JÖNKÖPING UNIVERSITY
School of Engineering

Lifting Heterogeneous Data into RDF/OWL Knowledge Graph: A Semantic Approach

Mohamad Monir Taktak
Nischal Bhattarai

MASTER THESIS 2024
Master in Built Environment with a specialization SUSTAINABLE BUILDING INFORMATION MANAGEMENT

This thesis work has been carried out at the School of Engineering in Jönköping in the subject area Built Environment - Sustainable Building Information Management. The work is a part of the Master of Science with a major Built Environment, specializing in Sustainable Building Information Management.

The authors take full responsibility for opinions, conclusions and findings presented.

Examiner: Peter Johansson

Supervisor: Rahel Kebede

Scope: 30 credits (second cycle)

Date: 16/05/2025

ADDRESS: School of Engineering, P.O Box 1026, SE-551 11 Jönköping, Sweden
VISIT: Gjuterigatan 5, Campus, Building E

PHONE: +46 (0)36 10 10 00
WEB: www.ju.se

Lifting heterogeneous Data into RDF/OWL Knowledge Graph: A Semantic Approach

Mohamad Monir Taktak, tamo23uk@student.ju.se / monertaktak@gmail.com
Jönköping University, Jönköping 551 11, Sweden

Nischal Bhattarai, bhni23vb@student.ju.se / nisbhswe@gmail.com
Jönköping University, Jönköping 551 11, Sweden

Abstract

Interoperability challenge in Architecture, Engineering and Construction (AEC) industry is often hindered by heterogeneous data which affects effective information exchange. This study investigates a semantic approach to lifting heterogeneous data to improve interoperability, automate validation and enable semantic reasoning. This project employed two methods, both the systematic literature review and use case focusing on light fixture where Level of Information Need (LOIN) was modelled reusing modular ontologies and validated through SHACL shape constraints. The use case depicts the integration of semantic web technologies in information exchange process. The results illustrated that the application of semantic web technologies and linked data principles enhances the exchange workflow where Shapes Constraint Language (SHACL) offers flexibility, logical reasoning and contextual checks during validation. Despite challenges in IFC to RDF transformation, manual enrichment and reuse ensured complete compliance. The results suggests that integration of semantic web technologies within BIM workflows supports semantic interoperability, data interoperability, enable flexibility, standard data driven validation, semantic querying and reasoning. Future work could explore real time data integration, automated semantic lifting and ontology population with Large Language Models.

Keywords

Semantic Web, SHACL Validation, RDF/OWL Knowledge Graph, Linked Data, BIM Interoperability.

1 Introduction

Building Information Modeling (BIM) is a digital method of integrating geometric as well as alphanumeric information into models to support collaboration, coordination and data driven decision making throughout the lifecycle of the built assets (Zhu and Wu, 2022). Regardless of the importance of BIM in the AEC sector, the sector is considered to have low productivity growth and is among the least digitalized industries. One of the current challenges in this sector is the interoperability issues that lead to data loss, misinterpretation, and contradiction (EU BIM Task Group, 2017). The need for enhanced interoperability to integrate various data types from multiple sources across different platforms is gradually increasing in the AEC sector for more efficient projects. Moreover, Dao, Ng and Kwok (2024) identified integrating IFC-BIM with CityGML-GIS as a pivotal and crucial technological pathway for enhancing the construction and management of urban environments. To address the issues of interoperability, different approaches and ways to manage AEC data and realize digital transformation are ongoing (Shen, Sepasgozar and Ostwald, 2024). The development and implication of Semantic Web technologies and Linked Data technology in AEC has gained popularity in the industry as a method to reduce interoperability issues and enhance information exchange (Sadeghineko and Kumar, 2022). According to Shen, Sepasgozar and Ostwald (2024), Semantic web technologies have been applied in the AEC sector to bridge data gaps, facilitate interoperability, and ensure seamless information exchange across various lifecycle stages.

1.1 Problem Statement

One of the standards used in AEC industry for data and information exchange is IFC model across different stakeholders (buildingSMART International, 2025). While IFC provides several advantages and is vendor neutral format, it falls short on the semantic reasoning, cross domain integration, real time integration with external sources such as IoT and sensor, as described in section 2.2. These limitations become evident when verifying model elements such as light fixtures against project specific requirements as the existing method of validation is rigid towards automated reasoning, cross domain integration. While this study investigates lifting heterogeneous data in construction industry into RDF/OWL knowledge graph to improve interoperability, a use case for information exchange using semantic web technologies was carried out, to showcase how semantic web can enhance the workflow. The hypothesis is that application of the semantic web technologies in BIM information exchange improves data consistency, interoperability, automates compliance checking processes, and semantic reasoning and inference.

1.2 Aim

The purpose of this study is to investigate how semantic web technologies such as ontologies and knowledge graphs in AEC industries can improve heterogeneous data interoperability. To achieve this, research questions were developed to gather knowledge about semantic web technologies, existing challenges and insights for lifting heterogeneous data into RDF.

RQ1. How can heterogeneous data be transformed into an RDF/OWL knowledge graph to improve interoperability?
RQ2. What are the challenges in lifting heterogeneous data to RDF/OWL and how can they be addressed?

2 Conceptual Background

2.1 BIM and Interoperability Challenges

Bucher et al., (2024) highlighted different data management technologies used in BIM namely, closed schema, open schema, open-source frameworks, Information Containers, and Traditional software. Closed schema are proprietary formats such as Revit, and Tekla that rely on closed data schema. Even though it provides standardization and consistency, it often leads to data silos and vendor lock-in issues. Moreover, the schema rigidity complicates the version control and hence conflict arises when the files must be merged across different disciplines. Studies and developments were made to address the problem of closed schema with open schema by introducing IFC. IFC was introduced by buildingSMART as a vendor neutral, standard for digital representation of built environment (buildingSMART International, 2025).

