A framework for using heterogenous data from digital building permits in urban mining

Marcel Heiß[1] and Timo Bittner[2]

[1] Institute of Numerical Methods and Informatics in Civil Engineering, Technical University of Darmstadt, Franziska-Braun-Str. 7, 64287 Darmstadt, Germany

[2] City of Darmstadt, Darmstadt, Germany

E-Mails: heiss@iib.tu-darmstadt.de

Abstract: Previous approaches to material inventories of buildings require complex surveys, complicated estimations or BIM models. The concept presented explores the use of the digital transformation of building permit in Germany to create a data collection layer for urban mining. By analysing certain aspects of building permits in Germany and using digital methods, relevant data can be extracted. Different forms of future building permits and their characteristics are defined and analysed in order to select specific digital methods such as artificial intelligence. A comparison of the different forms shows that the necessary data can be extracted from simple digitised documents and further developments of the building permit offering more potential in terms of reliability and accuracy.

Keywords: Urban Mining, Digital Building Permitting (DBP), Building Information Modeling (BIM)

Erschienen in Tagungsband 35. Forum Bauinformatik 2024, Hamburg, Deutschland, DOI: 10.15480/882.13528
© 2024 Das Copyright für diesen Beitrag liegt bei den Autoren. Verwendung erlaubt unter Creative Commons Lizenz Namensnennung 4.0 International.

1 Introduction

The built environment is society's largest resource stock and a major contributor to energy and resource demand. By reducing its high consumption, the construction sector offers considerable potential to achieve climate neutrality in Europe by 2050. Here, the recovery and reuse of building resources and components as well as their integration into a circular economy provides a significant opportunity [1]. The introduction of such concepts into the planning process of urban districts requires the coordination of material flows. This entails comprehensive knowledge of the materials in the built environment as an initial information basis for urban mining concepts. However, corresponding data is rarely available at the district level. Existing approaches require complex building surveys, costly estimations or BIM models to solve this problem. At the same time, the digital transformation is also gaining ground in many areas in Germany and machine-readable data is increasingly being generated. In the public sector, the Online Access Act facilitates digitising processes by offering administrative services electronically using digital forms and file exchange platforms. Likewise, the transition from paper-based handovers to digital exchanges also opens up opportunities to prevent data and information disruptions and thus enables development of new technologies. An example of this is the implementation of digital building permits and the information about the built environment associated with it. The approach presented in this work explores the use of the digital transformation of building permits in Germany to create a data collection layer for urban mining. For this purpose, relevant documents, which often implicitly contain information about the building resource inventory, are analysed. Current developments for the digital transformation of building permits and material inventories of buildings are presented. Based on this, different forms of potential future digital building permits are defined incorporated into a framework for extracting and evaluating the necessary data. The framework features different pipelines for processing non-machine-readable documents with methods from the field of computer vision and artificial intelligence (AI). The integration of future machine-readable formats is then explored to collect resource data without the need for complex information extraction methods. Additionally, the use of a BIM-based building permit is examined. The results of building material inventories are evaluated regarding their use for urban mining.

2 Related Work and Analysis

2.1 Digital transformation of building permit

The use case of Digital Building Permitting (DBP) is of growing interest with the level of development varying widely in different countries around the world. It can be summarised into four distinct levels of permitting based on different degrees of automation and integration [2]. These include traditional paper-based permitting practices, electronic 2D digital data-based permitting, automated digital BIM-based permitting, and a fully automated Geo Information System (GIS)-integrated level. One of the drivers for this development are the increasing requirements for newer buildings, resulting in more extensive inspections as well as increases in workload and time to approval. The primary aim of DBP is to automate processes, thereby reducing the manual workload of the approval procedure. The electronic, 2D digital data-based permit provides a central platform for communication and data exchange, but the PDF documents submitted are often analysed manually. For this reason, most research, such as the EuroSDR GeoBIM project, suggests a high-level harmonised workflow envisaging the use of GeoBIM information to automate the building permit process [3]. The EU-funded research project BRISE [4] is examining the use of BIM and AI, particularly Natural Language Processing (NLP), in the building permit process in Vienna, Austria. In Germany electronic, 2D digital data-based permits are currently being introduced in several projects at the state level, e.g. [5]. Further research [6] includes a concept for the integration of BIM in the building permit procedure. However, the current situation in practice is different. Not all building projects are planned using BIM and are therefore not eligible for BIM-based permitting without additional work. Processes that are not BIM-based but employ digital tools and methods are therefore still valuable for handling of building permit data. There are several examples for the value of reusing and analysing this data, such as study of land development [7]. This reflects the importance and wide range of possibilities of accessible information derived from DBP.

