333	0.333333	2	1

The calculated weights from the above matrix (Table 9) are as follows:
Volume Criterion = 0.435082
Area Criterion = 0.309207
Number of Floors Criterion = 0.105633
Location Criterion = 0.150078

When calculating the sum of distances and comparing them with the sample, assuming that all 4 criteria are equivalent, the result (Table 9) differs from the

Table 9.
Application of the Manhattan Method on normalized values without taking into account weights. The table highlights the three best results

	Volume	Area	Storeys	Location	Distances
LP3	0.009801376	0.047779694	0	0.57735	0.634931339
LP13	0.345030883	0.295675711	0	0.57735	1.218056863
LP103	0.078499764	0.245383508	0	0	0.323883272
LP107	0.14036178	0.139216197	0	0.57735	0.856928246
LP113	0.127204523	0.234230862	0	0	0.361435385
LP122	0.411607775	0.334640107	0.4082483	0.57735	1.731846442

Table 10.
Application of the Manhattan Method on normalized values. including calculated weights. The table highlights the two best results

	Volume	Area	Storeys	Location	Distances
LP3	0.004264398	0.014773838	0	0.086647	0.105685637
LP13	0.150116583	0.091425135	0	0.086647	0.328189119
LP103	0.034153802	0.075874411	0	0	0.110028213
LP107	0.061068826	0.043046687	0	0.086647	0.190762913
LP113	0.055344345	0.07242593	0	0	0.127770275
LP122	0.179082963	0.103473217	0.0431246	0.086647	0.412328177

calculations taking into account differentiated weights (Table 10). However, it's worth noting that under the assumption of equivalent weights – even though the 2 closest solutions don't overlap, they are among the three closest ones. The three "best" results are similar to each other.

Depending on the determination of the "importance" of the criteria, the results may vary, so one should carefully define the relationships between the criteria (their significance). The results of applying the above Manhattan method are the same or very similar to those obtained using the first comparative method and the Euclidean length method presented earlier.

4.2. Calculations based on the most similar referenced objects to "Sample" object

Using the entries of objects LP3 and LP103 which were highlighted as most similar reference objects to the sample object LP53 and confirmed using three different methods, the projected price for the tested object was computed. Based on the unit price of PLN 1,235.03 per m^3 for the object with the ID number: LP3, the cost for the volume of 38,422.20 m^3 of the tested object LP53 is calculated as PLN 47,452,569.67. Similarly, considering the rate per m^3 of – PLN 1,512.93 for the reference object LP103, the total amount were calculated to PLN as 58,130,099.05. The average of these two derived sums equates to PLN 52,791,334.36. Upon comparing this estimated value to the actual project value of PLN 50,899,568.95, it is evident that the estimation is close. The difference between the actual and the projected cost, calculated through the average calculation using data from the two reference objects, is approximately 4% for the presented pattern.

5. DISCUSSION OF CASE STUDY RESULTS AND FUTURE STUDIES ON THE PROBLEM

The paper introduces the Macro BIM concept and provides an illustrative calculation of estimated costs using real-world market data. The approach demonstrated in the example is straightforward and relies on a limited dataset. Nevertheless, the intended reference object database will encompass several hundred items, each described by characteristic variables that differentiate and define them.

Considering the Macro BIM procedure, as well as the overarching concept of furnishing a specialized BIM model and establishing the foundation of reference objects alongside their classification and distinguishing features, numerous questions and avenues emerge, each of which will be sequentially examined.

One of the crucial matters involves establishing clear guidelines for Macro BIM modeling. Drawing from the Roadmap and other resources, these guidelines could encompass grouping functions by floor or potentially adopting an even more generalized partitioning of spaces per floor. It is imperative to define best practices for determining how functions should be grouped (if applicable), how generalization should occur, how to filter data, which data to exclude, and which data requires special emphasis.

An essential matter involves conducting an analysis to identify which features of the facilities most effectively define their characteristics and exert the most significant influence on cost generation.

Another question pertains to the selection of modifiers for certain variables that could arise in both the estimated object and the reference objects. In cases where such attributes are present, the calculations (multipliers) need to be adapted to ensure that the projected outcome aligns as closely as possible with the anticipated actual value.

Addressing these specific inquiries and establishing a database of Macro BIM models alongside pertinent information will enable the development of a model for estimating costs of planned educational facilities. This model could leverage neural networks to identify and select the most analogous objects, which would directly inform estimated computations. In this model, the first step would be to identify the most comparable reference facilities to the intended investment. Anticipatedly, the cost prediction arising from computations grounded in cost data of the most akin real reference objects (factoring in potential additional modifiers) will boast a notably high level of accuracy concerning indicator calculations. Furthermore, when evaluating alternative projects of varying sizes and characteristic variables, using a single tool will facilitate comprehensive comparisons between different variants.

The outlined operational concept would involve inputting the characteristic variables of the estimated object (in accordance with the Specification of Contract Conditions) and then identifying, using these variables, the most analogous reference objects from the database. The characteristic variables employed for object selection must remain consistent and possess a significant impact on the diversity of objects and their associated cost value. This impact could stem from factors such as costly materials, intricate design, time-intensive implementation, or distinct technological solutions.

Utilizing data regarding the unit price per m^3 of the reference facility/objects, the projected cost of the facility would be computed. Additionally, this estimate could be refined by accounting for location discrepancies between the facilities or by standardizing the price fluctuations that have occurred over the periods corresponding to the estimated (tested) facility and the reference facility.

6. SUMMARY AND CONCLUSIONS

While reviewing the bids submitted in individual tenders, a notable disparity in the proposed implementation costs from various bidders becomes evident. The method of rapidly assessing costs based on preliminary estimates in comparison to real market data will enable an initial evaluation of the proposed concept.