The integration of IFC based BIM models with GIS has gained attention as a means to support urban planning and infrastructure management (Lam et al., 2024). IFC also has been instrumental in enabling seamless data exchange in digital twins and smart city applications as an important tool in built environment (Lam et al., 2024). IFC is widely adopted and continues to serve as standard data model for the AEC industry; however, emerging demands for real time connectivity and data integration present new challenges. Specifically, its complex file-based schema and reliance on advanced STEP modeling techniques can hinder seamless integration with modern technologies such as IoT systems and Digital Twins (buildingSMART International, 2020). One of the drawbacks is that it lacks flexibility in representing unstructured data (Costa and Sicilia, 2020). Also, there is not much advancement in reasoning technique for IFC based workflow which limits its ability to support decision making and knowledge inferencing (Huitzil et al., 2024). To achieve interoperability in data exchange and address its limitation, this study investigates how transforming BIM data into RDF knowledge graph could be a potential solution.

2.2 Semantic Web Technologies

Adoption of Semantic web techniques such as ontology based modeling and linked data can help achieve interoperable framework for handling BIM data (Shen, Sepasgozar and Ostwald, 2024). The transformation of IFC models into RDFs graph can support better data integration, and querying capabilities (Costa and Sicilia, 2020). Semantic web technologies were pushed forward to standardize the generic modeling and represent information in a decentralized as well as uniform manner (Borrmann Markus König Christian Koch Jakob Beetz Eds, 2018). The technology allows data to be stored on the web itself and build and write rules for handling these web-based file exchanges (W3C, 2025). Lam et al., (2024) conducted a study to address the interoperability gap between IFC & GIS data using semantic technology and identified improved interoperability, complex querying with RDF/OWL graphs, and enhanced data integration. Similarly, Sadeghineko and Kumar (2022) devised a framework to improve information exchange in existing buildings using semantic web technologies. Similarly, Khudhair, Li and Ren, (2023) developed automatic data exchange frameworks to incorporate data from the BIM model using semantic web technologies and eliminate fragmented data exchanges, lack of holistic decision-making, and manual data processing. It helped improve data interoperability using ontologies & RDF data, develop multi-objective knowledge base automating design analysis, and enable machine-readable BIM data exchange with automated reasoning capabilities. Several domain specific ontologies have been developed. For example, to standardize and structure product properties in accordance with EN ISO 23386:2020 and CEN/TS 17623:2021, Building Product Property Ontology (BPPO[1]), Lighting Product Property Ontology (LPPO[2]) were developed (Tan *et al.*, 2024). Similarly, Filardo et al. (2024) developed standard based aligned ontology called Data Template Ontology (TempO[3]) by aligning it with existing ontologies such as LOIN[4], IsoPropos[5], as well as frameworks like Data Template and properties from building codes. These ontologies ontologies are leveraged in this study to transform heterogeneous data into semantic data for better information exchange workflow in BIM process.

2.3 Information Requirements & LOIN

The information requirements and LOIN has significant role in this project as this project demonstrates how information requirement can be semantically modelled as a LOIN and used in the information exchange workflow through a case study. According to ISO 19650-2, information requirements form basis for the development of Exchange Information Requirements (EIR), which are crucial in setting expectations for the information to be delivered at various project stages (ISO, 2019). Well-structured information requirements enable a robust foundation for validation process and enable facilitation of structured data exchange across the teams (UK BIM Framework, 2021). LOIN is defined as framework which defines the extent and granularity of information. It ensures the right amount and quality of information is delivered at the right time throughout the lifecycle of built asset (CEN, 2020). According to Filardo et al. (2024), LOIN provides the description of properties and enhances digital workflows by providing accurate definition of what information is needed and how it should be structured, validated and exchanged between stakeholders as well as between the system. Tan et al (2024) also emphasized that LOIN plays a pivotal role in achieving semantic interoperability across different platforms by enabling shared understanding of data requirements and ensuring that BIM data is suitable for machine reasoning and automation.

2.4 Information Delivery Specification (IDS)

The buildingSMART defines IDS as a structured approach to specifying the exchange of information in BIM workflows, ensuring that the right data is delivered to the right stakeholder, in the correct format, and at the appropriate time. According to buildingSMART International, it provides "a standardized way to define information requirements in a machine-readable format, which can be validated against IFC models" to ensure consistency and clarity in data exchange processes. In this project, IDS is utilized to ensure that specification requested is compiled by the architect. It is particularly useful in supporting openBIM workflows, where interoperability between software platforms and project participants is required. IDS serves as a connection between the information need and the information provided which enables automatic checking of whether a digital model meets defined information requirements (Filardo *et al.*, 2024). By clearly outlining what information is needed for each object, property, or classification in a model, IDS reduces ambiguity and increases accountability in digital construction processes (Tan *et al.*, 2024).

2.5 SHACL for Semantic Validation

SHACL is a standard that allows RDF graph validation through constraints and rules. SHACL constraints are described with triples to create shapes and identify invalid data (W3C, 2017). It supports logical consistency and semantic reasoning while providing a highly customizable way to define complex validation logic beyond what is possible through syntactic checks, supporting use cases such as performance-based filtering, rule chaining, and compliance diagnostics (Yankulov, 2025). SHACL in this project allows semantic validation of the information requirements and thus helps to maintain data quality.

3 Methods

This thesis employed both systematic literature review to synthesize existing knowledge on semantic technologies, RDF/OWL knowledge graph, information exchange & requirements and presented a practical use case to demonstrate a workflow.