2.2 Material inventory of buildings

One of the biggest challenges for society is containing the climate crisis. In order to reduce the environmental impact of the construction industry, the introduction of a circular economy offers considerable potential for savings [1]. The material inventory of individual buildings and their GIS-based representation at city level through material cadastres is necessary for the development of a knowledge-based decision-making platform. However, there is still a lack of tools that allow automated access to the data to create such inventories in form of so-called building material passports (BMP). There are several approaches that use BIM models for initial information. The research project BIMaterial [8] developed a methodology for the semi-automatic creation of a BMP by linking BIM and eco-databases. The research project BIMstocks [9] aimed to digitally record the material composition by mapping on-site recordings such as laser scans, surveys and photo documentation onto BIM models. These approaches are only applicable to buildings with a BIM model, which is not the case for most existing buildings. While BIM models tend to contain geometric information for specific components as well as superficial material names, such a database provides more in-depth properties. To improve the reliability of the process and avoid blind mapping, additional surveys and recordings are required, which need to be automatically interpreted to avoid manual intervention. One approach is using a machine learning algorithm for automated material recognition [9]. Other possibilities include automatic classification of common building materials from 3D terrestrial laser scan data [10], which is limited because only the visible layer can be estimated. On-site recording could also include laser scanning, which allows the derivation of point clouds and automated reconstruction of BIM models [11]. However, this is very time-consuming and costly, and therefore not a tool for area-wide digital capture of urban mining data. Another option is to analyse existing construction documents. While the automated analysis of plans, so called Plan2BIM approaches like CubiCasa5K [12], and further automated BIM reconstruction are well researched, the extraction of additional material data is often out of scope. When dealing with heterogeneous construction documents, the use of NLP has been proven in the application of document classification and organisation [10] and is increasingly finding its way into the processing of unstructured text data for cost estimation [13].

2.3 Building permit forms

In many countries, building permit information is still handled in paper-based formats or, at best, in PDF documents. Moreover, in practice, the data involved is very heterogeneous and often relatively informal. Previous research in DBP has focused strongly on the use of BIM and its Industry Foundation Classes (IFC) format in digital permits while their use is not sufficiently established. Other specific data exchange formats are rarely considered in the context of DBP. The suggested approach aims to close the research gap between digital document-based and BIM-based by introducing an additional variant of DBP. A framework for urban mining is developed using different forms of DBP, leveraging them as a data source for digital urban twins [14] and corresponding knowledge-based decision-making platforms. The concept is based on a step-by-step model for the digital transformation of the building permit. The lowest level is the analogue, paper-based approach, which is still the most widely used in Germany and includes paper forms, documents and 2D plans. This is being replaced by the digital building permit, which often includes the same information sources in digital form submitted via an online platform, but mostly as non-machine-readable PDF documents. Little machine-readable data is captured by web forms. The intermediate step, the so-called Digital+ permit, represents a shift as most of the data is captured via machine-readable formats and files that are well-established and used in practice. The highest level of DBP+ focuses on the data provided by a BIM model in IFC format. The BIM model can replace other previous documents such as plans, as the information contained in the plans can be found in the model. The analysis of other construction documents and further semantic enrichment of BIM models is rarely researched. The work presented focuses on digital document-based, digital data-based and digital BIM-based versions as well as the enhancement of information using a developed "PerDoc2Data" framework.

3 Methodology

3.1 Material inventory processes and elements

The approach taken for the comprehensive material inventory of buildings based on DBP is a bottom-up approach. In contrast to the top-down concept commonly used, it describes the existing stock in detail and extrapolates the total quantity from these partial quantities. Using as much data as possible, the aim is to quantify resources directly and make them locally addressable. The document-, data- and BIM-based building permits introduced are starting points, as shown in Figure 1. Different steps are necessary to get the information for a material inventory for each input separately.

