DOI: 10.2478/sspjce-2025-0020

A Data Model Design for the Historic Buildings Classification and Management

Andrea Hrubovcakova[1*], Maria Grazianova[1], Peter Mesaros[1], Marcela Spisakova[1]

[1]Technical University of Košice, Slovakia
Faculty of Civil Engineering, Institute of Technology, Economics and Management in Construction
[*]e-mail: andrea.hrubovcakova.2@tuke.sk

Abstract

Documentation of historic buildings, not only in Slovakia, is often chaotic, incomplete, and the information is unclear. Managers of historic buildings and monuments offices preserve historical and building information mainly in text documents, which are fragmented and unstructured. The aim of this article is to propose a consistent data structure in the form of a data model that will enable the systematic classification and management of specific data on the historic buildings. The model also makes it possible to identify which types of data are relevant for different user groups and supports more efficient evidence, analysis, and utilization of information in decision-making processes. The developed data model reflects the interconnection of several entities, such as Building, Architecture, Technical Specifications, Legal Protection, and Economic Data. The proposed approach provides a basis for the digitization and integration of data on historical objects, contributes to improved renovation planning, and offers a framework for more efficient decision-making in the field of cultural heritage protection.

Keywords: data model, historic building, digitization, data classification, decision-making process

1 Introduction

Historical buildings pose a major challenge in terms of their renovation, management, and documentation. With their complex architecture, diversity of materials used, and historical data, they create collections full of information that are difficult to systematically monitor, organize, and continuously update. In Slovakia, the structuring of data on historical buildings is still inadequate, with data often only available in unstructured text documents, where locating specific information is time-consuming and inefficient. In addition, the databases of monument authorities lack a uniform style of documentation, which complicates their interpretation and causes that important information does not always reach experts capable of working with it in a usable form [1,2].

One possible solution to this challenge is the implementation of systematic data management, which creates mechanisms for managing, storing, and controlling data. Data in such an environment must be related to the extraction, preservation, sharing, application and utilization of data. DAMA International defines a six-level model of data management maturity that

categorizes organizations from level 0 (no data capabilities) to level 6, characterized by data optimization and effective use of data resources. This framework allows to assess the current state of data management and identify areas for improvement [3,4].
A data model represents a formalized structure for organizing and interrelating information within a specific domain, enabling consistent storage and analysis of complex datasets. In the built environment, data models are typically derived from heterogeneous sources such as BIM, life cycle assessment data, and digital twin frameworks, requiring semantic alignment and normalization across different information levels. Well-designed data models support interoperability, reuse, and automated analysis, transforming unstructured or semi-structured construction data into analytically usable knowledge [5-9].
A key tool in the process of improving data management is the creation of a suitable data model. A data model is a formal way of describing the structure of data, its relationships, and the rules that determine how information should be captured and interpreted. In the context of historic buildings, a well-designed data model can significantly simplify the management of information about buildings, support their interconnection with existing archival or technical materials, and enable more efficient data exchange between experts. At the same time, it provides a basis for decision-making systems that can assist in renovation planning, building condition monitoring, and risk assessment, thereby contributing to the systematic and sustainable protection of cultural heritage [10-12].

2 Methods

The objective of the scientific research paper is to develop a data model that will be used to classify and structure data on historical buildings. To obtain relevant data as a basis for the model, a theoretical analysis of information on historical buildings will be used in combination with the requirements defined by the Monument Office in Košice. By monitoring and defining these requirements, it is possible to analyze which data needs to be recorded for historical buildings and how it is divided into specific groups. The designed research approach will allow us to verify how it is possible to establish a comprehensive and reusable data model from unstructured data.

To enhance methodological clarity, the process of transforming unstructured textual descriptions into a structured data model was conducted in several steps. First, all textual records were reviewed and segmented into thematic blocks using content analysis (identification data, historical information, material characteristics, technical condition, legal data, etc.). In the next step, recurring information patterns were identified and normalized to eliminate inconsistencies across documents. Each data element was then assigned to a specific entity according to semantic relevance, frequency of occurrence, and analytical usefulness. In cases of ambiguous or inconsistent information, expert consultation with staff from the Monument Office in Košice was used to verify the appropriate classification. This structured approach ensured consistency and reliability of the resulting data model.