[1] http://www.w3id.org/ppon/bppo#
[2] http://www.w3id.org/ppon/lppo#
[3] https://w3id.org/tempo#
[4] https://w3id.org/loin#
[5] https://w3id.org/isoprops#

3.1 Literature Review

The objective of the literature review was to identify the existing methods, knowledge gaps, and tools related to the contribution of semantic web technologies in transforming data into RDF/OWL knowledge graph. It also contextualizes the study within a broader field while eliminating the probability of working on the existing solutions and studies forming a base for the design of case study. As shown in the Fig 1, a total of 1615 papers were initially added for screening where paper selection criteria were used and papers that discusses transformation of heterogeneous data, improvements and interoperability through semantic web technologies and linked data principles, were retrieved. For instance, to gather knowledge about interoperability challenges within BIM, GIS, Product Data for Information Exchange, keywords “("Interoperability Issues" OR "Interoperability Challenges") AND ("BIM" OR "GIS" OR "Product Data") AND ("Information Exchange" OR "Data Exchange")” was queried from scientific database “Scopus” for Language as “English” with focus on “Subject Area: Energy, Engineering” and included only the papers ranging from 2020-2025.

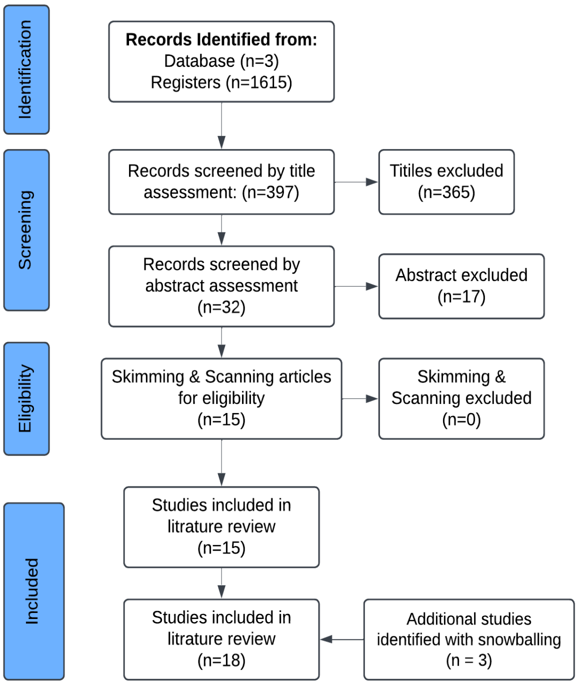


Fig 1. Paper Selection Process (PRISMA flow diagram)

Due to the criteria for producing this paper and page limit, a complete description to the strategies was not included in this report, however, it has been made publicly available[7]. Out of 1615 papers selected for screening, only first 100 papers per query were reviewed based on titles, relevance to the research questions to manage the volume of search results effectively and ensure a focused and practical screening process. Total of 397 papers were added for further screening where after reviewing the abstract, only 32 papers were short listed. Among this, 17 papers were excluded for not meeting the inclusion criteria leaving 15 papers for further study. In addition to that, 3 relevant papers to this study were included by looking at the references cited in a selected paper, making it total of 18 papers.

3.2 Case Study:

To demonstrate the applicability of semantic web technologies for improving information exchange in BIM, a case study was carried out by observing the knowledge gap identified through literature review study. The emphasis was placed on mapping light fixture information requirements to RDF/OWL format for semantic validation and comparison against conventional validation techniques. This involved transforming information requirements into RDF/OWL format for semantic validation using a light fixture as the case study.

Context and Problem Statement: In context of BIM, it is crucial that the data modelled complies with the standard (Filardo *et al.*, 2024). Using LOIN, IDS information requirements are handled during different data drops to facilitate data driven decision making. The use of modular ontologies and RDF/OWL knowledge graph in lifting the validation of these compliance check hasn’t been much explored even though various tools are utilized to transform the data into semantic data (Filardo *et al.*, 2024). This structured approach of semantic validation was aimed by leveraging existing ontologies such as BPPO, LPPO, LOIN, TempO, data dictionaries to support transparent, machine readable compliance of light fixture requirements within the workflow. As discussed in section 1.1, there is a need to integrate semantic web technologies, therefore a light fixture use case was presented to analyses how information requirements such as LOIN can be lifted into RDF.

Objectives: The objective of the case study is to lift the heterogeneous data into RDF knowledge graph utilizing existing ontologies. To achieve the objectives of the study, a structured workflow was devised. as shown in **Fig 2**. It involved modelling information requirements as LOIN in RDF format and validating the requirements using SHACL. The study conducted aimed to investigate how the information requirement could be lifted into RDF/owl knowledge graph for the semantic validation of data through a use case.

Case Study Design:

The case study used a real-world example: the *Notor 65 Beta Opti* light fixture within an office space BIM model. The following ontologies and data sources were employed: BPPO, LPPO, TempO, and LOIN. The tools and technologies applied are summarized in Table 1.

Table 1. Case Study Parameters

Parameter	Description
Use case context	Office room lighting simulation
Object	Notor 65 Beta Opti light fixture
BIM Model	HUS G IFC4ARCH.ifc
Tools used	Revit, ifcOpenShell, Protégé, Python, Cobuilder, GraphDB
Ontologies used	BPPO, LPPO, TempO, LOIN
Validation methods	IDS (Cobuilder + Python), SHACL (Protégé)

The workflow provides an overview of the process and tasks carried out. The process includes two parts where the “left” part includes light simulation, data enrichment, and validation of data against IDS. While the “right” part includes create LOIN, validation of data against SHACL, and store RDF files in GraphDB database. The process begins with customizing a LOIN ontology for the specific use case addressed in this study, light fixture in an office room, by reusing existing ontologies and data dictionaries. Prior to modeling of LOIN in RDF form, all the necessary structure of LOIN was defined according to ISO 19650-1 such as geometric information, alphanumeric information, documentation. Modeling of LOIN can be done in protégé or in python.

