An end-to-end Asset Life Cycle Knowledge Graph

Conor Shaw (corresponding author), (conor.shaw@ucdconnect.ie)
Energy Institute, University College Dublin, Ireland

Cathal Hoare, (cathal.hoare@ucd.ie)
Energy Institute, University College Dublin, Ireland

Martijn de Riet, (martijn@bimforce.com)
Bimforce, Deflt, Netherlands

Flávia de Andrade Pereira, (flavia.pereira@ucl.ac.uk)
University College London, United Kingdom

James O'Donnell, (james.odonnell@ucd.ie)
Energy Institute, University College Dublin, Ireland

Keywords: Built Environment, Asset Management, Information Management, Knowledge Graph, Life Cycle Cost

Abstract
The current information landscape is not fit for the evolving requirements around sustainability reporting and the management of built assets over their life cycle. This paper builds on prior stakeholder requirements-gathering and ontology development work, detailing the implementation stages to arrive at an end-to-end Asset Life Cycle Knowledge Graph and sample queries to support several analytical use cases. The prototype graph and queries are evaluated qualitatively through broad stakeholder focus groups and the study receives resounding positivity that the technical contribution pushes the IT trajectory in a suitable direction, but that a significant shift is required in the commercial paradigm around information management and sustainability reporting incentives. The domain insight, generated via the expert focus groups, provides useful future research directions, pointing to feasible alignment with bSI initiatives such as bsDD for intuitively visualising asset data and ensuring long-term information sustainability.

1. Introduction
The imminent climate crisis is driving a need to dramatically reduce resource consumption in the built environment (International Energy Agency, 2023). Environmental policy is one mechanism which is changing the way assets are valued in the built environment by making the reporting of sustainability indicators mandatory where previously it was voluntary and market driven (European Commission, 2023).

Standardisation within the built environment presents significant challenges, a topic extensively explored within the scientific community (Gao & Pishdad-Bozorgi, 2019). Due to the inherent complexity of Life Cycle Sustainability Analysis (LCSA) methods, reporting frameworks are still in development and lack consensus. Although these frameworks commonly rely on Life Cycle Analysis (LCA) as a foundational methodology, LCA's effectiveness is often constrained by the availability and quality of data (Gao &

Pishdad-Bozorgi, 2019). Moreover, the current software landscape exacerbates these issues, as many analytical tools are proprietary, costly, and ill-suited to the scale required for broader implementation (Shaw et al., 2024a). Asset Life Cycle (ALC) Information Management (IM) is rapidly becoming a priority for governments, portfolio owners and Asset Managers (AMs) who are scrambling to future-risk-proof their assets and avoid *asset stranding*. There has been significant work in building consensus and standardising concepts to support advanced IM, for example, by encoding domain knowledge as ontologies (Gao et al., 2020; Ghose et al., 2022; Hammar et al., 2019; Röck et al., 2024; W3C, 2022), in part for their ability to facilitate logical reasoning over graph databases on the web .

Graph databases are gaining interest, in both the research community and industry, over traditional relational databasing techniques for their ability to efficiently integrate, store and retrieve heterogeneous data. (Guyo & Hartmann, 2024) compare the approaches and suggest that graph databases are highly suited for applications where complex, interconnected data needs to be efficiently traversed, a future strategic need of ALC information systems (Fang et al., 2022). Bringing together graph databasing, and shared vocabularies results in the concept of a Knowledge Graph (KG). Previous work by the authors (Shaw et al., 2024b) provides a harmonisation of governance and practitioner requirements for sustainability policy-aligned IT landscape in the near future, which includes the need for ALC KGs, and is the point of departure for this paper.

The primary objective of this study is to develop and evaluate an end-to-end ALC KG prototype as an effective underlying data source for future IM needs in the built environment. We make use of the implementation stages of the Linked Open Terms (LOT) ontology development methodology (Poveda-Villalón et al., 2022) and Design Science Research (Holmström et al., 2009) to implement and query a knowledge graph, and evaluate the functionality through multidisciplinary focus groups (Figure 1). An additional objective of the study is the generation of domain insight and directions for future work to support in-practice uptake.