2.1 Data sources

When creating a data model, it is essential to have complete and accurate initial information in order to clearly structure the data and transform unstructured text data on historical buildings into a uniform data structure suitable for analytical processing. Even though there are official databases and registers of historical buildings and protected monuments in Slovakia, these data are not in a uniform format, and searching for specific data is often chaotic and complicated.
The main source for data collection were the requirements defined by the Monument Office in Košice, which collects and archives comprehensive information on individual historical buildings. The office has approximately 469 text records on historical buildings in the city center of Košice. However, these documents are unstructured, written by different authors and in a different architectural and artistic style, which makes them inconsistent. It is precisely this diversity that provides an opportunity to summarize and categorize the data, thus creating a more efficient management system and a more transparent data structure.
The starting point for summarizing and classifying the data was the scientific article The use of AI in the digitization of cultural heritage: the potential of LLMs in analyzing textual information on historical buildings. In this research, information about historical buildings was divided into ten categories (Table 1): Identification data, Historical data, Material characteristics, Structural and technical condition, Photographic documentation, Maintenance and monitoring, Legal data, Digital models and records, Research and analysis, and Financial data.

Table 1: Information about historical buildings

Category	Subcategory	Format
Identification data	Name of the building	Dom meštiansky
	Address	Hlavná 90/48, Kosice
	GPS coordinates	NOT RECORDED
	Year of construction	Before year 1500
	Current owner	NOT RECORDED
Historical data	Year of classification	NOT RECORDED
	Historical importance	National cultural monument
	Renovation records	Portal replacement (1972)
		Monument research (1980)
		Central heating (1990)
		Adaptation of the building (1992)
		Entrance to the store (1996)
Material characteristics	Exterior facade material	The street facade is made of exposed brick masonry and stucco decoration with neo-Renaissance elements
		The rear facade is smooth, without any distinctive morphological details
	Roof type	The main wing has a gable roof
		The courtyard wing is covered with a pult roof

	Interior material	Vaults made of stone and brick masonry
		The flat ceilings are wooden beam ceilings with plastered flat soffits
	Other materials	Stone masonry in the basement, gradually supplemented with brick screens
		Cast iron railing of the main staircase
		Rod railings on balconies
Structural and technical condition	Current status	Good
	Critical areas	Inappropriate interventions on the rear façade and some modern modifications have disrupted the historical integrity of the building
	Necessary renovations	Disturbed heritage value - professional research and subsequent restoration of historical elements and structures is recommended
Photographic documentation	Actual photos	NOT RECORDED
	Historical photos	NOT RECORDED
	Floor plans and layouts	NOT RECORDED
Maintenance and monitoring	Maintenance schedule	Replacement of the portal, adaptations, and heating (1972–1996)
	Inspection records	NOT RECORDED
Legal information	Protected areas	Košice Urban Conservation Area (MPR)
	Permits for interventions	Portal replacement (1972)
		Monument research (1980)
		Adaptation of the building (1992)
	Legislative restrictions	Heritage law, the need to preserve the basic architectural structure, artistic and craft elements, and the layout on all floors
Digital models and records	HBIM model	NOT RECORDED
	3D scans	NOT RECORDED
	Digital Plans	NOT RECORDED
Research and analysis	Archaeological research	NOT RECORDED
		Survey research by Mgr. Vlk (PÚK Bratislava) in 1980, which revealed medieval and Renaissance elements of the building
		Block focus unit VI-D (Stavoprojekt Bratislava, 1972)
		Monument survey block 15 A (PÚK Bratislava, 1989)
	Chemical analyses	NOT RECORDED
Financial data	Maintenance costs	NOT RECORDED
	Renovation costs	NOT RECORDED
	Project costs	NOT RECORDED

Based on the results of the article *The use of AI in the digitisation of cultural heritage: the potential of LLMs in analysing textual information on historical buildings*, it was possible to further analyze which data should be available to individual user groups. Access to information proved to be a key factor in defining what data needed to be structured. Different professions (e.g., architects, historians, administrators, the public) require different types of data, which is why a table was created dividing users into four main categories: General, Technical, Art-Historical, and Economic (Table 2).
This classification allows easier understanding of which data is relevant for individual user groups, while also improving the efficiency of the decision-making process when planning the renovation of historic buildings. In addition, it contributes to the protection of sensitive data and compliance with GDPR principles, as individual user groups only have access to information that is necessary for their activities, while the public only has access to general and informational data.

