Abstract
Knowledge of the HVAC system topology is required in several design and operation activities, including setting up building energy simulation models, fault detection and control applications, and facility management tasks. In theory, we can use the geometric context captured in Building Information Models (BIM) to infer these topological relationships; in practice, these models are far from perfect, with modelling errors and omissions. The implication is that manual remodelling is required, which results in unnecessary and redundant work. This paper discusses first the data quality issues faced in real-world data transformation workflows. We propose a methodology for generating good-quality HVAC topological models from imperfect BIM models. Our method’s core is a rule-based model-checking framework to identify missing elements. Once an issue is detected, a geometric-relationship enrichment tool is invoked to infer missing topological information. The proposed approach is tested in a real-world complex BIM model, and we discuss lessons learnt.

Highlights
- Used ontologies of the building domain to represent HVAC system topological information.
- Developed a rule-based model checking mechanism to identify missing relationships.
- Used geometric processing to enrich the models with such information.

Introduction
Information about Heating, Ventilation and Air-Conditioning (HVAC) systems’ geometry, operational characteristics and parameters are required in many design and operation tasks, including design coordination, building energy simulation, systems modelling, fault detection and diagnostics (FDD), and control applications (Andriamamonjy et al. (2018); Wetter et al. (2019)). Building Information Models (BIM) are an information-rich resource, especially concerning the geometric representations of building components. However, diverse user requirements across the building lifecycle, modelling errors, and complex BIM data representation (Pauwels et al. (2021)) might hinder using this information.

The knowledge of the HVAC system topology can be beneficially used in several contexts: in operation, this information is needed when implementing control strategies or configuring FDD and other analytics. This information might also be used to set up energy simulation models so that they can be used to understand energy performance. Information discontinuities occur at various stages in current practice, from design, handover, operation, to retrofitting. This can result in information loss and incurs inefficiencies related to the need to re-create the data. This can be time-consuming, error-prone, and laborious, especially for larger buildings.

By using BIM models, specific tasks can be automated. For example, in setting up energy models, it is possible to extract geometric (and other aspects) from the BIM models and set up energy models. Such an approach can reduce modelling time and costs, facilitate the rapid evaluation of design alternatives, and help the modeller focus on important modelling parameters, which can lead to higher-quality energy models (Bazjanac (2008)). Many efforts have focused so far on transferring passive element information (Gao and Pishdad-Bozorgi (2019)). Studies that attempted to extract HVAC system information from BIM are fewer, with the poor quality of BIM data often cited as a significant issue (Pinheiro et al. (2018)).

Figure 1: BIM issues examples: (a) Missing connectivity relationship between pipe and cooling coil, (b) Missing relationship between space and air terminal
data during the design and construction phases, operational requirements of HVAC systems are often unknown or omitted during the BIM authoring process. For example, in the model of Figure 1a, although the pipe segment and cooling coils appear visibly connected, their relationship is not explicitly modelled. Extracting this relationship would require a manual model inspection to interpret their connectivity. This requirement might be especially specified in Exchange Information Requirements or Modelling Guidelines, yet the BIM modeller might omit them. More importantly, they might not have been requested in the first place and, therefore, not included in the model. Furthermore, current authoring practice can result in fragmented building data generation, where different disciplines (e.g. architectural, mechanical, electrical etc.) generate and store building information in different models, thus reducing data interconnections among them (Ouyang et al. (2023)). In particular, although it is possible to inspect potential clashes between Architectural and MEP elements using the same 3D environment, the underlying semantic relationships between spatial and HVAC components — required to describe how HVAC systems operate — are often missing. For example, in Figure 1b, we can visually infer that the air terminal component feeds air to the space below. Yet, this relationship is not explicitly available between the two components. As a result of such imperfections and BIM’s focus on the geometric description of building elements, topological information remains insufficient for reuse in building energy simulation, systems modelling, or other OpM scenarios (Hauer et al. (2019)). Therefore, extracting and reusing HVAC system topology requires tools to automate retrieving existing information about HVAC components from BIM and mechanisms to enable users to check and enrich them with missing semantic relationships. This work addresses these imperfections to facilitate BIM data reuse for generating good-quality HVAC topological models usable in various use cases across the building lifecycle.