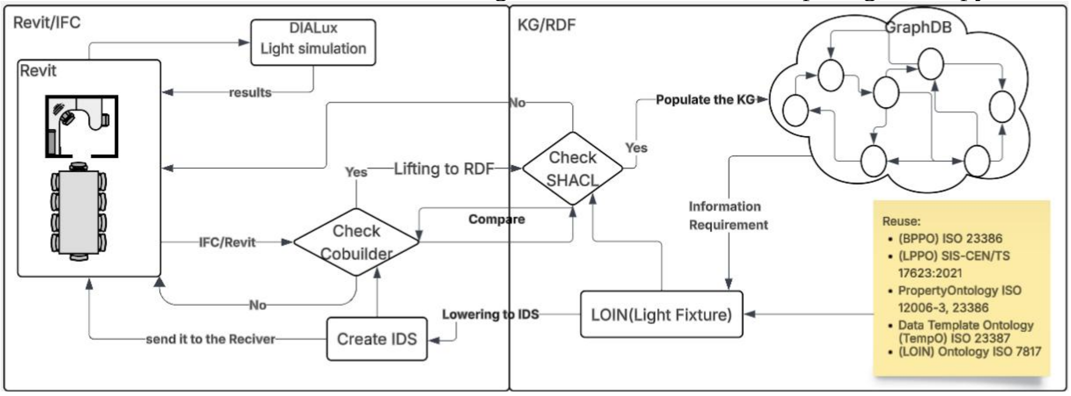
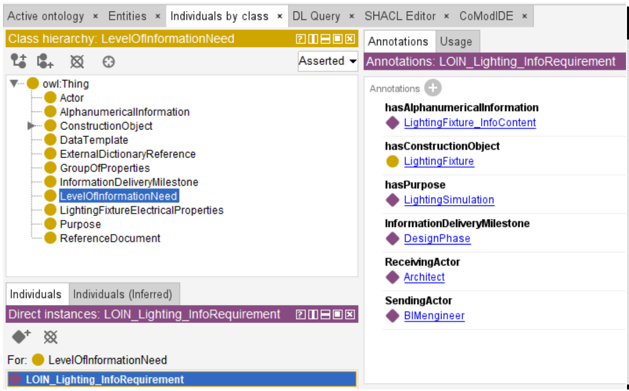


Fig 2. Illustration of process flow of the design

For the left part, the LOIN is used to create an IDS that formally describes the required attributes and data formats to be delivered in machine readable format. The IDS is submitted to the data provider (Architect and/or Light Designer) who then delivers the specified data. Based on information requirements shared from BIM Engineer outputs are first validated against the IDS in Co-builder, ensuring the data meets the predefined specifications. For the missing requirements, the ifc model is enriched with the required specifications by relevant actors and ensured compliance with the requirements. In the right part, the enriched model is thus lifted into RDF format using ifcOpenShell. The ifc data that was converted into RDF is checked against LOIN using SHACL. This semantic check confirms the logical consistency and completeness of the information within the graph. The outcomes of both IDS and SHACL are validated, and the KG is finally populated into GraphDB.

Implementation of the workflow: This section presents how the workflow was implemented in this project. Since the use case is concerned with LOIN for light fixture, Notor 65 Beta Opti was used in this workflow. The first step involved reusing existing ontologies and data dictionaries such as BPPO, LPPO, TempO and LOIN to model information requirement in the form of owl. prerequisites of LOIN, "Design Phase" for Information Delivery Milestone, "Light Simulation" for purpose, "Architect" for Receiving Actor, "BIM Engineer" for sending actor were used. To develop this LOIN, python and Protégé were used by reusing ontologies and dictionaries as shown in the Fig 3.



Property name*	Property set
illuminanceRange	Pset_FunctionalReq_PropertySet
GlareIndexUGR	Pset_FunctionalReq_PropertySet
SimulatedIlluminance	Pset_SimulationReq_PropertySet
DaylightAutonomy	Pset_SimulationReq_PropertySet
AnnualSunlightExposure	Pset_SimulationReq_PropertySet
luminousFlux	Pset_ProductReq_PropertySet
powerConsumption	Pset_ProductReq_PropertySet
mountingType	Pset_ProductReq_PropertySet
correlatedColorTemperature	Pset_ProductReq_PropertySet
colorRenderingIndex	Pset_ProductReq_PropertySet
averageRatedLifeOfLightSource	Pset_ProductReq_PropertySet

Fig 3. Modeling of LOIN in protégé (left) and associated alphanumerical information (right)

Alphanumeric Information such as Functional Requirements, Simulation Requirements, Product Requirements with their associated properties, granularity and semantic details for light product at design stage for Light Simulation were used to model the LOIN.

For the comparison of the effectiveness of data validation between IDS "Traditional BIM domain" and semantic technique using SHACL, LOIN was also created in "Co-builder Require[6]" using the excel sheet template provided where "IFCLightFixture" was added as IFC 4.3 Entity type. Thus, created LOIN was used to generate IDS which formally described the required attributes and data formats to be delivered. A building 3D model "HUS G IFC4ARCH" in ifc format was checked against the IDS in Cobuilder IDS check. The missing property set, and other data suggested by IDS could either be modelled in BIM authoring tool such as Revit or by using Co-builder. Co-builder also supports enrichment of ifc file by using the IDS. For ease, Co-builder was used to enrich the missing property and data sets. After enriching, data were populated for the property sets that was defined such as *Pset_FunctionalReq_PropertySet*, *Pset_ProductReq_PropertySet* in IDS. The additional property set "*Pset_ProductReq_PropertySet"* required light simulation to be conducted, so the IFC model was used to conducted light simulation using DIALux evo for remaining properties (*SimulatedIlluminance*, *DaylightAutonomy*, *AnnualSunLightExposure).* The values were extracted, added in the model and were validated against the IDS again, ensuring the data meets the predefined specifications.

