43rd International CIB W78 Conference
IT in Construction
New Delhi, India

Identifier Architectures for Lifecycle Traceability in Knowledge Graph-Based Construction Information Management

Peter Johansson[1,*], Khadija Foroogh[1], Reshmy Sasi Kumar[1], Rahel Kebede[1], Annika Moscati[1]

[1] Jönköping University, Jönköping, Sweden
[*] *Corresponding email: peter.johansson@ju.se*

ABSTRACT

The digital transformation of the construction industry increasingly relies on semantically linked data to support lifecycle information management. Emerging initiatives such as Digital Product Passports and the "Golden Thread of Information" highlight the need for consistent and traceable information across the lifecycle of built assets. However, current approaches do not provide persistent, instance-level identifier architectures capable of linking object representations across lifecycle stages.

This paper investigates how identifier architectures can support lifecycle-consistent traceability in knowledge graph-based construction information systems. The study is grounded in Systems Engineering lifecycle logic and structured through the Semantic Modelling and Linking framework, distinguishing between Planned Functional (PF), Planned Technical (PT), Realized Technical (RT), and Realized Functional (RF) states. A qualitative, design-oriented approach informed by Design Science Research is applied.

The analysis indicates that separating identifier roles across lifecycle states improves traceability between requirements, technical solutions, realised objects, and operational performance. A key finding is the trade-off between semantic expressiveness and identifier persistence. The proposed architecture addresses this by anchoring each lifecycle role to a persistent identifier, while semantic meaning is represented through associated properties.

1. INTRODUCTION

The construction industry is rapidly adopting Building Information Modelling (BIM), digital twins, and data-driven lifecycle management to improve information continuity, automation, and circular economy practices (Zhu and Wu, 2022; Sacks et al., 2025; König et al., 2018). At the same time, initiatives such as the "Golden Thread of Information" emphasise consistent, accessible, and traceable information throughout the lifecycle of built assets (Hackitt, 2018; UK Government, 2024).

Yet construction information remains fragmented across requirements, BIM models, product databases, construction records, and facility management systems, often using different structures and identifiers for the same physical object (Chen and Jupp, 2018; Pauwels et al., 2017). As objects evolve from requirements to solutions, realised products, and operational assets, these relationships are rarely maintained explicitly, limiting traceability, verification, and lifecycle decision-making (Davari et al., 2023; Corallo et al., 2020). This limitation becomes particularly critical in lifecycle scenarios

involving design changes, product substitution, and long-term asset management, where maintaining consistent object identity across system boundaries is essential.

Classification systems such as OmniClass, Uniclass, MasterFormat, and CoClass support type-level categorisation, while ISO 19650 defines information management processes (CEN, 2019). However, these approaches do not provide persistent, instance-level identifiers capable of maintaining links between individual object representations across lifecycle stages (Rogage and Greenwood, 2020). They are therefore insufficient for lifecycle-consistent traceability required by the Golden Thread (Hackitt, 2018; UK Government, 2024).

Semantic Web technologies and knowledge graphs offer machine-interpretable representations of entities and relationships across heterogeneous data sources (Pauwels et al., 2017; Rasmussen et al., 2019). Persistent and globally unique identifiers are central to such environments, but identifier architectures that support lifecycle traceability while remaining both semantically meaningful and persistent remain underexplored.

The paper makes three contributions: it conceptualises lifecycle-aligned identifier roles for a physical object across its different lifecycle states; analyses how these roles support traceability across heterogeneous representations; and identifies the trade-off between semantic expressiveness and identifier persistence as a key implementation challenge.

The central argument is that lifecycle traceability cannot be achieved by assigning a single identifier to a physical object alone. Instead, traceability requires explicit links between the functional intention, the planned technical solution, the realised technical instance, and the observed functional performance of the same lifecycle object.

This argument is developed by first reviewing related work on lifecycle traceability, classification and identification approaches, Semantic Modelling and Linking (SML), and knowledge graph technologies.

2. BACKGROUND AND RELATED WORK

This section reviews lifecycle traceability, classification and identification approaches, Systems Engineering, SML, and Semantic Web technologies to position the need for lifecycle-oriented identifier architectures.