**Ontology-based BIM data exchange to HVAC models**

Semantic web technologies offer an alternative ontology-based paradigm for modelling building data in the form of knowledge graphs based on the Resource Description Framework (RDF) introduced by W3C (2014). Representing building data in RDF graphs has shown potential for cross-linking information across disciplines (Pauwels et al. (2021)). In particular, recent studies have highlighted the potential of using RDF to enrich building data with missing object associations between BIM elements (Ouyang et al. (2023)), leading to semantically richer, and thus more usable, building data for building energy modelling purposes (Panagoulia et al. (2021)). Nonetheless, RDF is not considered an ideal format for storing all types of building data — in particular, RDF’s verbosity can be limiting in geometric representations (Pauwels et al. (2021)). Yet, unlike building envelope’s geometry, detailed geometric descriptions of HVAC components are not required as input to set up a HVAC model. Yet, topological relationships between spatial and HVAC components can be used as input both in energy simulation software such as EnergyPlus (i.e. through HVAC templates) (Pinheiro et al. (2018)) or Modelica (Andriamamonjy et al. (2018)) and data-driven models for control or fault detection applications.

Aiming to separate the representation of geometric and topological information of BIM elements, the Building Topology Ontology (BOT) (Rasmussen et al. (2021)) has been devised as a lightweight, general-purpose ontology that captures basic topological aspects of buildings and can optionally be aligned with complementary ontology-based layers of information, known as Linked Building Data (LBD) ontologies (Pauwels et al. (2021)). In contrast to IFC, this alternative paradigm allows more flexible retrieval of building information through SPARQL queries, omitting the need to produce an IFC model for importing data in building energy models.

To describe how HVAC systems operate within a building, we take a holistic view of HVAC system topology, not only including HVAC components (Ducts, Pipes, Coils, AHU, etc.) but also their spatial distribution, downstream/upstream connectivity, and logical interaction with Spatial components (i.e. how they “serve” building spaces). In particular, we aim to detect two types of relationships between HVAC components (Wang et al. (2022)): (1) topological or geometric-based relations capturing the spatial distribution of HVAC components within a building and (2) logical or upstream/downstream connectivity between HVAC components which describe the operational logic of the systems. We adopt a hierarchical approach in which increasingly more information is required to represent the relationships mentioned above between HVAC and Spatial components — as they organically appear during the modelling process. Leveraging the recent advancements in building data modelling, we propose an ontology-based workflow for extracting HVAC information from BIM data and checking the data quality of HVAC system topology to identify semantic errors on specific HVAC components. We develop the Geometric Relation Checking (GRC) tool and employ a semantic inferencing engine to enrich further the data with the missing topological relationships automatically. Yet, recognising the possibility of inferring “false-positive” relationships, we aim to support the entire workflow with a Graphical User Interface (GUI) to enable users to approve any additional relationships.
The main objectives of this work are to (1) Devise a workflow for generating usable HVAC system topology from BIM data and (2) Demonstrate how the proposed workflow can enable the reuse of BIM data.

Relevant Work

The complex nature of the problem has led to the adoption of graph-based representations of HVAC system topology (Han et al. (2022); Wang et al. (2022)), geometry-processing tools (Xiao et al. (2019)) to infer some of the missing relationships, and rule-based checking mechanisms to ensure good data quality (Wang et al. (2022); Hu et al. (2019)) — yet, these studies present limitations.

Han et al. (2022) introduced a method for generating graph-based HVAC models, highlighting the lack of topological connections between MEP components as a significant limitation. Such missing information issues can be partially addressed through geometry-based processing, according to Xiao et al. (2019), who proposed a method for generating HVAC logic chains using geometry inputs and expert system operation rules. However, their approach is limited in detecting connectivity between adjacent MEP elements, omitting more advanced relationships (e.g. clashes) between ARC and MEP components that can not be fixed without a checking mechanism.