The final ifc file that was populated with data, was converted into RDF patterns using ifcOpenShell. It should be noted that ifc RDF file was directly translated into standard IFC concepts which preserved basic properties only. Therefore, RDF model was manually enriched with modular ontologies relevant to light fixtures. The converted ifc model into RDF was validated against the LOIN in RDF format in protégé using rule-based validation, SHACL. This semantic check confirms the logical consistency and completeness of the information within the graph. The outcomes of both the IDS check and the RDF/Knowledge Graph reasoning are then compared. Since both validations were successful, the validation comparison was done against SHACL and IDS. If discrepancies are observed, the process can be looped back, allowing for iterative correction and revalidation and if the data is formally accepted, it can be stored in the Graph DB. This case study demonstrates the practical application of semantic web technologies in enhancing the accuracy, traceability, and

[6] https://cobuilder.com/en/cobuilder-require/

interoperability of product-level data in Information exchange workflows, particularly for performance-driven scenarios such as light.

Evaluation: The evaluation phase of this study focused on validating the practical effectiveness of using existing ontologies and RDF/OWL knowledge graph for information requirement modeling and compliance checking in the AEC sector. The validation of LOIN requirements was twofold:

1. IDS Validation: The IDS was lowered from RDF using an automated Python script and validated using Co-builder's IDS Checker. The data is said to have complied with the IDS if all the required specification is satisfied by the provided data achieving 100% pass rate. If the compliance rate is below 100%, the data fails to meet the requirements.

2. SHACL Validation: The IFC model enriched and converted into RDF format was semantically validated against the requirement stated in LOIN ontology using SHACL in Protégé. The evaluation of SHACL validation is done by observing the number of violations flagged by it against the constraints provided. The constraints are node shapes that define the requirements of LOIN.

In addition to that, more SHACL validation were conducted to illustrate and emphasize the advantage of leveraging semantic data, reasoning and inference.

4 Results

4.1 Literature Review

R.Q1: How can heterogeneous data be transformed into an RDF/OWL knowledge graph to improve interoperability?

Data such as IFC models can be transformed into RDF graphs using IfcOpenShell or IFCtoRDF tools. For example, Kebede et al. (2022) utilized product information datasets in Excel format to transform into RDF while Wu *et al.* (2022) transformed CSV energy data into RDF using Python script and RDFLib. Similarly, Lam et al., (2024) transformed building components such as doors, windows, walls and CityGML LOD4 model into RDFs. Wang and Zhao, (2024) used LLMs to extract text data from interviews, literature sources such as books, papers, databases, websites. GIS data such as CityGML or JSON based geospatial datasets can be mapped to RDF using R2RML framework or ETL tools like FME to streamline the integration of semantic properties into knowledge graph as Hansson, (2024) and Lam et al. (2024) performed in their study. CityGML 2.0 for urban planning or BOT for spatial relationship provides foundational vocabularies that standardize terminology use across datasets as seen in the study conducted by Sadeghineko and Kumar, (2022) to improve information exchange in existing building using semantic web technology. In addition to that, Kebede et al., (2022, 2024) demonstrated how using modular ontology enhances flexibility by breaking ontologies into reuseable components. The conversion of raw data into RDF/OWL structures can be automated using semantic lifting and mapping techniques. Script based tools such as SPARQL-Generate enables transformation of non-RDF formats such as XML, CSV into triples without intermediate steps (Costa and Sicilia, 2020), while rule-based reasoning with Semantic Web Rule Language or SQWRL infers implicit relationship such as connecting indoor and outdoor pedestrian network in urban navigation systems (Dao, Ng and Kwok, 2024). Validation and storage are other crucial parts to ensure the reliability and accessibility of the knowledge graph. The use of automated ontology reasoners such as HermiT, Pellet, TrOWL can be used to infer relationship and validate the data while eliminating logical conflicts (Huitzil et al., 2024). The use of LLMs also adds automated computation especially in ontology population and reducing manual errors. For example, zero shot prompting with Claude or GPT-4 extracts entities and relationships from unstructured texts, while hybrid approaches like KOnPoTe combined LLM output with semantic rules to reduce inaccuracies (Sahbi, Alec and Beust, 2025).

R.Q2: What are the challenges in lifting heterogeneous data to RDF/OWL using semantic technologies and how can they be addressed?

Despite the advancements in semantic technologies in AEC sector, literature study identified several challenges in lifting heterogeneous data to RDF/OWL knowledge graph. One of the significant challenges is ontology alignment and standardization due to the use of diverse schema that makes integration difficult (Wu et al., 2022). Moreover, semantic interoperability exists due to heterogenous source of data with varied formats such as CSV, JSON, XML, that requires complex preprocessing before transforming it into RDF. Data quality and completeness also are underlying challenges as heterogeneous data often are incomplete, inconsistent, redundant that affects the quality of RDF graphs. Another challenge is the complexity in integrating diverse data source, managing evolving, noisy data and handling scalability (Ibrahim et al., 2024). Ontology alignment techniques and schema mapping can reduce these issues. Knowledge graphs require continuous updates to reflect real world changes but having noisy or redundant data impacts its accuracy (Kommineni, König-Ries and Samuel, 2024). There is high cost associated with the transformation of large-scale heterogeneous datasets that impacts querying efficiency as well (Kirstein et al., 2020). Furthermore, transforming large scale dataset into RDF is computationally intensive as it must query and process large set of graphs. For example, IFC model contains millions of entities which leads to slow RDF conversion and querying performance (Huitzil et al., 2024).There is no universally accepted standards for BIM-GIS integration even though ontologies such as ifcOWL, BOT, PRODUCT, PROPOS exist (Huitzil et al., 2024). To address these issues, reusing ontology and expanding existing ontologies are emphasized. On top of that, the use of automated data conversion tools such as RDFlib, R2RML mapping etc. are recommended to be used to streamline the process and reduce the manual errors (Hansson, 2024). Studies also suggest optimizing pipelines to include only the relevant information to minimize unnecessary complexity. The use of ontology mapping techniques and rule-based inference can also be used to align heterogeneous data sources while preserving semantic meaning (Wang and Zhao, 2024). For BIM-GIS integration, the use of hybrid approach combining semantic reasoning with external spatial processing tools can improve geometric data while preserving critical spatial relationship as well (Dao, Ng and Kwok, 2024). Implementation of Retrieval Augmented Generation (RAG), LLMs can be used to update the knowledge graph and maintain accurate representation. Another challenge of integrating multilingual data sources while maintaining semantic consistency can be addressed using multilingual ontologies and cross lingual entity linking techniques (Ibrahim et al., 2024). The problem (Khudhair, Li and Ren, 2023) of dealing with real time data streams for continuous RDF conversion and update can be solved by using automated ontology engineering using LLMs (Shimizu and Hitzler, 2025).