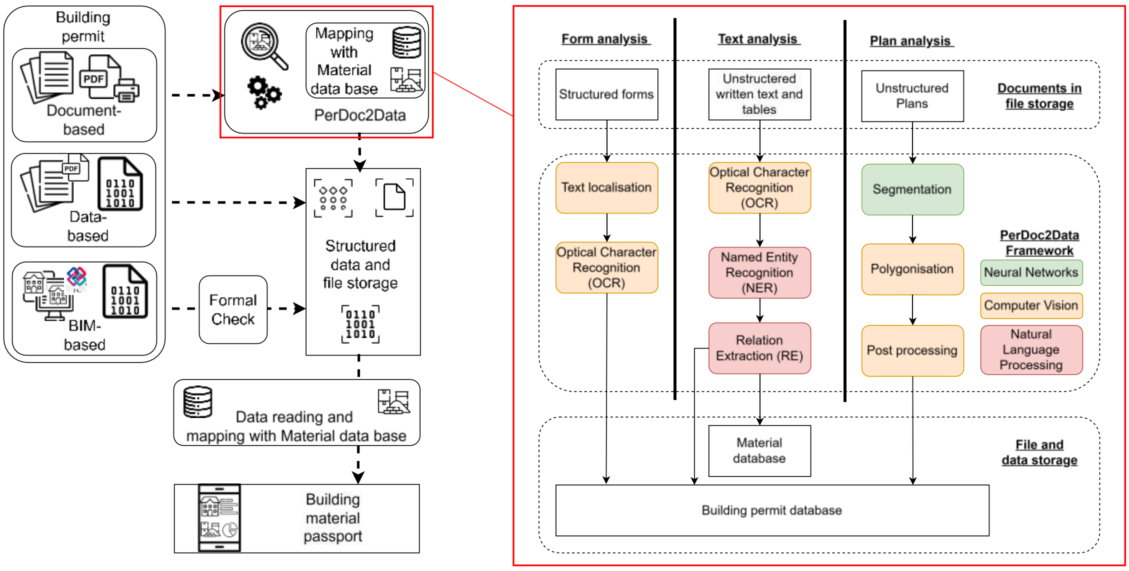

Figure 1: Material inventory processes and PerDoc2Data Framework

Similar to existing approaches, a necessary additional data source is a material database to enrich information from building permits. The structure of the database used here is based on the IFC material description, distinguishing between layered components, described by material layer sets, and non-layered components, described by material constituent sets. Layered components consist of one or more layers of a given thickness. Unlike IFC, the material database allows more than one material to be associated with a material layer. Non-layered components with different subcomponents consist of a collection of individual materials. In order to prepare the document-based permit for automated processing and to match it with the material database, a PerDoc2Data algorithm is developed to extract structured, machine-readable data. Information extracted from PerDoc2Data is stored alongside data generated during a data-based permit, e.g., using web forms. The result is a collection of heterogeneous documents and a structured, machine-readable database for further processing. A reading and mapping process combine data from different sources such as the materials database, permit database and files. Building material passports are then generated for urban mining efforts to visualise the material inventory of a building.

3.2 Implementation

The development of the process pipelines was preceded by an analysis of the documents related to the current building permit in order to identify possible sources of information, revealing heterogeneous, mostly unstructured documents such as forms, descriptions and certificates depending on the building classification. Documents regarding thermal insulation, sound insulation and additional building description are particularly useful sources of data to identify the materials used. Plans are often the only source of data for component quantities. For document-based permits and the associated PerDoc2Data approach, the degree of structuring and the types of elements used within the documents are of crucial importance. Identified sources are therefore categorised according to these criteria in order to select suitable methods for data extraction. Forms with specific queries are considered structured. Unstructured forms are further separated according to the contained elements defined by written text, tables and plans. To extract data from these, three pipelines are developed shown in Figure 1. Structured forms are analysed using a text localization process and optical character recognition (OCR). The text localisation process involves scaling the input form and dividing it into subparts in order to locate specific information in a particular position. The subparts are read using OCR and transferred to a building permit database. Written text analysis starts with reading using OCR. To analyse this text, a Named Entity Recognition model is created based on a review of building permit documents with the aim of creating a knowledge base for urban mining. Through patterns and keywords, additional data such as component thicknesses, number of levels, level height, component types and construction method is recognised. Relation extraction then associates component data with material layers and thicknesses. Results are shown in Figure 2 in the middle and on the right. After selecting a material layer from the material database based on the detected component type, material and thickness value, the layer set associated with this material is linked to the component within the building permit database. For simplicity, the component types roof, external walls, internal walls, slabs and foundation are extracted. Floor plans are analysed within the plan analysis pipeline to extract quantities for component types. The CubiCasa5K neural network [12] is first used in a segmentation process to detect walls on the floor plan. In a subsequent polygonization process, the detected wall pixels are combined into polygons. A final post-processing uses computer vision methods to differentiate between the component types slab, external and internal wall and roof. Results are shown in Figure 2 on the left. In conjunction with extracted metadata such as average level height area quantities are calculated. Fusing this quantity data and previous extracted material data allows for the determination of material inventory. Within the data-based permit, the same information on materials and component type quantities is obtained, but from different data sources. Energy performance certificates and formats for the digital provision of energy simulation models required for thermal insulation certificates play an important role. The use of these standardised, machine-readable formats such as Green Building XML and XML-based energy certificates and their combination with a material database allows the material inventory to be determined without the use of a complex framework. Compared to the other forms, the BIM-based permit uses the BIM model in IFC format as its only data source. In order to support the material mapping and data combination process and to avoid missing information, modelling guidelines and their verification are necessary. BIM models are checked for the specification of quantity data and compatibility with the material database in order to fuse with material data.