Recognising the need for data quality checking, a few studies have introduced mechanisms for checking topological or logical relationships between MEP components. In particular, Hu et al. (2019) developed a clash detection mechanism to identify errors between intersecting MEP components using geometry. Whereas, Wang et al. (2022) devised a mechanism using Revit’s API and Neo4j platform to extract BIM data, generate a knowledge graph of the MEP systems, and check the existence of logical relationships through queries on the graph database. Despite the benefits of those mechanisms, they remain limited in detecting and guiding users to correct errors manually. Instead, our proposed workflow aims to minimise the dependency on manual user efforts, leveraging BIM geometry and semantic inferencing rules to enrich poor data quality of HVAC system information.

Proposed Workflow

In this section, we present a novel knowledge graph-based workflow which aims to deliver good-quality information on HVAC system topology from BIM data. An overview of the proposed workflow is illustrated in Figure 2 and is based on three main processes:

- **Extract** and transform HVAC system topology from IFC files into a knowledge graph database.
- **Check** information quality of the generated graph using topological and logical rules.
- **Enrich** the graph with missing information using geometry processing and semantic inferencing tools.

As a first step, information about the HVAC system topology (i.e. HVAC, Spatial components and their relationships) is extracted both from architectural (ARC) and MEP IFC models, using an extended version of the Knowledge Graph Generator (KGG) tool developed by Mavrokapnidis et al. (2021). This tool is available online.

This initial version of the knowledge graph (in RDF format) is then validated over pre-defined validation rules that express the required topological and logical relationships among the extracted HVAC and Spatial elements. Through this rule-checking mechanism, we can produce a list of IFC elements that contain errors or, in other words, do not have the required relationships as defined by the validation rules.

After identifying those elements, we enable the graph enrichment with topological relationships. For this purpose, we use the Geometry Relation Checking (GRC) tool, developed as part of this study, to infer automatically missing associations between HVAC and Spatial components. In practice, the GRC tool

---

1. https://kgg.openmetrics.eu/
takes the list of IFC elements with errors as input, identifies geometric relations between other elements and infers additional topological relationships automatically. In addition, a logical inferencing engine is used to infer other logical relationships by implementing pre-defined if-then rules in the knowledge graph. Finally, any further unresolved issues can be resolved through manual inspection of the data and user enrichment with the aid of a Graphical User Interface (GUI).

Knowledge Graph Generation

Unlike the "one-size fits all" BIM approach, LBD ontologies have been devised to represent buildings as independent layers of information in a simple, modular and extensible way (Pauwels et al. (2021)). Embracing these principles, we select to represent HVAC systems using: (1) the Building Topology Ontology (BOT) (Rasmussen et al. (2021)) to capture the topological aspects of building components; (2) the Flow Systems Ontology (FSO) (Kukkonen et al. (2022)) to describe interconnected systems and components with material or energy flow connections and; (3) Brick Schema (Balaji et al. (2018)) to demonstrate the ability of our approach to inform practical applications during the operational phase of buildings.

Table 1: IFC-to-KG sample mappings

<table>
<thead>
<tr>
<th>IFC</th>
<th>bot:</th>
<th>fso:</th>
<th>brick:</th>
</tr>
</thead>
<tbody>
<tr>
<td>DuctSeg.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>DuctFit.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>PipeSeg.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>PipeFit.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>AirTerm.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>FlowTerm.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>Valve</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>Coil</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>UnitaryEq.</td>
<td>bot:Elem.</td>
<td>fso:Elem.</td>
<td>brick:</td>
</tr>
<tr>
<td>Space</td>
<td>bot:Zone</td>
<td>fso:Component</td>
<td>brick:</td>
</tr>
<tr>
<td>Storey</td>
<td>bot:Zone</td>
<td>fso:Component</td>
<td>brick:</td>
</tr>
</tbody>
</table>

We extended KGG, an ETL (Extract-Transform-Load) tool with additional mappings between IFC and BOT, FSO and Brick ontologies to generate the knowledge graph. KGG extracts HVAC and Spatial components from IFC, represented as subclasses of IfcDistributionElement and IfcSpatialElement and transforms them into BOT, FSO and Brick classes and relationships, according to Table 1 and Figure 3, and load them in the knowledge graph database. Detailed mappings between IFC and BOT, FSO, Brick and other ontologies are available online³.