4.2 Case Study

a. A complete validation of LOIN requirements from ifc RDF model.

A Total of four light fixture of same type *Notor 65 Beta Opti* were instantiated in the model where each of them were required with 11 properties, totalling 44 checks. Validation of ifc model with IDS in Cobuilder resulted 44 out of 44 checks passed implying 100% compliance rate. However, when the ifc model was converted into RDF graphs using IfcOpenShell, and validated using SHACL constraints, 44 violations were reported. Upon investigation, the ifc RDF graph contained generic entities and relationship but lacked semantic connection for light fixture triples with its corresponding values. The missing property values were enriched to ifc RDF graph manually according to LOIN RDF defined in previous stages. After enrichment, the validation of ifc RDF graph highlighted 0 violation for total of 44 requirements.

b. Validation of Light Fixture with Luminous Flux greater than 3000.

To validate that a light fixture must have a luminous flux greater than 3000, a SHACL NodeShape named ex:HighFluxFixtureShape was created and targeted to instances of ifc:IfcLightFixture. Within this shape, a property constraint was defined using sh:property on the path lppo:luminousFlux, specifying all values must have datatype as "xsd:float" and luminous flux > 3000 enforced via sh:miinInclusive.

During validation, one of the four IfcLightFixture instances in the RDF graph had a luminous flux value of 2500, which violated the defined constraint. As a result, a SHACL validation report flagged this instance with a message indicating the violation of the required minimum luminous flux.

c. Validation of all the Light Fixtures has averageRatedLifeOfLightSource greater than 40,000 and an illuminanceRange less than 350, both as xsd:float values.

Here, *ex:LightFixtureAdvancedValidationShape* is added as *NodeShape* and was targeted to all the instance of light fixtures. For float datatype, the *averageRatedLifeOfLightSource* should be greater than 40000 as constrained by *sh:minExclusive.* In this case, a total of three violations were raised where two of the instances had illuminance Range more than 350 and *averageRatedLifeOfLightSource* was more than 40000.

d. Validation which ifc:IfcLightFixture instances have mounting type Ceiling and illuminanceRange must be less than 350.

This SHACL code validates which light fixture instances in this RDF model has mounting type “Ceiling” and illuminance Range less than 350 lux. Here, a Node shape targeting instances of IfcLightFixture is added and two property constraints were applied where lppo:mountingType stated that property should be “Ceiling” while lppo:illuminanceRange must be less than 350. Only one instance “LightFixture03” had illuminanceRange 300 lux in the ifc RDF model and mounting type Ceiling at the same time.

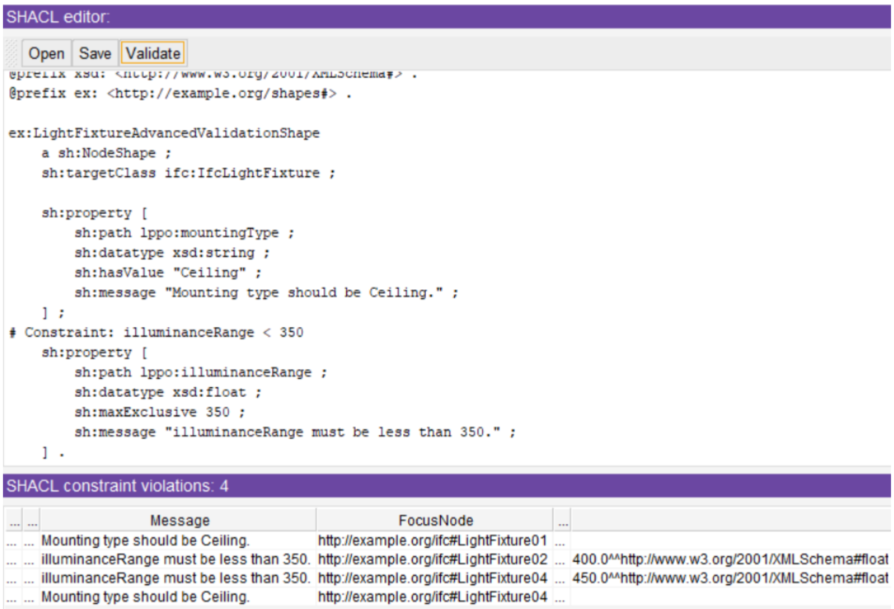


Fig 6. This SHACL shape validates each Light Fixture with mounting type Ceiling and Illuminance Range less than 350.

e. Validation of Ceiling Light Fixture in office space only.

Due to the limitation of data, complex validation wasn’t possible for this ifc-RDF file, however, more complex validation can be done using SHACL. It can also be extended for complex data validation also. For an instance, SHACL can be used to validate if all light fixture located in space labelled as “office” complies with requirements such as mounting type “Ceiling”, Illuminance Range >350 and Energy Class “A or higher”.

5 Discussion

This study revolved around the hypothesis that utilizing semantic web technologies and linked data principles in AEC domain significantly enhances data interoperability, enables semantic validation, and information traceability across the workflow. The findings the use case validates the insights that were gathered from the literature review. The use of semantic web technologies and linked data principles illustrated flexible, rule based validation mechanism than traditional technologies.

