Building system data integration using semantic

Teclaw W. [1,4], Oraskari J. [2] , Rasmussen M.H. [3] , Labonnote N. [1] , Hjelseth E. [4]
1) SINTEF, Norway 2) RWTH Aachen University, Germany 3) NIRAS, Denmark 4) NTNU, Norway
wojciech.teclaw@sintef.no

Abstract. In recent decades the construction industry started a rapid digitalisation process resulting in the widely used BIM methodology. Despite developing multiple Industry Foundation Classes (IFC) schemas, the interoperability and data integration between models remains an issue. This study aims to fill this gap by providing a framework for data integration between various models, achieved by analysis of available ontologies, IFC schema limitations, and their relations. This methodology is tested based on a case study containing multiple BIM models. The study's findings are expected to enhance the connectivity between MEP components placed in different models and provide knowledge representation for developing the Digital Twin concept.

1. Introduction

1.1. Context

The origins of Building Information Modelling (BIM) can be traced back to the 1970s (Eastman, 1974, 1975). Eastman proposed using virtual building models as an alternative to traditional drawings and integrating databases for data management, visualisation techniques, and quantitative analysis. This approach was initially referred to as Building Description System (BDS). Around the turn of the century, the term BIM began to gain traction and eventually became widely adopted (André Borrmann *et al.*, 2018). A key step was made in 1995 when the International Alliance for Interoperability (IAI) was established to enhance interoperability and productivity in the construction industry. One of the key initiatives of the IAI was the development of the Industry Foundation Classes (IFC), a data model for the exchange of building information. The IAI published the first release of IFC in 1997, marking a significant milestone in the history of BIM.

In recent years, governments have placed significant emphasis on addressing the issue of rapidly increasing CO2 emissions. This contributes to the popularity of the digital twin concept as a potential solution to the problem. The digital twin's primary objective is to provide a comprehensive and structured representation of building information and equip the industry with autonomous tools capable of controlling, managing, and detecting anomalies in a construction object throughout its lifecycle, encompassing the construction and operation phases (Boje *et al.*, 2020; Hu *et al.*, 2021). Achieving such a structure and linkage between an object's components requires the development of machine and human-readable data representation. Consequently, in recent years, many research initiatives have proposed using semantic web technologies as a data-representation model as a key to successfully implementing the Digital Twin concept.

Creating a semantic representation of BIM models is laborious when done manually. AECO professionals and researchers have developed automation tools to convert BIM models into their semantic equivalent. However, challenges arise due to the characteristic of BIM models, existing workflows and included data. Independent models expressing domain knowledge are combined to create a federated BIM model, aligning with the ISO 19650 standard and enabling easy interpretation of standalone models.

The industry frequently employs the method described, which enables domain designers to work independently on their respective areas of expertise and later combine their knowledge by uploading their work to a Common Data Environment (CDE). However, this method is not without its limitations. One major drawback is the absence of all connections between the final asset. In particular, no relations are observed between two or more mechanical, electrical and plumbing (MEP) models. Mechanical equipment is being modelled in one domain while relying on resources from several others. Therefore, when exported from its native format to IFC, only relationships within one file are preserved, resulting in a lack of interdisciplinary connections between MEP systems and domains.

1.2. Research questions and overview

This research proposes a methodologic interdisciplinary framework linking MEP components in multiple domain BIM models. Therefore the following research questions have been formulated:

RQ1: What are the BIM model's undefined relations between MEP elements?

RQ2: How to complement missing knowledge between graph representation of BIM models?

Section 2 gives a background. Section 3 presents the methodology and applied rules. Section 4 discusses the results and limitations of the framework. Finally, the last two sections answer the research questions, draw a conclusion and give an overview of the further work.

2. Background

2.1. Ontologies

In recent years, various research initiatives focused on developing a broad range of ontologies standardising existing workflows and providing knowledge interpretability by computer systems. Semantic web technologies aim to establish systematic associations among various information sources by utilising URIs as identifiers and ontologies as a framework to describe relationships, commonly implemented with the Resource Description Framework (RDF). The W3C Linked Building Data community has designated multiple ontologies, including ifcOWL, as the standard format for exchanging information in a form satisfying the requirements of the linked data concept (Beetz, van Leeuwen and de Vries, 2009). IfcOWL is an ontology representing the Industry Foundation Classes data model using Web Ontology Language (OWL). It includes various information about geometry, property sets, elements, systems, and relations. As the ifcOWL specification closely mirrors the IFC standard and serves as a serialisation of IFC, it shares the inflexibility and does not provide the modularity offered by the new LBD ontologies.

To address the challenges of achieving interoperability and flexibility with ifcOWL, researchers have proposed methods to streamline the schema and create a more user-friendly structure. A notable example was the development of SimpleBIM. This effort culminated when introducing the core of a new modular ontology, the BOT. By relying on just six fundamental classes, the BOT substantially simplifies the expression of the BIM model (Rasmussen *et al.*, 2020). The other studies focused on applying a similar approach but providing a more understandable structure for the MEP system and its connectivity. The Flow System Ontology (FSO) provides a coherent and extendable form that might be successfully applied to most of the building system (Kukkonen *et al.*, 2022). Both ontologies are aligned and provide a holistic representation of the building MEP systems, significantly increasing the IFC models' potential. Nevertheless, Pauen et al. (Pauen *et al.*, 2021) noticed that the FSO ontology does not provide direct control and knowledge about the system's state and fluid mass proposing a Tubes System Ontology (TSO).