Table 2: Main categories

	General	**Technical**	**Art - historical**	**Economic**
Identification data	X			
Historical data	X		X	
Material characteristics	X	X		
Technical specifications	X	X		
Photo documentation	X		X	
Facility management	X	X		
Legal data	X	X	X	X
Digital models	X	X		
Research		X	X	
Analysis			X	

2.1.1 Data model development

Developing a data model is a key step in the efficient management of information about historical buildings. The aim of the process was to transform unstructured source data into a logical data structure suitable for analytical processing and visualization. The model was based on the concept of a relational data model, in which each table represents a separate entity linked by logical relationships. The model designed in this way allows data on historical objects to be classified, filtered, and linked according to their type and purpose.
Various digital tools were used in the design of the model, enabling more efficient work and faster visualization of results. The basic data structure was created in Microsoft Excel, where an extensive table containing various categories of information was created. This data was classified by type into: basic information about the building, historical data, architectural

characteristics, data on materials used, legal information, location data, digital data, economic information, technical specifications, user data, and information about the category to which the information belongs (Table 3).

Table 3: Data classified by type

	Column	**Type**	**Description**
Buildings	ID_Building	Whole Number (PK)	Building identifier in the database
	Building_Name	Text	Name of the object
	Buiding_Type	Text	Type of building (residential, religious, etc.)
	Year_Built	Whole Number	Year of construction
	Architect_Name	Text	Name of architect or designer
	Curent_Use	Text	Current function of the building
	Owner_Name	Text	Private individual or legal entity
	Manager_Name	Text	Facility manager
	Heritage_Status	Text	Level of monument protection
	Created_At	Date	File creation date
	Updated_At	Date	Date of file modification
Architecture	Arch_ID	Whole Number (PK)	Architectural record identifier
	ID_Building	Whole Number (FK)	Connection to the building
	Style	Text	Architectural style
	Art_Period	Text	Art period
	Requir_Category_ID	Whole Number (FK)	"Art-Historical"
Materials	Material_ID	Whole Number (PK)	Material identifier
	ID_Building	Whole Number (FK)	Connection to the building
	Material_Type	Text	Material type
	Last_Renovation_Year	Whole Number	Year of last renovation
	Requir_Category_ID	Whole Number (FK)	"Technical"
Legal_Protection	Legal_ID	Whole Number (PK)	Legal protection identifier
	ID_Building	Whole Number (FK)	Connection to the building
	Protection_Level	Text	Type of protection (national cultural monument, zone, etc.)
	Protection_Basis	Text	Description of protection
	Legal_Notes	Text	Legal regulations
	Requir_Category_ID	Whole Number (FK)	"Legal"
Location	Location_ID	Whole Number (PK)	Location identifier
	ID_Building	Whole Number (FK)	Connection to the building
	Address	Text	Street, number, city, postal code
	GPS_lat	Decimal Number	Latitude
	GPS_lng	Decimal Number	Longitude
	Cadastre	Text	Region/Municipality/Parcel
	Requir_Category_ID	Whole Number (FK)	"General"
Digit. Data	Data_ID	Whole Number (PK)	Digital record identifier
	ID_Building	Whole Number (FK)	Connection to the building

	Data_Type	Text	Digital tool, records, documentation
	Description	Text	Data description
	File_URL	Text	Reference to documentation
	Capture_Date	Date	Date of completion
	Requir_Category_ID	Whole Number (FK)	"General"
Economic_Data	Economic_ID	Whole Number (PK)	Economic identifier
	ID_Building	Whole Number (FK)	Connection to the building
	Operating_Cost_Est	Decimal Number	Estimated operating costs
	Revenue_Potential	Decimal Number	Income potential
	Funding_Sources	Text	Sources of funding
	Current_Use_Value	Decimal Number	Current utility value
	Requir_Category_ID	Whole Number (FK)	"Economic"
Technical_Spec	TechSpec_ID	Whole Number (PK)	Technical specifications identifier
	ID_Building	Whole Number (FK)	Connection to the building
	Current_Condition	Text	Current condition of the building
	Critical_Areas	Text	Critical areas
	Required_Interventions	Text	Necessary interventions
	Protective_Zones	Text	Protection zones
	Requir_Category_ID	Whole Number (FK)	"Technical"
Users_Roles	User_ID	Whole Number (PK)	User identifier
	User_Name	Text	User name
	Role	Text	Role or access data
	System_Admin	Boolean	Is he a system administrator?
	Requir_Category_ID	Whole Number (FK)	"Admin"
User_Links	Link_ID	Whole Number (PK)	Connection identifier
	ID_Building	Whole Number (FK)	Connection to the building
	User_ID	Whole Number (FK)	Connection to the user
	Responsibility	Text	User access
Requir_Cat	Requir_Category_ID	Whole Number (PK)	Request category identifier
	Category_Type	Text	General / Technical / Economic / Art-Historical / Legal / Admin
	Description	Text	Category description
Requir_Cat_ID	1	General	General identification and location data
	2	Technical	Technical specifications and interventions
	3	Art_Historical	Architectural style and artistic period
	4	Economic	Economic and operational data
	5	Legal	Legal/heritage protection
	6	Admin	Users and access management

