A conceptual system architecture for enriching Digital Twins with material performance data using symbolic and sub-symbolic Artificial Intelligence

Julia Kaltenegger[1*], Ekaterina Petrova[1], André Borrmann[2], Pieter Pauwels[1]

[1] Eindhoven University of Technology, Department of the Built Environment, Den Dolech 2 Eindhoven, Netherlands

[2] Chair of Computational Modelling and Simulation, Technical University of Munich, Germany

The built environment accounts for 50% of raw material consumption and 35% of all waste [1]. Optimising material efficiency and enabling reuse depends on the availability of sufficient and reliable information relevant to identifying reusable building materials and components and direct circular strategies. These strategies are guided by performance assessments such as Life Cycle Assessment (LCA) and representations such as Digital Product Passport (DPP) and rely on materials classifications, property definitions and impact categories. Currently, material (meta)data is hardly openly available, incomplete, scattered across systems or not recorded at all. It is crucial to provide material data that is machine-readable, findable, accessible, interoperable, and reusable (FAIR) [2] to enhance performance assessments and promote the prediction of missing material insights. The complex nature of material information can be captured using Material Information Modelling (MIM) and the corresponding data models [3]. MIM enables material classifications and property definitions at whole building or object-oriented performance simulation levels. BIM [4] represents the building systems; however, neglects the complexity of the material information [5]. Standardised and semantically uniform data models are needed to facilitate seamless data exchange between applications utilising materials properties and definitions. In that relation, recent advancements in Digital Twins (DTs) have shown the potential of integrating various static and dynamic data sources across domains. In this context, AI technologies hold significant potential for addressing challenges related to data modelling, prediction, enrichment and integration guided by domain knowledge. Integrating sub-symbolic (machine learning, data mining, pattern recognition, and probabilistic models) and symbolic (semantic data modelling, formal logic, automated reasoning, ontologies) AI approaches can help achieve data sensitivity and explainability that aligns with domain knowledge representation [6].

This work presents a conceptual system architecture of an AI-enabled framework to enrich DTs with material performance data and enable building material-related performance assessments. The aim is to overcome the lack of reliable material information. As such, the system architecture addresses three key elements: (I) the data layer that structures knowledge, (II) the data population module that predicts and infers (missing) material data, and (III) the integration layer to semantically enrich DTs with material information. More specifically, that includes:

- **(I)** A multimodal material information knowledge graph that structures heterogeneous, federated, openly available and FAIR materials data using multi-scalar data models. Openly available databases such as 3DBAG, ECOPlatform and SLiCE will be accessed via open API and interpreted using ontologies such as BOT, BEO, and BMP. Further, sub-symbolic data processing techniques will be proposed to account for missing data and align entities from heterogeneous data sources according to the knowledge graph.
- (II) The data population module predicts material properties, indicators and performance indicators in the context of new and existing buildings. Robust machine learning methods and semantic graph inference techniques are used, leveraging historical data and domain

knowledge. This model relies on the structured information from the knowledge graph to enrich the predictions.

(III) The data integration layer envisions the automated semantic enrichment and matching of the material information from the knowledge base (I) and the predicted data (II) with the DT. Semantic similarity matching using Natural Language Processing (NLP) techniques and determining semantic closeness in the graph representations of the DTs and material properties are a included as a part of the framework.

[1] The Swedish National Board of Housing, Building and Planning. (2023). Available at: https://www.boverket.se/sv/byggande/hallbart-byggande-och-forvaltning/miljoindikatorer---aktuell-status/vaxthusgaser/

[2] Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). *Comment:* The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 1–9. https://doi.org/10.1038/sdata.2016.18

[3] Kaltenegger, J., Frandsen, K.M. & Petrova, E. (2024a). A data management perspective on building material classification. *Journal of Building Engineering.* Accepted

[4] Borrmann, A., König, M., Koch, C., & Beetz, J. (2018). *Building Information Modeling: Technology foundations and industry practice* (1st ed.). Springer. https://doi.org/10.1007/978-3-319-92862-3

[5] O'Donnell, J., Maile, T., Rose, C., Mrazović, N., Morrissey, E., Regnier, C., Parrish, K., & Bazjanac, V. (2013). *Transforming BIM to BEM: Generation of Building Geometry for the NASA Ames Sustainability Base BIM* (LBNL-6033E). Ernest Orlando Lawrence Berkeley National Laboratory. https://simulationresearch.lbl.gov/sites/all/files/lbnl-6033e.pdf

[6] Hoehndorf, R. & Queralt-Rosinach, N. (2017). Data Science and symbolic AI: Synergies, challenges and opportunities. *Data Science, 1,* 27–38