2.1 Lifecycle information and traceability

Lifecycle information management in construction is challenged by fragmented data environments across BIM models, product databases, and facility management systems, which limits interoperability and traceability between requirements, design decisions, and realised assets (Kebede et al., 2022; Chen and Jupp, 2018).

Traceability, understood as the ability to establish and follow relationships between requirements, design elements, and realised systems, is central in Systems Engineering (Gotel and Finkelstein, 1994). In construction, it supports requirement verification, change management, lifecycle decision-making, and circular economy applications where material and component information must be preserved (Corallo et al., 2020; Davari et al., 2023).

2.2 Classification systems, standards and identification approaches

Classification systems such as OmniClass, Uniclass, MasterFormat, and CoClass structure construction information through standardised taxonomies and support BIM workflows and lifecycle-oriented requirement management (Royano et al., 2023; Tang et al., 2025; Sadrinooshabadi et al.,

2021). However, together with ISO 19650, they primarily operate at class level and do not maintain persistent identification of individual object instances across lifecycle stages.

Global identification frameworks such as GS1 provide identifiers including GTIN, GLN, and GS1 Digital Link for products, locations, and supply chain information (GS1, 2026a; GS1, 2026b). Digital Product Passport initiatives similarly emphasise the linking of product and lifecycle information (Kebede et al., 2023; Kebede et al., 2024). Yet these approaches are not directly aligned with lifecycle modelling of construction objects across design, realisation, and operation.

2.3 Systems Engineering and Semantic Modelling and Linking (SML)

Systems Engineering provides lifecycle logic for linking requirements, design solutions, and realised systems through forward and backward traceability (Chen and Jupp, 2018; Gotel and Finkelstein, 1994).

In construction, SML structures object information into four lifecycle states: Planned Functional (PF), representing functional requirements; Planned Technical (PT), representing design solutions; Realized Technical (RT), representing realised objects; and Realized Functional (RF), representing operational performance (CEN, 2022; CEN, 2024).

Although SE and SML define lifecycle logic and semantic relationships, they do not prescribe how object identity should be maintained across these states in a machine-interpretable way.

2.4 Knowledge graphs and Semantic Web technologies

Semantic Web technologies and knowledge graphs support machine-interpretable integration of heterogeneous construction information (W3C, 2026; Pauwels et al., 2017; Rasmussen et al., 2019; Kebede et al., 2022). They have been applied to link BIM data, product information, DPP data, and lifecycle analysis in the built environment (Kebede et al., 2024).

A key principle is the use of globally unique and persistent IRIs to identify entities across systems (Hogan et al., 2021). However, existing work has focused more on integration and interoperability than on systematic identifier architectures aligned with lifecycle modelling.

2.5 Research gap

The literature shows progress in lifecycle modelling, information structuring, semantic data integration, global identification, and DPP-related lifecycle information. However, these developments remain disconnected and do not provide a unified identifier design for lifecycle-consistent traceability of construction objects.

The gap is therefore the absence of identifier architectures that maintain persistent, semantically interpretable links between functional requirements, technical solutions, realised assets, and operational performance across lifecycle stages.

3. RESEARCH APPROACH AND METHODOLOGY

This study adopts a qualitative, design-oriented approach informed by Design Science Research to develop and analyse a conceptual identifier architecture for lifecycle traceability in knowledge graph-based construction information systems (Hevner et al., 2004; Peffers et al., 2007). The approach is appropriate because the study addresses a design problem, *how persistent identifiers should be structured and related*, rather than testing an existing implementation. Formal evaluation in an operational setting is left for future work.

The analysis is structured by the lifecycle object-state framework defined in EN 17632 (SML) and grounded in Systems Engineering lifecycle logic (Chen and Jupp, 2018; CEN, 2022; CEN, 2024). The

framework distinguishes four lifecycle object states (PF, PT, RT, RF) based on lifecycle stage and level of abstraction (Figure 1). The relations defined in SML link these states and enable traceability across the lifecycle.