Each part of the table contains a set of key data necessary for capturing and managing information about historic buildings. The table also includes the Requir_Category_ID field, which defines the user group for which the information is intended. Each building also has its own identification number (Building_ID), which can be used to efficiently manage data on individual buildings and ensure that all information is linked to a specific building. Building_ID

will be a specific identifier consisting of 4 to 6 characters, which will be assigned to each building in order to ensure its unique and consistent identification throughout the database.
The Type section defines the type of value (text, integer, decimal number, date), and the Description column describes what specific type of data should be found in a given category and what information it should contain. A whole number PK (Primary Key) represents a unique identifier for a record in a table, while a whole number FK (Foreign Key) is a numerical value that references the primary key of another table, creating a relationship between individual data items. The table structure created in this way allows all data on historical buildings to be clearly categorized, analyzed, and continuously supplemented. At the same time, it provides a solid foundation for creating a data model that visually illustrates the connections between individual data categories and supports their further analytical interpretation.

3 Results and discussion

To achieve the research objective and design a data model for historical buildings, Microsoft Power BI software was used. This tool enables the automatic generation of relational models from prepared data files and provides a clear visualization of the connections between individual entities.
The input data was originally processed in Microsoft Excel, where an extensive table with summary information on historical buildings was located. This data had to be transformed into smaller and clearer tables, each representing a separate entity. This division resulted in twelve entities containing specific groups of information. The prepared tables with corresponding data were linked using unique identificators, which allows a logical and unique connection of all data.
Thanks to this connection, it is possible to obtain all relevant information about each building – about its architecture, materials used, technical parameters, economic indicators, and legal protection.
In designing the relational model, four basic types of relationships between entities were used:

- ***1 : n (one-to-many)*** – the most commonly used type of relationship, which allows one entity to have multiple records on another. This type of link ensures that a single building can have multiple records of materials used, technical parameters, or legal documents.
- ***n : 1 (many-to-one)*** – the opposite direction of the connection, used mainly for reference tables. This type of link allows information to be classified according to user categories or types of requests; for example, architectural data belongs to the Art-Historical category.
- ***n : n (many-to-many)*** – used in cases where a single record can be linked to multiple entities on both sides. This relationship allows one building to have multiple users (e.g., administrator, architect, owner) and, at the same time, one user can be assigned to multiple buildings.
- ***1 : 1 (one-to-one)*** – this type of link is used only occasionally in the model. This is used where it is necessary to directly link supplementary data without duplication. In the data model, this link is found when linking a building to a location, because each building has exactly one location.