The remainder of this paper is structured as follows. Section 2 and it's sub-sections describe the steps to develop and evaluate the ALC KG prototype. Section 3 presents the results of the evaluation which are then discussed in section 4, and the paper summarised with high-level reflections in section 5.

2. Developing an Asset Life Cycle Knowledge Graph

As discussed, KGs are gaining popularity as a method for IM in the built environment but there is a paucity of studies which build applications on top of the underlying structured domain knowledge. Some examples include the work of Mavrokapnidis et al. (2023) and Wang et al. (2021). Based on the specific aims of this study and the harmonised governance and practitioner requirements for asset IM from our previous work, we propose a ALC KG as the codified, underlying KG for broad stakeholder IM and analytical needs, including LCA functions. The following sub-sections detail this research process.

2.1. Asset life cycle knowledge graph requirements

Previous work by the authors provides a concept schema for ALC IM representing a harmonisation of governance-practitioner requirements for a future environmental policy-aligned information landscape for the built environment (Shaw et al., 2024b), which include an need to;

- ***access*** information over the web;
- ***integrate*** distributed data sources;
- ***aggregate*** information consistently (spatial / inheritance / composition etc.);

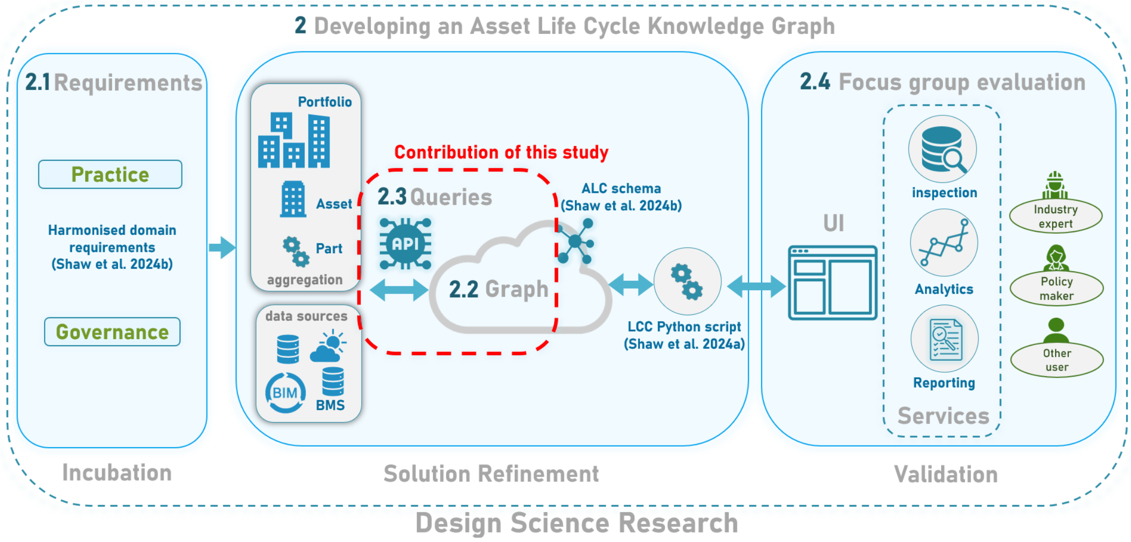

Figure 1: Sections of this paper are mapped to their methodological steps for developing an Asset Life Cycle Knowledge Graph using a Design Science Research framework.

- ***support extensibility*** to add cost/resource items;
- ***analyse*** multi-criteria LCSA scenarios;
- ***visualise*** analysis in a variety of mediums base on stakeholder preference; and
- ***demonstrate data quality***, transparency and reproducability.

In accordance with the LOT methodology, and in order to support development and validation of the technical aspects of this study, suitable Competency Quesitons (CQs) are selected (from Shaw et al., 2024b) for which the prototype KG should be able to provide answers as a proof-of-concept.

- CQ1 - What is the [Lifetime] of Asset with [AssetID]?
- CQ2 - Which Assets [List] have a [ResidualValue] > '0'?
- CQ3 - What is the average [AnnualEnergyCost] of Assets with [AssetType]?
- CQ4 - Return all assets which have an initial cost > '500000'
- CQ4 - Retrieve inputs for conducting LifeCycleCost analysis for asset [List]?