In more detail, Figure 3 illustrates the ontological description of Spatial and HVAC components and their relationships, where topological relationships are captured using BOT and FSO ontologies, as described by


Semantic Rule Checking

When these relationships exist in IFC, KGG can extract, transform, and load them in RDF, yet real-world models lack this information, most often due to modelling imperfections or IFC exportation issues of BIM authoring tools. In this section, we present a rule-checking mechanism that aims to validate the existence of the aforementioned relationships in the generated knowledge graph.

Figure 4: Knowledge Graph validation process
We define and express the rules using the SHApes Constraint Language (SHACL) introduced by W3C (2017) as "a language for validating RDF graphs against a set of conditions". These conditions are expressed as a SHACL shape which is then executed over the RDF graph and produces a validation report including the detected violations. In other words, a SHACL shape includes constraints about the structure and content of the knowledge graph; whenever a node in the RDF graph does not conform with these constraints, a validation error is produced. Figure 4 illustrates an overview of the knowledge graph validation process.

Table 2: Rules for checking Topological relationships

<table>
<thead>
<tr>
<th>Node</th>
<th>Edge</th>
<th>Node</th>
<th>Card.</th>
</tr>
</thead>
<tbody>
<tr>
<td>bot:Zone</td>
<td>bot:hasElem.</td>
<td>bot:Elem.</td>
<td>1</td>
</tr>
<tr>
<td>fso:Segment</td>
<td>fso:con...With</td>
<td>fso:Segm.</td>
<td>1</td>
</tr>
<tr>
<td>fso:Fitting</td>
<td>fso:con...With</td>
<td>fso:Comp.</td>
<td>2...*</td>
</tr>
<tr>
<td>fso:Compon.</td>
<td>fso:con...With</td>
<td>fso:Comp.</td>
<td>2...*</td>
</tr>
</tbody>
</table>

In table 2, we propose a set of topological rules to identify elements (i.e. nodes in the graph) and the number (i.e. cardinality) of missing relationships (i.e. edges). We have defined those rules to reflect expert knowledge about HVAC systems topology. For example, considering that terminals (e.g. Air Diffuser) are the latter elements in a distribution network, we want to validate that every fso:Terminal element is located in exaxtly one (cardinality: "1") bot:Zone. Furthermore, on the grounds that fso:Fitting and fso:Segment are internal distribution elements, they must be connected with two or more (i.e. cardinality: "2...*") fso:Component elements. Through the validation process, we can identify terminals without any connection or segments only connected with one element, thus inferring the existence of a topological error.

Table 3: Rules for checking Logical relationships

<table>
<thead>
<tr>
<th>Node</th>
<th>Edge</th>
<th>Node</th>
<th>Card.</th>
</tr>
</thead>
<tbody>
<tr>
<td>fso:Compon.</td>
<td>fso:feedsFl.</td>
<td>fso:Compon.</td>
<td>1</td>
</tr>
<tr>
<td>fso:Term.</td>
<td>fso:feedsFl.</td>
<td>brick:Loc.</td>
<td>1</td>
</tr>
<tr>
<td>brick:Equip.</td>
<td>brick:feeds</td>
<td>brick:Loc.</td>
<td>1...*</td>
</tr>
</tbody>
</table>

Similarly in Table 3, we propose a set of logical relationships in order to locate elements missing downstream/upstream connectivity with other HVAC or spatial elements. For example, we want to ensure that every brick:Equipment serves at least one or more (i.e. cardinality: "1...") brick:Location/bot:Zone elements. Whenever an instance of brick:Equipment does not serve a building space, the rule-checking mechanism produces a logical error related to this instance. The following listing describes those three examples, as expressed using SHACL. We execute those rules over the RDF graph, using the pySHACL library as validation engine (https://github.com/RDFLib/pySHACL). Through this process, we generate a validation report which provides feedback about specific rule violations on the instances of the RDF graph.