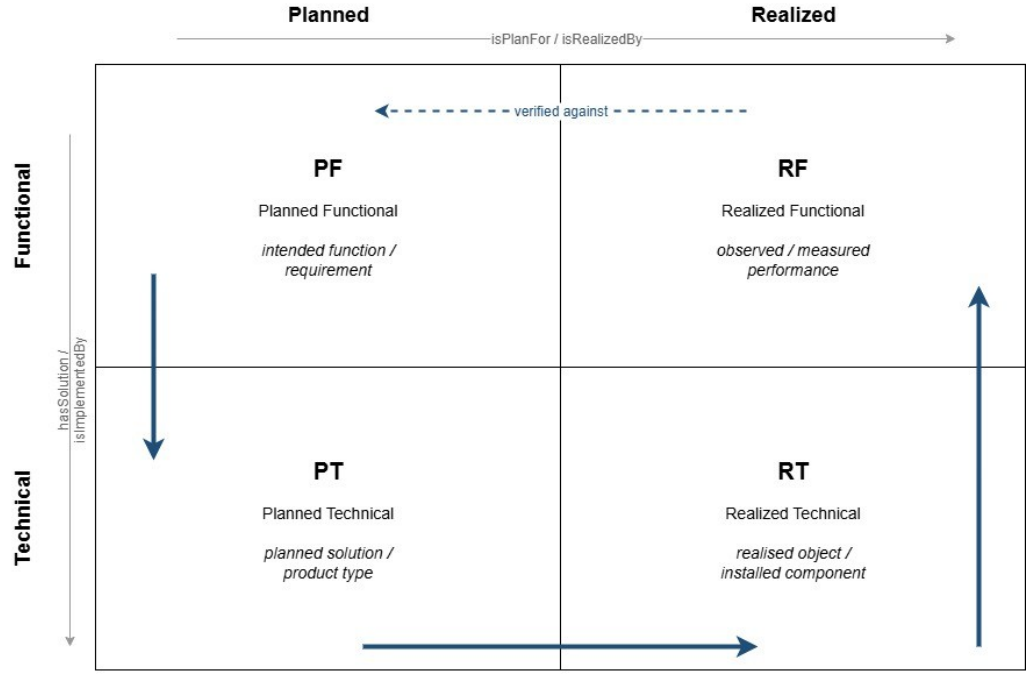

Figure 1. The four lifecycle object states (PF, PT, RT and RF), formed by the lifecycle stage (Planned or Realized) and the level of abstraction (Functional or Technical). The blue arrows show the typical process flow, from intended function (PF) to planned solution (PT), realized object (RT) and observed performance (RF), which is then verified against the original function. Adapted from EN 17632-1:2022, Figure 6 (CEN, 2022).

The study draws on expert-informed material from a collaborative research project, including a technical report on Semantic Web-based object identification (Eckerberg, 2026), conceptual modelling notes and a Turtle-based knowledge graph example (Wikström, 2026), and an industry-oriented report on lifecycle-oriented identifier structures (Jansson et al., 2026).

The materials are treated as conceptual and technical input rather than primary empirical data. The analysis extracts concepts related to lifecycle modelling, identifier persistence, and semantic representation, maps them to the four lifecycle object states, and assesses the resulting architecture in terms of traceability, semantic clarity, and long-term persistence. Particular attention is given to whether identifiers remain stable when object properties, classifications, locations, or system contexts change over time.

The resulting architecture is presented in the following section as a set of identifier roles and relations aligned with the four SML lifecycle states (PF, PT, RT and RF).

4. IDENTIFIER ARCHITECTURE

The proposed identifier architecture follows the lifecycle object states introduced in Section 3, assigning distinct identifier roles to PF, PT, RT, and RF representations across the lifecycle (Chen and Jupp, 2018; CEN, 2022; CEN, 2024).