Based on the schema and Ontology Requirements Specification Documentation (OSRD) detailed in Shaw et al. (2024b), the next step is to develop a prototype ALC KG.

2.2. Integrating distributed data sources using a graph database

Graph databases consist of nodes (entities) and edges (relationships), and are enriched with attributes and metadata, allowing for complex interconnections and the encoding of domain logic with semantic richness. Graph databases support the integration of data from distributed sources, provide ground truth for interdisciplinary applications and support deconstruction of domain specific information silos (Ji et al., 2022).

Sample asset data to test the ALC KG consists of the asset register (∼4000 assets) from a large infrastructure case study presented in Shaw et al. (2024a). A KG is created and managed using Neo4j[1], a state-of-the-art graph-based Database Management System (DBMS), which provides an overarching data organisation for the ALC KG. Features of Neo4j's query language are used to fuse sample data from distributed sources

[1] Neo4j documentation

including the building management system (BuildingManagementSystem.csv), supplier information (Supplier.csv) and asset register (AssetRegister.csv). Information about the study scope is also captured to organise the provenance of analysis events in a separate 'ApplicationServer.csv' dataset. Data from each information source is read and organised based on domain logic described by Shaw et al. (2024b). Four entity types are represented, including *Energy Use*, *Asset*, *Cost* (of asset) and *Condition* (of asset); *Cost* is further subclassed by several cost types. The analysis study period is described by an *Analysis* entity and this is further contextualised through the inclusion of *inflation* and *discount rates* which capture prevailing economic conditions. Each of the entities is related through a named relationship, allowing it to be included when reasoning about the data.

The graph database schema is described in Figure 2 and the instantiated database for the case study can be accessed through a browser[2].

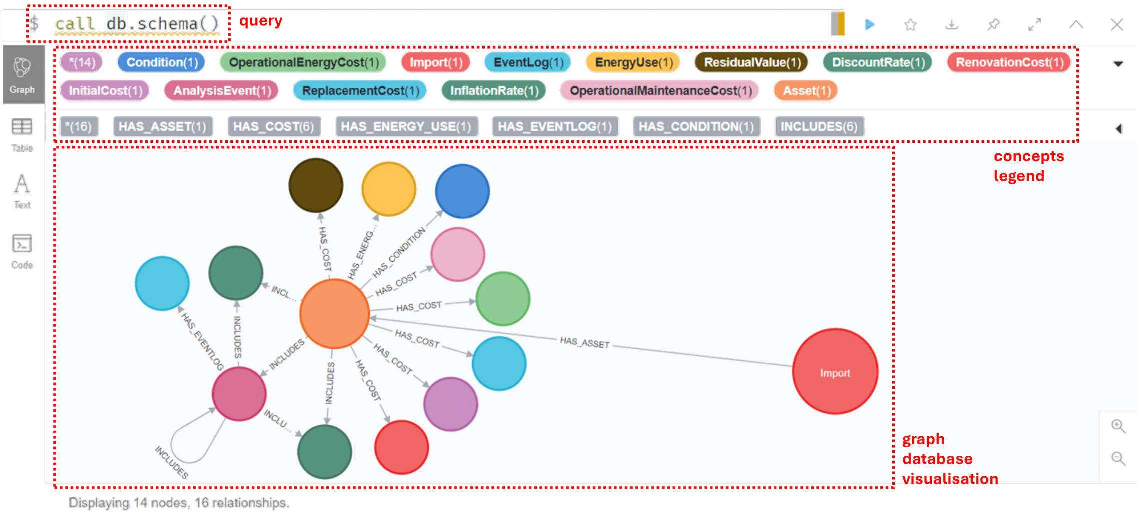

Figure 2: The LCAIM schema is retrieved via a Cypher query within Neo4j. The 'import' concept represents the asset register.