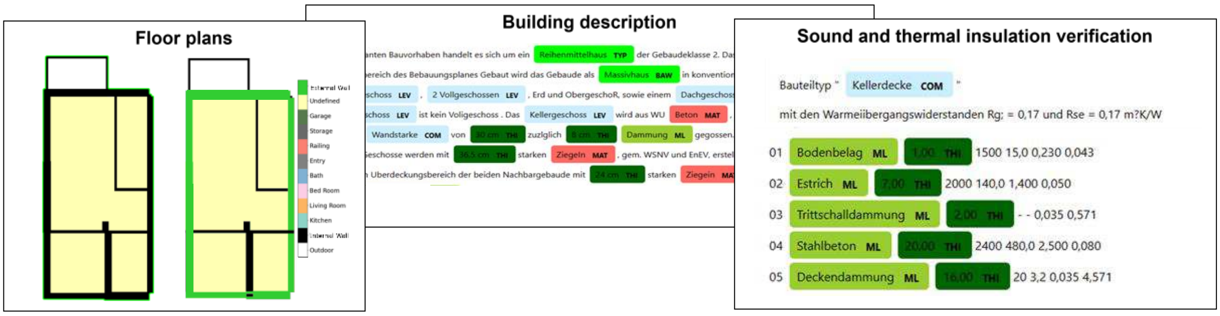

Figure 2: Results of PerDoc2Data framework

4 Results and Discussion

In order to compare the results of the digital and BIM-based permits, the characteristic values of the building material inventory of an example building are determined using both methods. The results, excluding window and door materials, are shown in table 1.

Table 1: Results for the material inventory of an example building

	Digital		BIM		Digital		BIM
	Total [m²]	[%]	Total [m²]		Total [t]	[%]	Total [t]
Gross floor area	279,7	15,1	243,3	Total mass	446,5	16,0	384,83
Net floor area	230,5	14,1	201,4	Concrete	231,6	3,9	222,8
Exterior wall area	196	43,8	141,7	Metal	87,8	3,9	84,5
Interior wall area	260,4	18,2	223,2	Insulation	21,44	21,2	17,69
Ceiling area	279,7	24,6	223,8	Wood	11,1	-12,1	12,63
Roof area	279,7	15,1	243	Gypsum	20,71	49,0	13,9

A material is assigned to each component so that all masses are documented in both approaches. The aforementioned component types are considered and a standard floor height is used. The results of the Digital+BIM approach were verified through manual investigations and are consistent with them, so they are used as a baseline. The Digital+ approach was not compared with the other two concepts due to unavailable data, although its accuracy and reliability should be comparable due to the data formats used. The Digital+BIM approach results in a very accurate material inventory with deviations only occurring in conjunction with unrecorded changes during construction. The uniform deviation of area values obtained with CubiCasa5K (Digital) in one direction can be attributed to an inaccurate scaling factor when converting pixels to actual dimensions. Wall areas also show a higher deviation due to the omitted window and door openings. Consequently, deducting these openings improves the results. A similar, but less significant effect occurs for ceiling areas over openings caused by stairs and lifts. The 16% overestimation of the total amount of material can be attributed to an overestimation regarding components. A closer look at the material composition reveals some anomalies with a considerable impact on the total amount of material because of differences in bulk density. This is caused by specifications being standardised across component types, meaning that only one composition is assigned to each type, even if the actual occurrence diverges. Further improvements can be achieved by integrating the results for windows and doors from CubiCasa5k and incorporating them in the post-processing presented, allowing component quantities to be determined more accurately and door and window materials to be considered. Additional segregation of component types in terms of composition is limited by the level of detail in the approval plans, as further information on materials and layers is often not shown in the plans. More detailed plans would provide an opportunity for a more precise mapping between the aforementioned certificates and the plans, for example using machine learning methods. The use of machine learning can also add significant value when mapping between IFC or extracted data and a materials database.