Conceptually, the identifier architecture can be understood as a four-node lifecycle structure in which each construction object is represented by distinct but explicitly linked identifiers

corresponding to the PF, PT, RT, and RF states. Rather than assigning a single identifier to an object, the architecture defines a set of coordinated identifier roles and relations, enabling traceability across transformations from functional requirements to technical solutions, realised instances, and observed performance. The relationships between these identifier roles follow the lifecycle transitions defined in SML, including realisation, refinement, and verification links between states. These relations include refinement from PF to PT, realisation from PT to RT, and verification from RF to PF, together forming a closed lifecycle traceability loop.

Each lifecycle state has a complementary identifier role: PF represents the intended requirement, PT the selected design solution, RT the realised object instance, and RF the measured performance. This separation enables explicit traceability between requirements, design decisions, realised assets, and operational outcomes (Gotel and Finkelstein, 1994; CEN, 2022; CEN, 2024). It also makes it possible to preserve relationships when a requirement is revised, a design alternative is replaced, or a realised object is substituted during construction or operation.

The primary identifier in the knowledge graph should be globally unique and persistent, and URI-based identification is therefore adopted as the canonical mechanism in line with Linked Data principles that emphasise globally unique, persistent, and dereferenceable identifiers (Hogan et al., 2021). Existing identifiers, such as classification codes, reference designations, IFC GlobalIds, product identifiers, and system-specific IDs, are not replaced but explicitly linked to the persistent URI (Kebede et al., 2022). This approach enables the integration of established industry identification schemes while preserving interoperability and avoiding dependency on any single modelling tool, classification system, or product data platform.

The architecture separates persistent identity from semantic description. Identifiers that encode meaning, such as IEC 81346-1 reference designations, can improve readability but may become unstable when objects change function, location, or configuration. Representing meaning through properties instead supports more robust lifecycle traceability (Hogan et al., 2021; IEC, 2022). For example, the same realised component may be reclassified, moved, or assigned to another system without requiring its persistent URI to change.

Figure 2 illustrates the architecture through an example of an office room lighting system. At the PF level, the identifier denotes the lighting requirement. At the PT level, it denotes the selected design solution. At the RT level, it denotes the installed components, to which existing identifiers such as IFC GlobalIds, GTINs, and serial numbers are linked. At the RF level, it denotes measured performance, enabling verification against the original requirement.

The example illustrates that identifiers at different lifecycle states should not be treated as duplicates. Rather, they identify different but related information objects: a requirement, a design solution, an installed technical instance, and a measured functional outcome. The value of the architecture lies in maintaining explicit links between these objects so that verification, replacement, and performance follow-up can be traced over time.

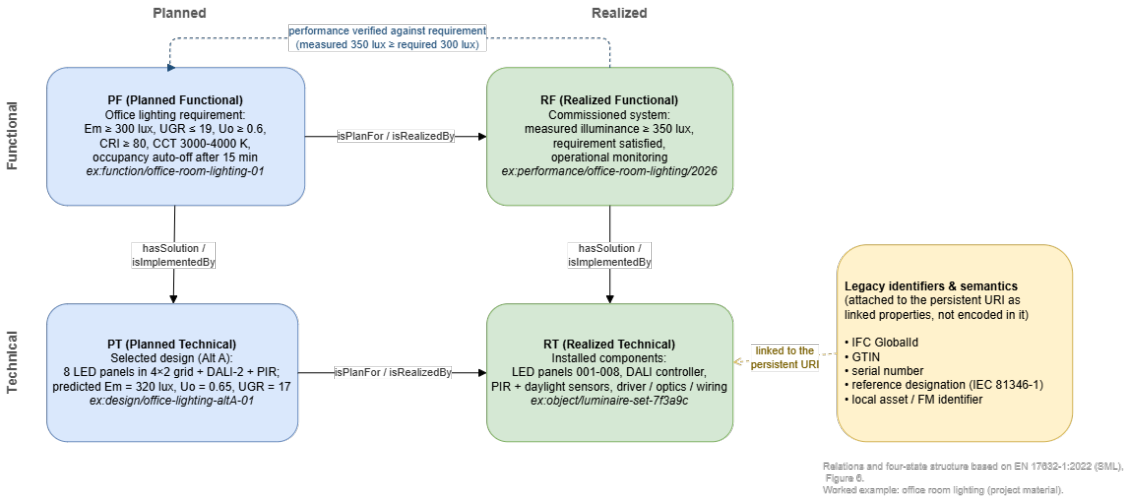

Figure 2. The identifier architecture applied to an illustrative office room lighting example, showing the four persistent URIs (PF, PT, RT, RF), their SML relations, and the legacy identifiers linked to the realised object. Structure based on EN 17632-1:2022 (CEN, 2022).