2.3. Querying the knowledge graph

The next step in the development process is to verify technical functionality of the KG by running queries. In order to demonstrate the retrieval of structured data from the graph DB a number of Cypher queries are developed based on the CQs[3] from Section 2.1. Listing 1 demonstrates CQ-1 which returns lifetime information for a selected asset, and CQ-4 which searches the DB for costs based on a given threshold.

In order to validate the ALC KG functionality so that it supports industrial application, it should be tested in practical settings. In practice, AM stakeholders need to perform complex lifecycle calculations using asset data as inputs, and so CQ-5 is developed as a Cypher query which retrieves the information requirements to conduct a comparative (financial) Life Cycle Cost (LCC) analysis between two assets (Figure 3). Due to the operational nature of the query strategy, the approach supports potential for scale in line with domain requirements for aggregation and scalability across organisational types and portfolio sizes. The returned data is used as input for a LCC analysis script provided by Shaw et al. (2024a) which uses the Python programming language. This practical,

[2] Neo4j graph database for the case study asset register (click 'connect' to access the graph)

[3] sample Cypher queries

end-to-end demonstration of the data storage, retrieval and analysis is used as a basis for verification with AM stakeholders in the following step.

```
// CQ1 - What is the [Lifetime] of Asset with [AssetID]?
:param asset_id => "82341808";
MATCH (a:Asset) WHERE a.asset_id = $asset_id
RETURN a.lifetime AS Lifetime
   "Lifetime"
------------
  15  .0

// CQ4 - Return all assets which have an initial cost >
      500000
MATCH (a:Asset)-[:HAS_COST]->(ic:InitialCost)
WHERE ic.value > 500000
RETURN a, ic
   "AssetID"    "InitialCost"
--------------------------------
  9658              800000

  9657              800000

```

Listing 1: Cypher query generated for CQs 1 and 4 demonstrating asset and attribute retrieval

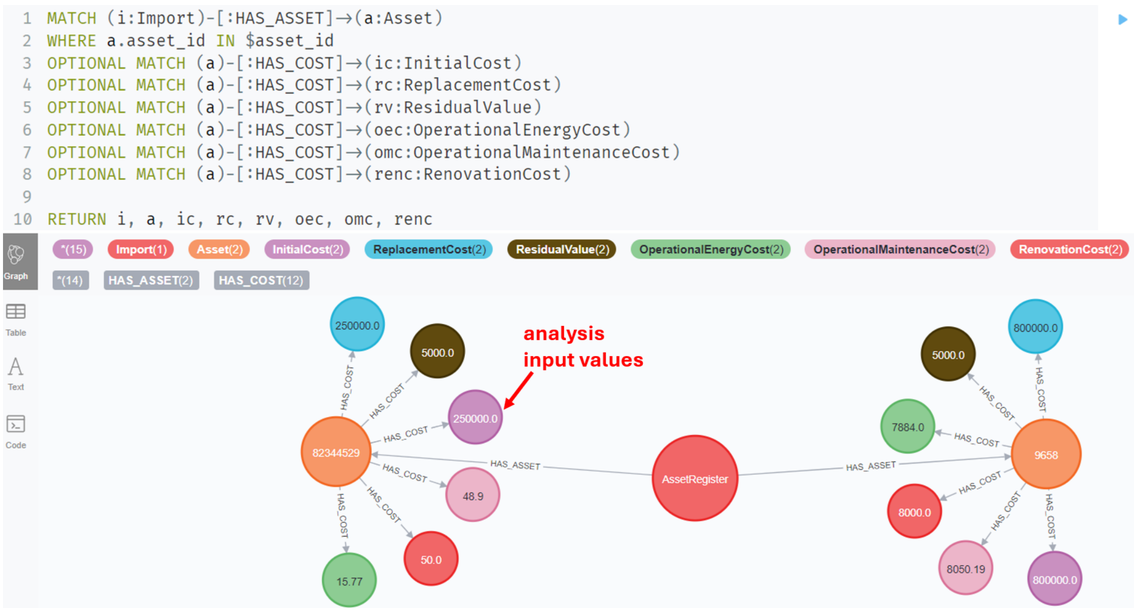

Figure 3: Query generated from CQ-5 to retrieve LCC analysis inputs from the graph database for two selected assets.