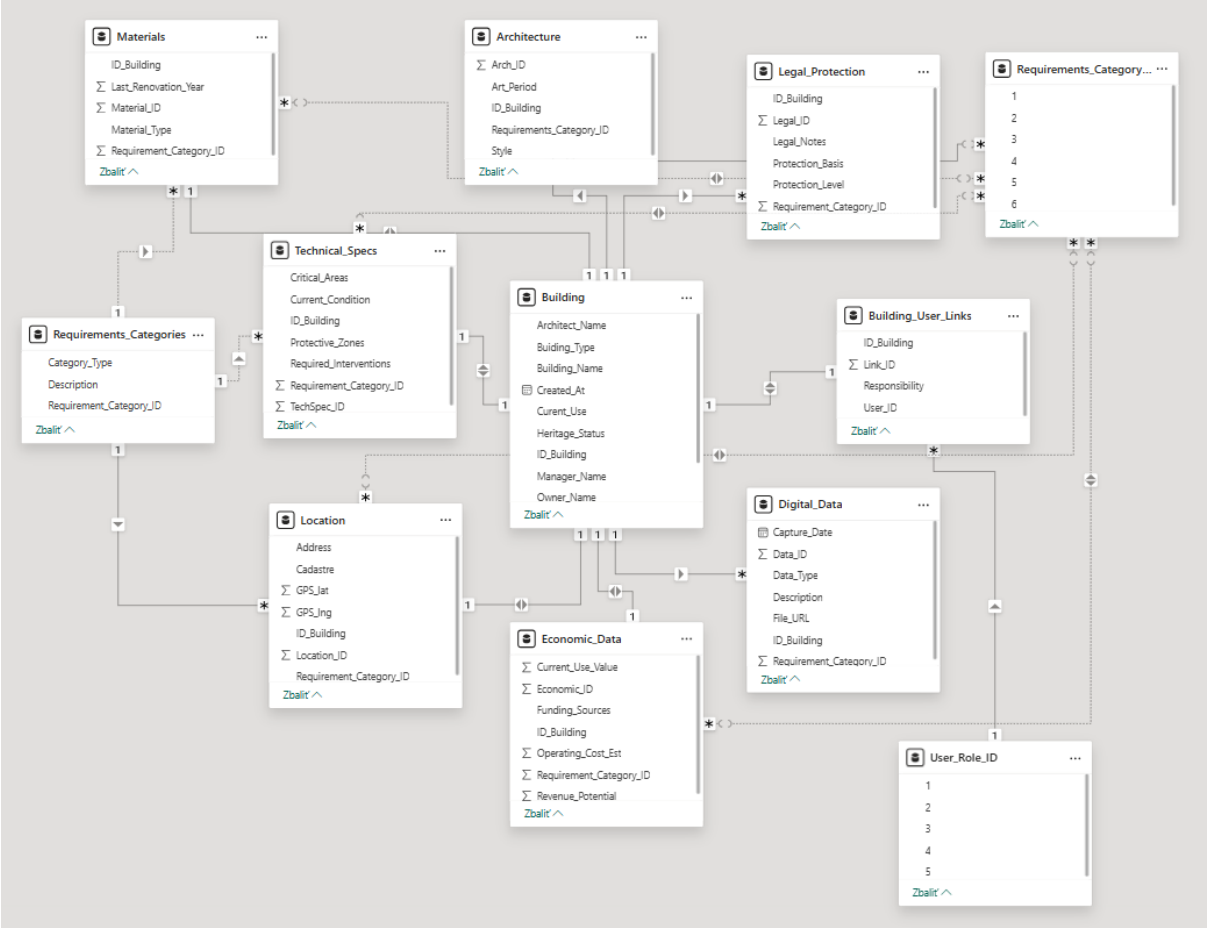

Figure 1: Data model

As can be seen in the figure (Figure 1), the data model is designed centrally, with the main Building table at its center, which is linked to all other entities through the ID_Building identifier. This approach ensures a clear and logical structure, thanks to which data can be easily expanded with new entities without interfering with the existing model architecture.

The use of 1:n and n:n connections creates a flexible system that allows to manage multiple records assigned to a single building and supports multi-level user access management. This solution is crucial for information protection and compliance with GDPR principles, as each user group has access only to relevant data.

The proposed data model represents a clear, uniform, and systematic method of recording information about the historic buildings. It supports analytical data processing, enables visualization in Power BI, and serves as a basis for digitizing cultural heritage and linking multiple databases. The model is scalable, so it can be expanded with new data types or entities while adapting to the needs of different user groups, from monument authorities, architects, and historians to the general public.

3.1 Comparison with international data standards

Although the proposed relational model was developed based on the specific needs of the Monument Office in Košice, it aligns with several principles found in internationally recognized standards for cultural heritage documentation.

CIDOC Conceptual Reference Model is an ontology used primarily in museums and heritage collections to describe events, temporal relations, provenance, and object histories. The presented model shares a similar conceptual approach in its separation of identification, historical, material, legal, and digital information. However, CIDOC CRM is designed as an ontological framework with semantic relationships, whereas the proposed model follows a relational database structure optimized for practical data management and analytical use within heritage buildings administration. The model can be extended with semantic layers if interoperability with CIDOC-based systems becomes necessary.

HBIM focuses on the geometric and parametric representation of historic structures in a 3D environment. While HBIM integrates geometrical, material, and diagnostic data, it typically requires advanced BIM platforms. The proposed model is complementary: it offers a structured tabular repository of descriptive, legal, technical, and economic information. It can serve as a metadata source for HBIM applications and supports integration with digital models by including dedicated entities for Digital Data, 3D scans, and HBIM records.