Each lifecycle role (PF, PT, RT, RF) carries its own globally unique, persistent URI. The URIs are linked through the SML relations, providing explicit traceability from requirement to performance. Existing identifiers and semantic meaning are represented as properties of the URI rather than encoded in the identifier string, which keeps the identifier stable when objects are relocated, replaced, or reclassified.

This structuring introduces a fundamental trade-off between semantic expressiveness and identifier persistence. While semantically rich identifiers may improve interpretability, they risk becoming unstable as objects evolve across lifecycle stages. The proposed architecture resolves this by prioritising identifier persistence, assigning stable URIs to each lifecycle role while expressing semantic meaning through associated properties and relations rather than within the identifier itself.

5. DISCUSSION

The proposed architecture addresses the lack of lifecycle-aligned identification by linking PF, PT, RT, and RF states in a machine-interpretable way, consistent with Systems Engineering traceability logic (Chen and Jupp, 2018; Gotel and Finkelstein, 1994). Compared with classification-based approaches, it maintains identity at instance level, allowing realised objects to remain traceable even when classification, configuration, or location changes. The main implication is that lifecycle traceability depends not only on persistent identifiers, but also on the explicit relations that connect different lifecycle representations of the same object.

For example, in a construction project, a lighting requirement defined at the PF level may initially be linked to a planned luminaire type at the PT level. If the selected product is later replaced during procurement or installation, the RT identifier changes, while the PF and PT identifiers remain unchanged. By maintaining explicit links between these identifiers, the architecture enables traceability of requirement fulfilment despite changes in technical solutions. This illustrates how lifecycle-aligned identifier roles enable not only traceability, but also systematic handling of change and verification across lifecycle stages.

A key insight is the tension between semantic expressiveness and identifier persistence. Meaningful identifiers can improve human readability but risk instability when objects evolve. By

separating persistent identity from semantic description, the proposed architecture allows meaning to be expressed through properties and relationships, supporting both semantic clarity and robust lifecycle traceability.

The architecture also has implications for Digital Product Passports and the Golden Thread, both of which depend on reliable links between product data, requirements, design information, realised assets, and operational performance (Kebede et al., 2024). In practice, it must accommodate multiple existing identification systems, including classification codes, product identifiers, RFID-based tracking, and system-specific IDs, by linking rather than replacing them (Pyke and Madan, 2013; Huo et al., 2023). This linking strategy is especially important in construction, where information is typically produced and maintained by many actors over long periods.

From an implementation perspective, the architecture implies that governance is as important as technical modelling. Organisations need rules for URI minting, ownership, versioning, deprecation, and maintenance of links between lifecycle states. Without such governance, persistent identifiers may exist technically but fail to support reliable traceability in practice.

The study remains conceptual, and the architecture must be evaluated in real-world implementations involving large-scale data integration, organisational boundaries, lifecycle change management, and long-term information continuity. Such evaluations should examine not only technical feasibility, but also usability for project actors, compatibility with existing information management processes, and the cost of maintaining identifier links over time.

6. CONCLUSION

This paper examined how identifier architectures can support lifecycle-consistent traceability in knowledge graph-based construction information systems. Building on Systems Engineering and SML, it addressed the lack of persistent, instance-level identification strategies for linking requirements, technical solutions, realised objects, and operational performance across lifecycle stages.

The main contribution is a lifecycle-aligned identifier architecture distinguishing PF, PT, RT, and RF states. By separating these identifier roles, the architecture enables explicit, machine-interpretable traceability across heterogeneous data sources while preserving object identity over time.