2.4. Focus group evaluation

The evaluation strategy is informed by a variety of qualitative methods, and developed specifically to address the research objectives of validating the prototype ALC KG and

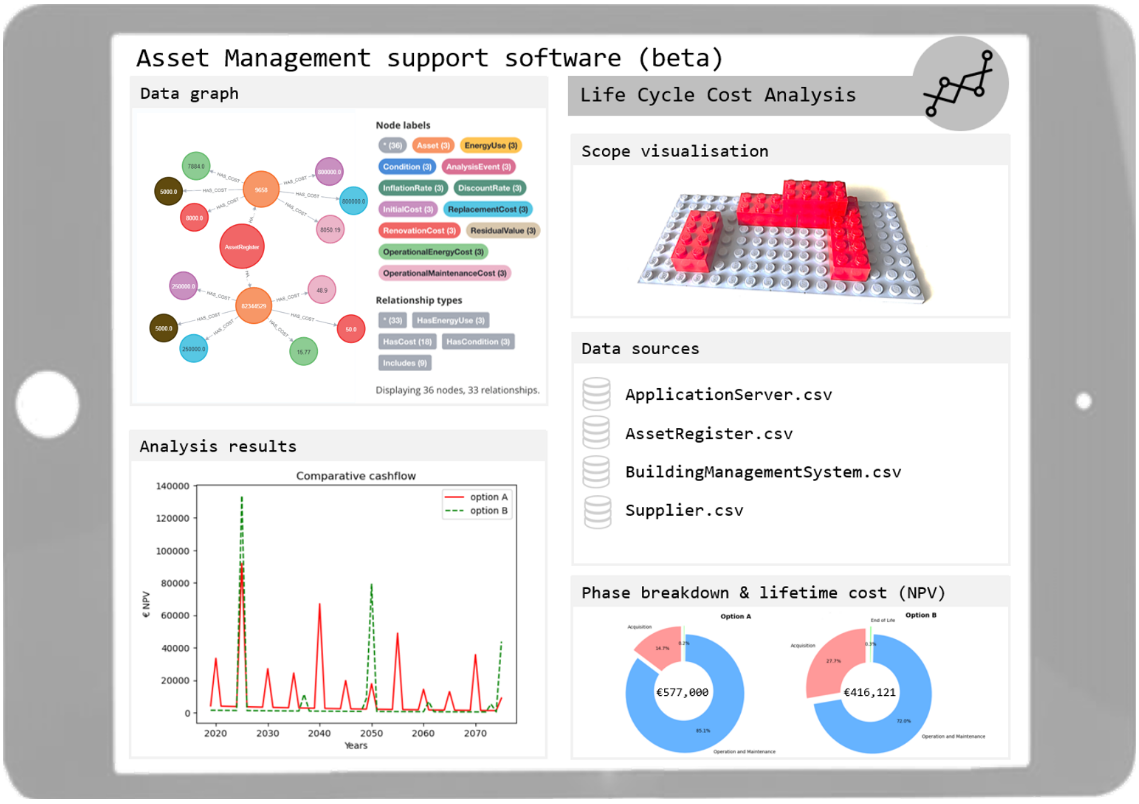

Figure 4: A mock-up decision-support system UI is presented to domain experts, facilitating evaluation of the knowledge graph functionality. In this use case the user can investigate linked databases, which supports their analytical requirements, via a graph visualiser
.

generating directions for future research and domain insight for in-practice uptake. Due to the *harmonisation* aims of the research (targeting wide adoption among a diverse stakeholder group), an immersive focus group experience is designed to gather input from a variety of AM stakeholders (Hevner et al., 2004; Onwuegbuzie et al., 2009).

Participants were sought for their specific expertise via the research group's extensive industry and pubic-sector network, and the cohorts were designed to be interdisciplinary, encouraging multiple viewpoints and perspectives for debate. To enable comprehension and interaction by non-technical domain experts in evaluating the functionality of the prototype ALC KG, experts are presented with a mock-up decision-support IT system User Interface (UI) which simulates an analysis scenario (Figure 4).