Literature suggested that even though IFC is widely used across BIM information exchange workflow and extends several use cases, it offers limited support for semantic inference, real time cross domain integration (Huitzil et al., 2024; Lam et al., 2024). This was seen when the intial RDF transformation using IfcOpenShell preserved structure but lacked to semantic connections. It was seen during the validation of ifc RDF model where model initially failed due to generic property connection highlight the importance of semantic binding between entities and their properties as pointed out by Wu et al., (2022) and Khudhair, Li and Ren, (2023). After enriching ifc RDF model with URIs and properties as described in LOIN, all 44 constraints passed. This demonstrated the need of semantic integrity for complete and consistent data exchange along the workflow. From the results, it was observed that the semantic validation such as SHACL performs well in terms of flexibility and depth as compared to IDS. For example, some of the validation required that threshold of luminous flux for light fixture be 3000, which was beyond just validating the presence of values. Similarly, some validations as described in the results show how SHACL can perform multiple property logic validation and semantic inference. This emphasized the strenght of using SHACL for logical reasoning, enabling constraint chaining and dependency based checks during design and validations as claimed by Yankulov, (2025). In addition to that, conceptualized validation which leverages rule driven, context aware data checks that can transverse through spatial and functional hierarchies.

Modeling of LOIN for light fixture reusing modular ontologies such as LPPO, BPPO, TEMPO, LOIN and data dictionaries in the case study provided flexibility and reusability across the domain which was also pointed out by Kebede et al., (2022). The LOIN based RDF graph allowed structured information requirements to be validated symantically using SHACL. This suggested that machine readable LOIN can be effectively implemented in information exchange workflow. In contrast to the advantages, challenges were also encountered similar to as described in the literatures, especially the conversion of IFCtoRDF, loss of semantics and meaningful link (Ibrahim et al., 2024). The transformation of IFC was achieved with IfcOpenShell which required enrichment of data manually. This agreed with the fact that current tools fall short on seamless mapping of data into RDF graphs. Manual enrichment and use of standard ontologies were a possible solutions which was also observed by .Hansson, (2024) and Lam et al., (2024). Even though the IFC model used for this use case consisted of architectural elements only, the conversion of this ifc to RDF resulted several triples. For a model that mimics the real world, the transformation could result in scalability issues as noted by Huitzil et al., (2024). It should be noted that this study managed the scalability issue by focusing on light fixtures only.

6 Conclusions

The investigation of transforming hterogeneous data into RDF/OWL knowledge graphs demonstrated a semantic approach which overcomes interoperability challenges often encountered during information exchange process. A use case connected to light fixture was impemented and its information requirements was modeled using LOIN. This study reused existing ontologies and data dictionaries to enable structured and machine readable data and its validation through SHACL. It was then investigated how these ontologies and RDF/OWL KG play a crucial role in ensuring better information exchange workflows. From the results, it can be acknowledged that while IDS validates the presence and structure of required information within the model, SHACL provides deeper insights by suppporting logical constraints, contextual checks and thresholds. It also emphasized the advantage of RDF and SHACL to ensure that semantic relationships is maintained across the data. On the other hand, it also revealed the limitations of existing tools which failed to seamlessly align and connect entities with their associated properties with correct relationship while converting ifc to RDF witout human intervention.

In conclusion, the study demonstrated that lifting heterogeneous data into RDF/OWL formats and integrating it within BIM based information exchange workflow improves data consistency, automated compliance checks, flexible validation and itneroperability across different domains. Lifting of Information requirements such as LOIN and its validation using SHACL highlight potential for BIM workflow. While several limitations were present, it should nevertheless noted that semantic approach lays foundation for advancing data driven, intelligent, and interoperable construction practices.

While this study presented a promising semantic workflow for information exchange in BIM using RDF/OWL and validation using SHACL, future work can further enhance integration of semantic web principles across workflow. The workflow used in this work relied on intermediate steps and traditional tools. Future work could re-engineer the workflow such as automated semantic lifting of pipelines, creation of SHACL compliant data with authoring tools. Furthermore, automating the enrichment of IFCtoRDF transformation with reasoning capabilities. In addition, this study used static data to illustrate the integration in the workflow, however, future work can investigate how dynamic data can be be used. Moreover, the use LLMs can also be investigate for dynamic ontology population.

7 Data Availability

All project files including the thesis manuscript datasets, ontologies (BPPO, LPPO, LOIN), SHACL validation scripts, and workflow diagrams are publicly available on GitHub[7].

[7] https://github.com/tamo23uk/LOIN-for-Lighting-Simulation

8 References

buildingSMART International, 2022. *Information Delivery Specification (IDS) Technical Documentation.* [Online]
Available at: https://www.buildingsmart.org/standards/bsi-standards/information-delivery-specification-ids/
[Accessed 12 December 2024].

buildingSMART International, 2022. *Methods to specify information requirements in digital construction projects.* [Online]
Available at: https://www.buildingsmart.org/methods-to-specify-information-requirements-in-digital-construction-projects/
[Accessed 06 May 2025].

buildingSMART International, 2025. *Industry Foundation Classes (IFC).* [Online]
Available at: https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/
[Accessed 05 May 2025].

buildingSMART International, 2025. *Industry Foundation Classes (IFC) – An Introduction.* [Online]
Available at: https://technical.buildingsmart.org/standards/ifc/
[Accessed 12 March 2025].

CEN, 2020. *EN 17412-1:2020 – BIM – Level of Information Need – Part 1: Concepts and principles,* Brussels: European Committee for Standardization.

EU BIM Task Group, 2017. *Handbook for the Introduction of Building Information Modelling by the European Public Sector: Strategic Action for Construction Sector Performance - Driving Value, Innovation and Growth..* [Online]
Available at: https://www.eubim.eu/wp-content/uploads/2017/07/EUBIM_Handbook_Web_Optimized-1.pdf
[Accessed 29 January 2025].

ISO, 2019. *BS EN ISO 19650-2:2019- Organization and digitization of information about buildings and civil engineering works, including building information modeling- Information management using building information modeling - Part 2: Delivery phase of the assets.* 1 ed. s.l.:British Standards Institution.