The study is conceptual and based on expert-informed material, but it clarifies an important design principle: persistent identity should be separated from semantic description while explicit relations maintain lifecycle meaning. Future research should validate the architecture in large-scale implementations involving BIM models, Digital Product Passports, operational systems, versioning, change management, governance models, and real-time data integration.

ACKNOWLEDGEMENTS

This work was supported by the Smart Built Environment projects “Digitala datablad och Datamallar” and “Interoperabilitet – Digital samverkan för den byggda miljön”, which aims to strengthen digital collaboration and interoperability in the built environment through frameworks and recommendations for information exchange, standards, and linked data approaches.

REFERENCES

CEN (2019) SS-EN ISO 19650-1:2019 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling – Part 1: Concepts and principles (ISO 19650-1:2018). Stockholm: Swedish Standards Institute (SIS).

CEN (2022) EN 17632-1:2022 Building information modelling (BIM) – Semantic modelling and linking (SML) – Part 1: Generic modelling patterns. Brussels: European Committee for Standardization.

CEN (2024) EN 17632-2:2024 Building information modelling (BIM) – Semantic modelling and linking (SML) – Part 2: Domain-specific modelling patterns. Brussels: European Committee for Standardization.

Chen, Y. and Jupp, J. (2018) 'Model-based systems engineering and through-life information management in complex construction', in Product Lifecycle Management to Support Industry 4.0. IFIP Advances in Information and Communication Technology, vol. 540. Cham: Springer, pp. 80–92. https://doi.org/10.1007/978-3-030-01614-2_8

Corallo, A., Latino, M.E., Menegoli, M. and Pontrandolfo, P. (2020) 'A systematic literature review to explore traceability and lifecycle relationships', International Journal of Production Research, 58(15), pp. 4789–4807. https://doi.org/10.1080/00207543.2020.1771455

Davari, S., Jaberi, M., Yousfi, A. and Poirier, E. (2023) 'A traceability framework to enable circularity in the built environment', Sustainability, 15(10), 8278. https://doi.org/10.3390/su15108278

Eckerberg, K. (2026) Identifiering av objekt med semantisk webbteknik [Identification of objects using Semantic Web technologies, in Swedish]. Unpublished technical report, available from the authors upon request.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004) 'Design science in information systems research', MIS Quarterly, 28(1), pp. 75–105. https://doi.org/10.2307/25148625

Gotel, O.C.Z. and Finkelstein, A.C.W. (1994) 'An analysis of the requirements traceability problem', in Proceedings of the IEEE International Conference on Requirements Engineering. IEEE, pp. 94–101. https://doi.org/10.1109/ICRE.1994.292398

GS1 (2026a) An Introduction to GS1 Identification Keys. Available at: https://www.gs1.org/standards/id-keys (Accessed: 15 June 2026).

GS1 (2026b) GS1 Digital Link Standard. Available at: https://www.gs1.org/standards/gs1-digital-link (Accessed: 15 June 2026).

Hackitt, J. (2018) Building a Safer Future: Independent Review of Building Regulations and Fire Safety: Final Report. London: Ministry of Housing, Communities & Local Government. Available at: https://www.gov.uk/government/publications/independent-review-of-building-regulations-and-fire-safety-final-report (Accessed: 18 June 2026).

Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Neumaier, S., Polleres, A., Sequeda, J., Staab, S. and Zimmermann, A. (2021) Knowledge Graphs. San Rafael: Morgan & Claypool Publishers.

Huo, H. *et al.* (2023) 'Design and application of intelligent construction quality inspection system based on BIM-RFID', Advances in Civil Engineering, 2023, Article ID 5638265. https://doi.org/10.1155/2023/5638265

IEC (2022) IEC 81346-1:2022 Industrial systems, installations and equipment and industrial products – Structuring principles and reference designations – Part 1: Basic rules. Geneva: International Electrotechnical Commission.

Jansson, T. et al. (2026) Byggnadsverkets informationsmodell och tillämpningsexempel [Building asset information model and application examples, in Swedish]. Available at: www.smartbuilt.se/Media/13xaawas/byggnadsverkets-informationsmodell-och-tillampningsexempel.pdf (Accessed: 25 June 2026).