Functionality of the system to retrieve information, conduct analysis and visualise the results is demonstrated. The participants are then asked to discuss open ended questions about the system functionality, such as *"What are the advantages and limitations of the system in relation to your discipline"* and "What implications might the system have in terms of bringing LCSA methods into widespread use". Thematic analysis methods are employed in a similar approach as described in detail in Shaw et al. (2023). The results are then synthesised within the study context and in relation to the technical state-of-the-art and developing environmental policy landscape.

3. Results

This section summarises both the learnings from the development and verification process (graph development, data retrieval via queries, conducting demonstration analysis) and the focus group evaluation (validation and generation of relevant domain insights).

Table 1: Focus group cohort details and emergent themes

Group	Expert	Emergent themes
1	(1.1) SAP arch. - housing	Intuitiveness reduces training requirement
	(1.2) LCA consultant	Customisation very important
2	(2.1) Facility Manager - REIT	Closed data greatest barrier
	(2.2) Asset Manager - infra.	Business case for open source
	(2.3) Policy expert - energy	Significant potential/risk for gaming the sustainability reporting system currently
	(2.4) CTO - info. mgmt.	Graph visualisation not broadly intuitive

The functionality of the proposed methodology to support domain exchange requirements is verified by demonstrating successful data retrieval using queries based on the CQs. Furthermore, in-practice verification is satisfied on the grounds that financial LCC analysis could be conducted using the input data retrieved from the KG. In terms of validation with stakeholders, Table 1 provides details about the focus group cohorts and a summary of the emergent themes of the qualitative analysis.

It was found that, in general, participants see significant value in the demonstrated functionality. However, the majority are doubtful as to whether sufficient data is typically available in practice to support such analytics, with one participant explaining that *"even properties that are sold to us as LEED platinum are using massively excessive kWhs. It's just not true what they're being accredited as"* (Participant 2.1). This sentiment was echoed by another participant who highlighted the blatant commercial incentive for assessors to inflate ESG results, with little risk of discovery. *"As it stands with the current IT setup, this type of analytics is impossible to verify"* (Participant 1.1). In terms of usability of the graph database, as presented to the participants, there were mixed views regarding the intuitiveness of visualising graph databases for non-technical / no-code data navigation. Several described it as a compelling alternative to *"wading through spreadsheets"*, however, one participant who had experience with user research in this specific area, explained that "providing information consumption options based on user preferences is essential as it's not as intuitive as we data-people might like to think" (Participant 2.4). For this reason, the study is limited in that it does not connect asset data with geometric building information models for viewing in a 3D environment, which would be a highly-intuitive alternative (Yalcinkaya & Singh, 2018).

A theme which emerged repeatedly during the focus group sessions concerned *openness* of software and data. The cohort were in broad agreement that a fundamental underlying issue impeding broad adoption of LCSA techniques is the persistence of closed data practices in the industry. *"We just can't get the data out of our assets. The BMS is closed… All of them are spitting out different types of data and it's impossible to handle it without major capacity-building in terms of our IT skills"* (Participant 2.1). The group, which included significant large-scale IT architecture expertise, were adamant that open, semantically-structured data and processes are the only feasible answer to the unprecedented IM challenge posed by impending environmental policy. Though there was some pessimism about actualising such a significant paradigm shift in an industry which has, historically, seen slow technical adoption (Mischke et al., 2017), there was positivity in terms of envisioning a business-case for delivering open source IT. The participants discussed the potential for organisation-specific implementations and customisation based on extensible open-source applications and open data standards, as well as promising advances in ETL (Extract Transform Load) processes *"which are likely to remain unavoidable for the foreseeable future"* (Participant 1.2).

4. Discussion

This study provides an end-to-end prototype asset information storage, retrieval, analysis and consumption strategy, which aligns with the harmonised governance - practitioner requirements for future IT system to support the current environmental policy trajectory. An Asset Life Cycle Knowledge Graph and queries are presented which verify the data-retrieval functionality to support an analytical use case=. Due to the operational nature of the query strategy, the concept supports significant potential for scale in line with user requirements for composition/aggregation within future Asset Life Cycle information system.