Overall, the relational model is compatible with these standards conceptually, while remaining lightweight and practical for administrative and analytical workflows in heritage management. It can act as a foundation for future integration with both semantic (CIDOC CRM) and geometric (HBIM) approaches.

4 Conclusion

The aim of the research was to create a uniform data model for evidence and classification of data on historical buildings. Based on available source data, it was possible to identify weaknesses in the management of information on historical buildings in Slovakia and transform unstructured data into a clear and logically organized model. The information provided mainly by the Monument Office in Košice made it possible to define which data are key for the documentation of historical buildings, for which user groups they are most important, and how this data should be categorized.

Using the Microsoft Power BI desktop platform, a central data model consisting of 12 interconnected entities was created, resulting in a uniform structure for easier management, searching, and updating of information. The interconnection of these entities provides a comprehensive view of the historical building from a technical, legal, architectural, and economic perspective. The model is flexible, scalable, and provides a suitable basis for the digitization of historical buildings, as well as for more efficient decision-making processes for various types of users, especially in planning building renovations.

The research also showed that the quality of the resulting model depends on the availability and accuracy of the input data. Some source information was incomplete, which required manual updating and validation of the data. Another challenge for the future is the integration of the model into a real system, which requires the standardization of data formats and coordination between several institutions.

The developed data model represents a fundamental pillar in the digitization of historic building records and the efficient management of historic buildings. Future research should focus on expanding the model to include a time dimension for tracking changes and renovations over time, as well as linking the data model to a digital or web platform that will enable interactive visualization and easier access to data for different user groups.

Acknowledgments

This work was supported by the Slovak Research and Development Agency under the contract no. APVV-22-0576 “Research of digital technologies and building information modelling tools for designing and evaluating the sustainability parameters of building structures in the context of decarbonization and circular construction”. This paper presents partial research results of project KEGA 009TUKE-4/2022 “An interactive tool for designing a safe construction site in an immersive virtual reality”.

References

[1] An official website of the Monument Office of the Slovak Republic. (accessed 17 November 2025) https://www.pamiatky.sk/

[2] An official website of the Monument Office of the Slovak Republic - Monument Information System (PAMIS). (accessed 17 November 2025) https://www.pamiatky.sk/projekty/pamiatkovy-informacny-system-pamis

[3] Caballero, I,. Gualo, F., Rodriguez, M. & Piattini, M. (2023). A Maturity Model for Data Governance, Data Quality Management, and Data Management. *Cloud computing, big data & emerging topics. 1828*, 157-170. 10.1007/978-3-031-40942-4_12

[4] Bradley, Ch. (2024). *DAMA-DMBOK: data management body of knowledge*. https://bibliotek.nav.no/record/5716?ln=no

[5] Gan, V. J. L. (2022). BIM-based graph data model for automatic generative design of modular buildings. *Automation in Construction. 134.* https://doi.org/10.1016/j.autcon.2021.104062

[6] Rock, M., Passer, A. & Allacker. K. (2024). SLiCE: An open building data model for scalable high-definition life cycle engineering, dynamic impact assessment, and systematic hotspot analysis. *Sustainable Production and Consumption. 45.* 450-463.https://doi.org/10.1016/j.spc.2024.01.005

[7] Shi, J., Pan, Z., Jiang, L. &, Zhai, X. (2023). An ontology-based methodology to establish city information model of digital twin city by merging BIM, GIS and IoT. *Advanced Engineering Informatics. 57.* https://doi.org/10.1016/j.aei.2023.102114

[8] Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction. 114.* https://doi.org/10.1016/j.autcon.2020.103179

[9] Qi, B. & Costin, A. (2023). BIM and Ontology-Based DfMA Framework for Prefabricated Component. *Buildings. 13(2).* https://doi.org/10.3390/buildings13020394

[10] Zhang, J. P., Li, H., Zhang, X. P. Chen, M., Dai, Z. Y. & Zhu, M. (2019). Determination of the data Model for heterogeneous data Processing Based on Cost Estimation. *Artificial intelligence and algorithms in intelligent systems. 764,* 374-383. 10.1007/978-3-319-91189-2_37

[11] Wang, L. (2017). Heterogeneous data and Big Data Analytics. *Automatic Control and Information Sciences. 3,* 8-15. https://doi.org/10.12691/ACIS-3-1-3

[12] Castro, E., Cuadra, D. & Velasco, M. (2010). From XML to relational models**.** *Informatica: An International Journal of Computing and Informatics. 4,* 505-519. 10.15388/Informatica.2010.302