The initial analysis enabled an examination of individual implementation instances (gathering, analyzing, and organizing documentation) and highlighted areas that demand specific attention. Subsequent efforts will entail: broadening the scope of "characteristics" criteria, expanding and organizing the reference object database systematically, constructing a model using neural networks to assess input data (estimated object) and categorize the presented case, exploring the feasibility of incorporating modifiers and defining the circumstances under which they will be applied, and verifying the accuracy of predictions grounded in reference object data and supplementary modifiers.

In the process of developing a system in the form of a prototype module for estimating computations, the option of automating the batch data processing and presenting the outcomes will also be taken into account.

ACKNOWLEDGEMENTS

This work was created as part of a project financed by the program of the Minister of Education and Science entitled "Implementation doctorate", number of the agreement concluded between the Cracow University of Technology and the State Treasury – the Minister of Education and Science – DWD/6/0520/2022.

REFERENCES

[1] Czmoch I., Walendzik J. (2022). O korzyściach z wdrożenia MacroBIM+ w trakcie realizacji „międzywydziałowego projektu interdyscyplinarnego BIM", jak i w codziennej praktyce. *Builder*, 26, ISSN 1896-0642.

[2] Kassem M., Succar B. (2017). Macro BIM adoption: Comparative market analysis. *Automation in construction*, 81, 286–299, ISSN: 0926-5805.

[3] Succar B., Kassem M. (2015). Macro-BIM adoption: Conceptual structures. *Automation in construction*, 57, 64–79, ISSN: 0926-5805.

[4] Zima K. (2017), Kalkulacja Kosztów Robót Budowlanych z wykorzystaniem technologii BIM; Kraków: Wydawnictwo PK.

[5] Ministerstwo Rozwoju i Technologii (2020), Cyfryzacja procesu budowlanego w Polsce – Mapa drogowa dla wdrożenia metodyki BIM w zamówieniach publicznych; Ministerstwo Rozwoju i Technologii.

[6] Gajbhiye A. D. (2011). Empirical Study of Macrobim and Conceptual Estimation. (Doctoral dissertation, Texas A & M University).

[7] Khemlani L. (2008). DProfiler: A „Macro" BIM Solution; AECbytes.

[8] Juszczyk M. (2017). On the Neural Networks Applications in BIM-Based Macro Level Cost Analyzes in Construction Projects. *DEStech Transactions on Economics, Business and Management*, ISSN: 2475-8868.

[9] Szczepaniak R., Michalczyk. „MacroBIM – ewaluacja inwestycji budowlanej w jej początkowej fazie" https://www.youtube.com/watch?v=yy0eZNegv7k ;

[10] Zima K. (2021). Modele Macro i Micro BIM. Buduj Z Głową. *Poradnik Kosztorysanta*, 2, 91–97.

[11] Plebankiewicz E., Zima K., Skibniewski M. (2015). Construction cost and time planning using BIM-based applications. *Creative Construction Conference*, 537–545.

[12] Juszczyk M. (2010). Modelowanie kosztów realizacji budynków mieszkalnych z zastosowaniem zespołów sztucznych sieci neuronowych. *Czasopismo Techniczne, Budownictwo*, 176–185.

[13] Harness C. Macro BIM: When global parameters rule everything, TestFit Blog, https://testfit.io/blog/macrobim/, (acess: 14.06.2023).

[14] Juszczyk M. (2017). Studies on the ANN implementation in the macro BIM cost analyzes.

[15] Juszczyk M. (2015). Application of committees of neutral networks for conceptual cost estimation of residential buildings. AIP Conf. Proc., 1648, 600008;

[16] Bryszewski J. M. (2020). Multidisciplinary project of A1 Highway in Poland using BIM process, (Doctoral dissertation, Instytut Dróg i Mostów).

[17] Kurek S. (2023). CPK: BIM w Centralnym Porcie Komunikacyjnym. *Builder*, 76–78, ISSN 1896-0642.

[18] Miecznikowski P. (2017). Inwestor publiczny-dobre bliskie praktyki wdrożenia BIM. *Builder, 21*(5), 42–45, ISSN 1896-0642.

[19] Orlińska-Dejer K. (2023). PROJEKTOWANIE PRZYSZŁOŚCI: Konsylium BIM. *Builder*, 72-73, ISSN 1896-0642.

[20] Przepiórka G. (2023). buildingSMART Polska: openBIM 2023. *Builder*, 68–69, ISSN 1896-0642.

[21] BIM STANDARD PL – Projekt zasad przygotowania i realizacji inwestycji kubaturowych w Polsce zgodny z normą PN-EN ISO 19650 i krajowym prawem budowlanym. Wersja dokumentu nr 2.0, https://www.gov.pl/web/rozwoj-technologia/realizuj-inwestycje-zgodnie-z-metodyka-bim---opracowanie-bim-standard-pl, (dostęp: 17.06.23);

[22] Strategia Polskiej Izby Inżynierów Budownictwa w zakresie wdrażania BIM, miejsca i roli Izby w tym procesie oraz wskazania sposobów realizacji tej strategii. (2019). https://www.piib.org.pl/pliki/kr/Strategia_ds_BIM_24022020.pdf, (dostęp: 17.06.2023);

[23] Janota-Bzowski J. (2021). Mapa drogowa BIM?. Rynek Instalacyjny.

[24] StatSoft (2006). Elektroniczny Podręcznik Statystyki PL, Kraków, WEB: http://www.statsoft.pl/textbook/stathome.html.