The diverse stakeholders which comprise the focus group cohort make it clear that there are commonly applicable information categories and functionality between disciplines, but that *what* exactly is viewed by different stakeholders based on their expertise and goals varies. Future AIM systems need to support both aggregation and multi-medium *views* of information. Overall, the response was positive and the UI concept was found to align with the harmonised governance-practice requirements for policy-aligned IM systems for the built environment. Realising this fundamental shift, however, may require fundamentally reorganising business models (towards open ones) and commercial incentive structures (towards objective ones) related to IM and sustainability reporting. Such a paradigm shift is viewed by the experts with reasonable skepticism, but they agree that the work presented in this study *"moves the dial"* in the right direction.

Several limitations of this study suggest future research directions. Firstly, the validation stage demonstrates a limited set of queries and future iterations would benefit from implementing further business and analytical use cases. Second, user feedback was limited by the number of focus group studies, and future work could be improved by evaluating ALC KG developments with a wider and further-diversified cohort. Finally, though the UI developed shows both the graph database visualisation and an abstract suggestion of a 3D model, a technical solution to *link* the 3D model and asset data has not been provided. This is essential in addressing the concerns vocalised by stakeholders regarding data viewing-medium flexibility.

Linking asset data to IFC models using, for example the BuildingSMART Data Dictionary (bsDD), would allow graphical representation and fulfill the ambition for an open source - open data, risk-averse information strategy. Using the IFC schema to build ALC KGs is feasible due to the existence of overlapping concepts, including *IfcAsset*[4], *IfcCostItem*[5], and *Pset_EnvironmentalImpactIndicators*[6]. In short, since both *Assets* and *CostItems* are *concepts* in the schema (not tied to an object instance or type definition) and this suits the abstract nature of assigning costs to aggregations of assets based on domin logic. For the *PSet_EnvironmentalImpactIndicators*, it's somewhat different. Since these are attached to the *IfcElement* (and therefore apply to all physical building components by inheritance), they require a geometrical object definition. This, however, is quite logical since environmental indicators are a physically-related field. Future iterations of this work will benefit from exploring alignment with bSI initiatives, especially in the context of linking graph-based asset data with bsDD definitions.

[4] bSI documentation - IfcAsset

[5] bSI documentation - IfcCostItem

[6] bSI documentation - Pset_EnvironmentalImpactIndicators

5. Conclusion

This study demonstrates an end-to-end Asset Life Cycle Knowledge Graph prototype and associated queries, addressing the future needs for robust information management and sustainability reporting in the built environment. The developed methodology aligns with regulatory and practitioner requirements offering a scalable direction to address diverse domain stakeholder needs. Expert feedback from the focus group evaluation highlights the importance of open data and open source software initiatives, and emphasises the necessity for future systems to support aggregation and multi-medium information visualisation.

Despite the positive reception, the study underscores the need for a paradigm shift in business models and incentive structures to fully realise these advancements, but suggests a compelling business-case for organisation-specific implementation of open source solutions. Limitations identified in the study include the narrow scope of use cases and the restricted number of focus group evaluations. Future research should expand the range of business and analytical use cases, engage a broader stakeholder cohort, and develop technical solutions to link 3D models with asset data, leveraging standards like the bsDD. Aligning with such initiatives will enhance the practicality and sustainability of information management systems, moving towards an open and risk-averse strategy as we tackle environmental concerns in the built environment.

Acknowledgements

The authors thank Julia Kaltenegger (TU/e), Esra Bektas (TNO) and the Ibec facility manager network for their valuable input. This work was supported by the CBIM-ETN funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860555.