Listing 1: Semantic rule expressions in SHACL

```
@prefix sh: <http://www.w3.org/ns/shacl#>
@prefix fso: <https://www.w3id.org/fso#>
@prefix brick: <https://brickschema.org/Brick#>
@prefix kgg: <http://openmetrics.eu/kgg#>
@prefix bot: <https://w3id.org/bot#>

kgg:rule_1 a sh:NodeShape ;
  sh:targetSubjectsOf bot:Zone ;
  sh:name "Terminal’s location error" ;
  sh:property [ sh:path bot:hasLocation ; sh:class fso:Terminal ; sh:minCount 1 ; sh:maxCount 1 ].

kgg:rule_2 a sh:NodeShape ;
  sh:targetClass fso:Segment ;
  sh:name "Segment connectivity error" ;
  sh:property [ sh:path fso:connectedWith ; sh:class fso:Component ; sh:minCount 2 ].

kgg:rule_3 a sh:NodeShape ;
  sh:targetClass brick:Equipment ;
  sh:name "Equipment not serving space" ;
  sh:property [ sh:path brick:feeds ; sh:class brick:Location ; sh:minCount 1 ].
```

Enrichment through geometry checking

Based on the validation report output, we can identify the HVAC components with missing relationships. We apply a semantic enrichment process using geometry processing to discover these relationships. We have developed the Geometric Relation Checker (GRC) to support this process. This software tool identifies these additional links by examining the solid geometric representations of MEP BIM elements and related building space volumes and detecting the existence of certain geometric relations between them.

A. Geometric relations

```
| Adjacency | - OR - | Clash |
```

B. Graph relation (connectivity)

```
| fso:Component | fso:connectedWith | fso:Component |
```

Figure 5: Graph relation (connectivity) inference using geometric adjacency or clash detection

The diagram in Figure 5 illustrates how the solid geometric representations of BIM MEP components (part A) can be analyzed to identify geometric adjacency and clash relationships, which in turn can
be used to detect semantic links related to connectivity (part B). These links can be added to semantic graphs using the fso:connectedWith relation between fso:Component elements.

Figure 6: Graph relation (containment) inference using geometric clash or containment detection

Similarly, the diagram of Figure 6 shows how geometric clash and containment relationships among the solid geometric representations of BIM MEP components and related building space volumes (part A), can be used to detect missing semantic links related to containment (part B). These links can be established using the bot:hasElement relation between bot:Element and bot:Zone elements.

While the geometric relations between the solid geometric representations of MEP elements and their respective building space volumes can assist in identifying missing semantic links (topological relations), not all detected geometric relations indicate such missing semantic links. Clash errors in the design of MEP components cannot be used to infer a semantic relation link between them. Moreover, two adjacent MEP components related through a geometric adjacency relation should not be connected via a semantic relation (topological relation) unless they are physically connected. Therefore, all the geometric relations detected by GRC should be validated using an appropriate graphical user interface (GUI) to extract the desired missing semantic relations that accurately reflect proper connections.

Figure 7 illustrates GRC’s operational process. Initially, GRC receives a set of geometric representations of elements that fail to pass the semantic rule-checking (Figure 2). Alongside these elements, GRC receives the geometric representations of: the building’s space volumes and the remaining MEP elements (1). Subsequently, GRC performs geometric checking operations, and the checking results are reported in an OBJ format (2). Visual verification is performed to differentiate detected topological relations from design errors and element adjacencies. Once the verification process is complete, the list of verified topological relations (connectivity and containment) is combined with the output of GRC to generate the desired additional semantic graph links (3).

Figure 7: Semantic enrichment using GRC

**Enrichment through logical inferencing**

Representing our data using BOT, FSO, and Brick allows us to infer additional logical relationships between HVAC and Spatial elements using simple if-then rules. Eliminating the topological errors with the aid of GRC was an essential prerequisite step before deriving logical conclusions and further enriching our data.