5 Conclusion

Relevant data available in selected building permit documents provides an initial, sufficiently accurate information base for urban mining and other uses such as energy simulations. The data extracted from non-machine-readable documents is subject to a degree of uncertainty and inaccuracy, but may still be suitable for use in urban mining. For smaller and older buildings, the depth of information can be limited due to the lower requirements of previous permits, making additional data sources necessary. For larger and newer buildings, additional documentation eliminates the need for these extra sources, but the actual form might make it difficult to extract reliable data. However, certainty and accuracy will increase along with the proportion of machine-readable data, further facilitated by the possible integration of existing exchange formats and standards into the digital permit. While the reuse of materials and resources can already be assessed, the reuse of specific components requires a higher level of accuracy and information depth, often only achievable with BIM models. Here, the presented approach shows the potential of machine-readable formats and digital building permits to reliably capture relevant information for purposes such as urban mining.

Acknowledgements

This work was funded by the German Federal Ministry of Education and Research (BMBF) within the framework of the project RessStadtQuartier2 - Urban material flow management: tools for the resource-efficient development of urban districts, funding code 033W109AN-EN.

References

[1] R. Vogt *et al.*, “Ermittlung der Klimaschutzpotentiale in der Kreislaufwirtschaft für Deutschland und die EU,” Umweltbundesamt, 2023.

[2] K. Shahi, B. Y. McCabe, and A. Shahi, “Framework for Automated Model-Based e-Permitting System for Municipal Jurisdictions,” *J. Manag. Eng.*, vol. 35, no. 6, Nov. 2019.

[3] F. Noardo *et al.*, “EuroSDR GeoBIM project - Integrated Workflow using GeoBIM information for building permit process,” Jul. 2020, doi: 10.5281/ZENODO.3948493.

[4] (BRISE) Building Regulations Information for Submission Envolvement, Accessed: Apr. 10, 2024. [Online]. Available: https://digitales.wien.gv.at/projekt/brisevienna/

[5] Ministry of Regional Identity, Local Government, Building and Digitalization of the State of North Rhine-Westphalia, “BAUPORTAL.NRW.” Accessed: Apr. 10, 2024. [Online]. Available: https://bauportal-assistent-prod.kdvz-frechen.de/intelliform/forms/vba/vba/demo_64/index

[6] J. Tulke, M. Theiler, M. König, and A. Vonthron, “Konzept für die nahtlose Integration von Building Information Modeling (BIM) in das behördliche Bauantragsverfahren,” 2021.

[7] M. Shakro, “Tracking neighborhood development and behavioral trends with building permits in Austin, Texas,” *J. Maps*, vol. 9, no. 2, pp. 189–197, Jun. 2013.

[8] I. Kovacic *et al.*, “Prozess-Design für den „Building Information Modeling“ (BIM) basierten, materiellen Gebäudepass,” 2019.

[9] M. Honic, I. Kovacic, and D. Breitfuss, “Digital Urban Mining Platform: Assessing the material composition of building stocks through coupling of BIM to GIS,” 2023.

[10] L. Yuan, J. Guo, and Q. Wang, “Automatic classification of common building materials from 3D terrestrial laser scan data,” *Autom. Constr.*, vol. 110, p. 103017, Feb. 2020.

[11] Y. Perez-Perez, M. Golparvar-Fard, and K. El-Rayes, “Scan2BIM-NET: Deep Learning Method for Segmentation of Point Clouds for Scan-to-BIM,” *J. Constr. Eng. Manag.*, vol. 147, no. 9, p. 04021107, Sep. 2021, doi: 10.1061/(ASCE)CO.1943-7862.0002132.

[12] A. Kalervo, J. Ylioinas, M. Häikiö, A. Karhu, and J. Kannala, “CubiCasa5K: A Dataset and an Improved Multi-Task Model for Floorplan Image Analysis,” 21st SCIA 2019, Sweden, 2019,

[13] C. Gatto, A. Farina, C. Mirarchi, and A. Pavan, “Development of a framework for processing unstructured text dataset through NLP in cost estimation AEC sector,” EC3 2023 ,2023.

[14] N. Schubbe, “Urbane Digitale Zwillinge als Baukastensystem: Ein Konzept aus dem Projekt Connected Urban Twins (CUT),” *Zfv – Z. Für Geodäsie Geoinformation Landmanagement*, no. 1/2023, pp. 14–23, 2023, doi: 10.12902/zfv-0417-2022.