ISOb, 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.* 1 ed. Stockholm: British Standards Institution.

Mordor Intelligence, 2023. *(2023) Building Information Modeling (BIM) Market - Growth, Trends, COVID-19 Impact, and Forecasts (2025 - 2030)..* [Online]
Available at: https://www.mordorintelligence.com/industry-reports/building-information-modelling-market
[Accessed 28 January 2025].

UK BIM Framework, 2021. Guidance Part D Developing Information Requirements. In: *Information Management according to BS EN ISO 19650.* s.l.:UK BIM Framework.

W3C, 2017. *Shapes Constraint Language (SHACL).* [Online]
Available at: https://www.w3.org/TR/shacl/
[Accessed 14 May 2025].

W3C, 2025. *Semantic Web Standards.* [Online]
Available at: https://www.w3.org/2001/sw/wiki/Main_Page
[Accessed 12 March 2025].

Yankulov, M., 2025. *What is SHACL.* [Online]
Available at: https://www.ontotext.com/knowledgehub/fundamentals/what-is-shacl/
[Accessed 21 April 2025].

Borrmann Markus König Christian Koch Jakob Beetz Eds, A. (2018) *Building Information Modeling Technology Foundations and Industry Practice*. Edited by A. Borrmann et al. Available at: https://doi.org/https://doi.org/10.1007/978-3-319-92862-3.

Bucher, D.F. *et al.* (2024) 'From BIM to Web3: A critical interpretive synthesis of present and emerging data management approaches in construction informatics', *Advanced Engineering Informatics*, 62. Available at: https://doi.org/10.1016/j.aei.2024.102884.

buildingSMART International (2020) *Technical Roadmap buildingSMART Getting ready for the future*. Available at: https://www.buildingsmart.org/about/technical-roadmap/ (Accessed: 14 May 2025).

Costa, G. and Sicilia, A. (2020) 'Alternatives for facilitating automatic transformation of BIM data using semantic query languages', *Automation in Construction*, 120. Available at: https://doi.org/10.1016/j.autcon.2020.103384.

Dao, J., Ng, S.T. and Kwok, C.Y. (2024) 'Interlinking BIM and GIS data for a semantic pedestrian network and applications in high-density cities', *Developments in the Built Environment*, 17. Available at: https://doi.org/10.1016/j.dibe.2024.100367.

Filardo, M.M. *et al.* (2024) *A standard-based ontology network for information requirements in digital construction projects*. Available at: https://www.researchgate.net/publication/381408124.

Hansson, F. (2024) *Linked geodata: CityGML represented as a virtual knowledge graph*. Lund University.

Huitzil, I. *et al.* (2024) 'Semantic Building Information Modeling: An empirical evaluation of existing tools', *Journal of Industrial Information Integration*, 42. Available at: https://doi.org/10.1016/j.jii.2024.100731.

Ibrahim, N. *et al.* (2024) 'A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): models, evaluation metrics, benchmarks, and challenges', *Discover Artificial Intelligence*. Springer Nature. Available at: https://doi.org/10.1007/s44163-024-00175-8.

Kebede, R. *et al.* (2022) 'Integration of manufacturers' product data in BIM platforms using semantic web technologies', *Automation in Construction*, 144. Available at: https://doi.org/10.1016/j.autcon.2022.104630.

Kebede, R. *et al.* (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. Available at: https://doi.org/10.1016/j.spc.2024.05.007.

Khudhair, A., Li, H. and Ren, G. (2023) 'Knowledge-based OpenBIM data exchange for building design', *Automation in Construction*, 156. Available at: https://doi.org/10.1016/j.autcon.2023.105144.

Kirstein, F. *et al.* (2020) 'Piveau: A Large-scale Open Data Management Platform based on Semantic Web Technologies'. Available at: http://arxiv.org/abs/2005.02614.

Kommineni, V.K., König-Ries, B. and Samuel, S. (2024) 'From human experts to machines: An LLM supported approach to ontology and knowledge graph construction'. Available at: http://arxiv.org/abs/2403.08345.

Lam, P.D. *et al.* (2024) 'Digital Twin Smart City: Integrating IFC and CityGML with Semantic Graph for Advanced 3D City Model Visualization', *Sensors*, 24(12). Available at: https://doi.org/10.3390/s24123761.

Sadeghineko, F. and Kumar, B. (2022) 'Application of semantic Web ontologies for the improvement of information exchange in existing buildings', *Construction Innovation*, 22(3), pp. 444–464. Available at: https://doi.org/10.1108/CI-03-2021-0058.

Sahbi, A., Alec, C. and Beust, P. (2025) 'Semantic vs. LLM-based approach: A case study of KOnPoTe vs. Claude for ontology population from French advertisements', *Data and Knowledge Engineering*, 156. Available at: https://doi.org/10.1016/j.datak.2024.102392.

Shen, X. han, Sepasgozar, S.M.E. and Ostwald, M.J. (2024) 'Knowledge-based semantic web technologies in the AEC sector', *Automation in Construction*. Elsevier B.V. Available at: https://doi.org/10.1016/j.autcon.2024.105686.

Shimizu, C. and Hitzler, P. (2025) 'Accelerating knowledge graph and ontology engineering with large language models', *Journal of Web Semantics*, 85. Available at: https://doi.org/10.1016/j.websem.2025.100862.

Tan, H. *et al.* (2024) *Semantic Interoperability using Ontologies and Standards for Building Product Properties*. Available at: https://www.buildingsmart.org/.

Wang, H. and Zhao, R. (2024) 'Knowledge graph of agricultural engineering technology based on large language model', *Displays*, 85. Available at: https://doi.org/10.1016/j.displa.2024.102820.

Wu, J. *et al.* (2022) 'A semantic web approach to uplift decentralized household energy data', *Sustainable Energy, Grids and Networks*, 32. Available at: https://doi.org/10.1016/j.segan.2022.100891.

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