Listing 2: If-then rule as SPARQL Construct query

```
```

To illustrate the inferencing process, we provide a simple example of an Air Handling Unit (AHU) feeding air into two air terminals located in two different office spaces (Figure 8). To generate the more abstract brick:feeds relationship, we write a SPARQL Construct query (i.e. Listing 2) to express and implement the following rule: If some equipment feeds fluid indirectly (i.e. in a transitive way) to a terminal
device which is located in a zone, then the equipment is feeding air (brick:feeds) in this zone.

Although logical inferencing allows the automatic inference of additional relationships such as the brick:feeds, further logical errors might still exist after the process. To overcome those errors, manual inspection and correction of the remaining issues are expected from the user with the aid of a GUI.

Insights from a real-world application

To test the usability and identify limitations of the proposed workflow, we implement it in the 26,000 sq.m BIM model of the new Museum of London (MoL), currently the most extensive cultural development in Europe [https://museum.london/].

Figure 9: GUI-supported discovery of "Terminal's location" error

Firstly, spatial and HVAC information is extracted from the building’s Architectural and MEP IFC models and transformed into an RDF graph, excluding any redundant geometric information from the graph. The proposed rules are then executed over the RDF graph to produce a validation report containing topological and logical rule violations related to building elements. Although it has been possible to infer additional relationships using GRC, our tool remains limited in detecting intersection and containment relationships. As a result, in cases where building components are neither contained nor intersecting a space volume, the missing topological relationship, for example, about the location of a terminal (Figure 9), could only be manually inferred from the user through a GUI environment.

Figure 10: GUI-supported discovery of "Missing Connectivity" and "Missing segment’s flow" errors

Another common topological issue discovered in the model was the lack of connectivity between segments and flow terminal devices, such as the space heater in Figure 10. Apart from the topological relationships, we also discovered logical violations, such as the lack of flow descriptions between building elements. In particular, the fluid direction in the flow segments of Figure 10 is not explicitly included in the model. Thus, even if we resolve the topological connectivity error, generating the upstream/downstream connectivity relationships requires further input from the user to resolve logical errors.

Figure 11: Automatic inference of downstream/upstream connectivity of spatial and HVAC components

Thanks to the logical inferencing process, the manual resolution of a topological or logical issue can generate further relationships between HVAC and spatial components. For example, through the manual assignment of the air terminals of Figure 11 in the space below, it was possible to automatically infer more abstract upstream/downstream connectivity relationship between HVAC components such as the specific AHU that "serves operationally" a particular space. Overall, our method has been successfully applied to MoL’s BIM, making it feasible to identify and correct missing topological relationships in a large-scale and complex model. Although considerable effort is still necessary to manually address unresolved errors, the one-off nature of the method holds potential for reusing it across BIM models of similar or lower complexity and scale.

Conclusions and Future work

In this paper, we have introduced a novel workflow that aims to minimise the efforts of building practitioners to extract, check and enrich HVAC topological information from BIM models with real-world imperfections. In particular: (1) we extracted and stored HVAC and spatial components in an RDF knowledge graph database, (2) defined a set of expert rules and executed them over the RDF graph to identify building elements with errors, (3) resolved those errors in a semi-automatically through (a) the development of GRC, a software tool that enables geometry-based inferencing of topological relationships, (b) implementation of logical if-then rules to infer logical relationships, (c) both supported by a GUI guiding users to resolve any unresolved issues.

We employed our workflow to extract HVAC system topology from a real-world building, demonstrating
its potential to solve modelling imperfections, reduce manual tasks and improve BIM data quality, thereby enabling the reuse of HVAC system topology as input for energy simulation, fault detection and facility management tasks. Yet, this implementation uncovered limitations we aim to address in future work. In particular, our next steps include further development of GRC to identify more advanced geometric relations, focusing on spatial proximity, and evaluating the workflow’s ability to generate usable HVAC information for energy modelling purposes.

References