Kebede, R., Moscati, A., Tan, H. and Johansson, P. (2022) 'Integration of manufacturers' product data in BIM platforms using semantic web technologies', Automation in Construction, 144, 104630. https://doi.org/10.1016/j.autcon.2022.104630

Kebede, R., Moscati, A., Tan, H. and Johansson, P. (2023) 'Circular economy in the built environment: A framework for implementing digital product passports with knowledge graphs', in Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference, Heraklion, Greece, 10–12 July 2023, pp. 925–932. https://doi.org/10.35490/EC3.2023.245

Kebede, R., Moscati, A., Tan, H. and Johansson, P. (2024) 'A modular ontology modeling approach to developing digital product passports to promote circular economy in the built environment', Sustainable Production and Consumption, 48, pp. 248–268. https://doi.org/10.1016/j.spc.2024.05.007

König, M., Koch, C., Beetz, J. and Borrmann, A. (2018) *Building Information Modeling: Why? What? How?* Cham: Springer. https://doi.org/10.1007/978-3-319-92862-3

Pauwels, P., Zhang, S. and Lee, Y.C. (2017) 'Semantic web technologies in the AEC industry: A literature overview', Automation in Construction, 73, pp. 145–165. https://doi.org/10.1016/j.autcon.2016.10.003

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007) 'A design science research methodology for information systems research', Journal of Management Information Systems, 24(3), pp. 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pyke, C.R. and Madan, I. (2013) 'Breaking barriers to interoperability: assigning spatially and temporally unique identifiers to spaces and buildings', Annals of the New York Academy of Sciences, 1295(1), pp. 10–17. https://doi.org/10.1111/nyas.12225

Rasmussen, M.H., Lefrançois, M., Pauwels, P., Hviid, C.A. and Karlshøj, J. (2019) 'Managing interrelated project information in AEC knowledge graphs', Automation in Construction, 108, 102956. https://doi.org/10.1016/j.autcon.2019.102956

Rogage, K. and Greenwood, D. (2020) 'Data transfer between digital models of built assets and their operation and maintenance systems', Journal of Information Technology in Construction (ITcon), 25, pp. 469–481. https://doi.org/10.36680/j.itcon.2020.027

Royano, V., Gibert, V., Serrat, C. and Rapinski, J. (2023) 'Analysis of classification systems for the built environment: Historical perspective, comprehensive review and discussion', Journal of Building Engineering, 67, 105911. https://doi.org/10.1016/j.jobe.2023.105911

Sacks, R., Lee, G., Burdi, L. and Bolpagni, M. (2025) *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers*. 4th edn. Hoboken: John Wiley & Sons.

Sadrinooshabadi, S., Taheri, A., Yitmen, I. and Jongeling, R. (2021) 'Requirement management in a life cycle perspective based on ISO 19650-1 and CoClass as the new classification system in Sweden', Engineering, Construction and Architectural Management, 28(9), pp. 2736–2753. https://doi.org/10.1108/ECAM-07-2020-0538

Tang, S., Bito, T. and Shide, K. (2025) 'Automatic classification of BIM objects based on IFC data using the Uniclass classification standard', Buildings, 15(13), 2347. https://doi.org/10.3390/buildings15132347

UK Government (2024) Keeping information about a higher-risk building: the golden thread. Available at: https://www.gov.uk/guidance/keeping-information-about-a-higher-risk-building-the-golden-thread (Accessed: 25 June 2026).

W3C (2026) Semantic Web Wiki: Main Page. Available at: https://www.w3.org/2001/sw/wiki/Main_Page (Accessed: 25 June 2026).

Wikström, L. (2026) Anteckningar och kunskapsgrafsexempel för URI-baserad identifiering [Notes and knowledge graph example on URI-based identifier modelling, in Swedish]. Unpublished material and meeting notes, available from the authors upon request.

Zhu, J. and Wu, P. (2022) 'BIM/GIS data integration from the perspective of information flow', Automation in Construction, 136, 104166. https://doi.org/10.1016/j.autcon.2022.104166