References

European Commission. (2023, July). The Commission adopts the European Sustainability Reporting Standards. Retrieved October 17, 2023, from https://finance.ec.europa.eu/news/commission-adopts-european-sustainability-reporting-standards-2023-07-31_en

Fang, Z., Liu, Y., Lu, Q., Pitt, M., Hanna, S., & Tian, Z. (2022). BIM-integrated portfolio-based strategic asset data quality management. *Automation in Construction*, *134*, 104070. https://doi.org/10.1016/j.autcon.2021.104070

Gao, X., & Pishdad-Bozorgi, P. (2019). BIM-enabled facilities operation and maintenance: A review. *Advanced Engineering Informatics*, *39*, 227–247. https://doi.org/10.1016/j.aei.2019.01.005

Gao, X., Pishdad-Bozorgi, P., Tang, S., & Shelden, D. (2020). Machine Learning-Based Building Life-Cycle Cost Prediction: A Framework and Ontology. https://doi.org/10.1061/9780784482865.116

Ghose, A., Lissandrini, M., Hansen, E. R., & Weidema, B. P. (2022). A core ontology for modeling life cycle sustainability assessment on the Semantic Web. *Journal of Industrial Ecology*, *26*(3), 731–747. https://doi.org/10.1111/jiec.13220

Guyo, E. D., & Hartmann, T. (2024). Evaluating the efficiency and performance of data persistent systems in managing building and environmental Data: A comparative study. *Advanced Engineering Informatics*, *62*, 102582. https://doi.org/10.1016/j.aei.2024.102582

Hammar, K., Wallin, E. O., Karlberg, P., & Hälleberg, D. (2019). The RealEstateCore Ontology [Series Title: Lecture Notes in Computer Science]. *The Semantic Web – ISWC 2019*, *11779*, 130–145. https://doi.org/10.1007/978-3-030-30796-7_9

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research [Publisher: Management Information Systems Research Center, University of Minnesota]. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Holmström, J., Ketokivi, M., & Hameri, A.-P. (2009). Bridging Practice and Theory: A Design Science Approach. *Decision Sciences*, *40*, 65–87. https://doi.org/10.1111/j.1540-5915.2008.00221.x

International Energy Agency. (2023). *World Energy Outlook 2023* (tech. rep.). https://www.iea.org/reports/world-energy-outlook-2023

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition and Applications [arXiv:2002.00388 [cs]]. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Mavrokapnidis, D., Fierro, G., Husmann, M., Korolija, I., & Rovas, D. (2023). SeeQ: A Programming Model for Portable Data-driven Building Applications.

Mischke, J., Barbosa, F., & Parsons, M. (2017, February). *Reinventing construction: A route to higher productivity* (tech. rep.). McKinsey & Company.

Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A Qualitative Framework for Collecting and Analyzing Data in Focus Group Research [Publisher: SAGE Publications Inc]. *International Journal of Qualitative Methods*, *8*(3), 1–21. https://doi.org/10.1177/160940690900800301

Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., & García-Castro, R. (2022). LOT: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence*, *111*, 104755. https://doi.org/10.1016/j.engappai.2022.104755

Röck, 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

Shaw, C., de Andrade Pereira, F., Hoare, C., Farghaly, K., Hartmann, T., & O'Donnell, J. (2024a). Life cycle cost analysis at scale: A reference architecture-based approach [Emerald (Accepted)]. *Construction Industry as a Net-Zero Enabler: Driving Circular Economy and Sustainability through Innovation and Change Management, special issue of Built Environment Project and Asset Management*.

Shaw, C., de Andrade Pereira, F., McNally, C., Farghaly, K., Hartmann, T., & O'Donnell, J. (2023). Information management in the facilities domain: Investigating practitioner priorities. *Facilities*, *Vol. 41*(No. 5/6), pp. 285–305. https://doi.org/https://doi.org/10.1108/F-02-2022-0033

Shaw, C., de Andrade Pereira, F., & O'Donnell, J. (2024b). Multi-stakeholder information requirements to support life cycle asset management.

W3C. (2022, July). Linked Building Data Community Group. Retrieved October 19, 2023, from https://www.w3.org/community/lbd/

Wang, Z., Yeung, T., Sacks, R., & Su, Z. (2021). Room Type Classification for Semantic Enrichment of Building Information Modeling Using Graph Neural Networks, 9.

Yalcinkaya, M., & Singh, V. (2018). VisualCOBie for Facilities Management: A BIM integrated, Visual Search and Information Management Platform for COBie Extension. *Facilities*. https://doi.org/10.1108/F-01-2018-0011