2.2. IFC schema limitations

Governments increasingly promote digital transformation to minimise costly planning errors and construction delays. In this context, the IFC format has emerged as a crucial tool for process stakeholders to exchange information without divulging proprietary knowledge or unique modelling practices contained in native files. IFC is available in various Model View Definitions (MVDs), which are subsets of the schema that are tailored to specific use purposes like standardised export for facility management purposes, coordination view or a structure aligned with the Construction Operation Building information exchange format (COBie) (Laakso and Kiviniemi, 2012). The increasing popularity and the number of successful uses led to an analysis of potential defects and constraints in the IFC schema. One of the limitations of the IFC schema is the lack of all connectivity detail between MEP components (Xiao, Li and Hu, 2019). Such gaps might occur within a single model or in between them. Some missing connections are due to modelling mistakes, inaccuracies, errors during export or inappropriate modelling practices (Hu *et al.*, 2018). The second type of missing connections involves interdisciplinary connections occurring in the intersection between domains (Hosamo *et al.*, 2022). The modelling practice shows that every domain works in a separate native model, developing its content depending on task characteristic (e.g. number of domains and their complexity) and BIM-related documents (e.g. Master product delivery table and content) (Sacks *et al.*, 2018). According to the BIM Execution Plan (BEP) and defined data drops, stakeholders produce domain models according to the standards and rules of the project. Such an approach suits coordination, construction management and estimation (Sacks *et al.*, 2018). However, it does not provide all relations and detailed information required for the automated conversion of the model to its semantic representation (Andre Borrmann *et al.*, 2018).

2.3. Components relations

The IFC schema contains a substantial number of relations whose purpose depends upon the function and requirements associated with components. The IFC structure was created to provide a systematic and detailed way of describing the functions of building components while establishing specific and precise relationships between them (Lai and Deng, 2018). The high-level classes, such as an ifcBuilding or ifcSite and their relations are easy to follow and understand because one can certainly imagine that a building is placed on a site, and such relation is reflected in an IFC file. However, with the system components, complexity increases, including the representation of their properties or relations with other elements (Rasmussen *et al.*, 2019; Xiao, Li and Hu, 2019). Such complexity greatly benefits the industry, allowing accurate, reflecting features, even for layered elements such as Air Handling Units (AHU), explaining in detail its connection with other components through IfcDistributionPorts, as shown in Figure 1. The usage of these relations defined within the scope of a single IFC model was broadly investigated (Liebich, 2009; Hosamo *et al.*, 2022). However, none of the studies analysed a mechanism of a model extension by the automated discovery of new relations unexisting in the model but possible to define in the interdisciplinary intersection of models.

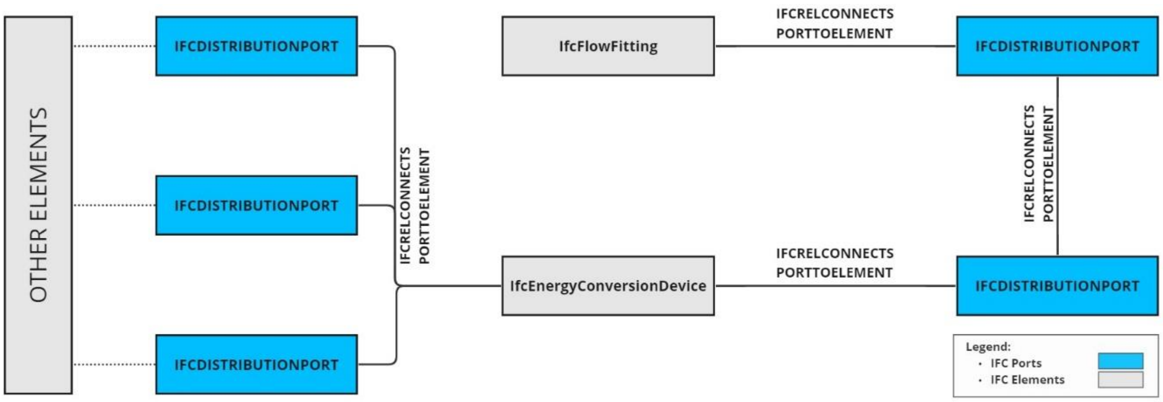

Figure 1 The relation between mechanical equipment and a flow segment element in IFC 2x3 model

3. Methodology

3.1. Model representation

The first point of solving the interoperability issue and gaps in the connectivity between MEP systems was selecting a convenient, extensible and machine-readable structure. Therefore using the IFC-LBD tool proposed by Rasmussen (Rasmussen, 2023), IFC models are converted into RDF files using the FSO ontology requirements. The FSO ontology describes the relations between components more extensively than IFC does. It contains information about flow directions, heat transfer and electric flow (Kukkonen *et al.*, 2022). FSO relations between components in a tree structure form are shown in Figure 2. The diagram illustrates the hierarchical order of relations, starting with a top-level concept describing the relation of two connected components and finishing at a low level specifying the affiliation to a system characteristic. Moreover, the relations of FSO ontology, accessible in an IFC file, were marked in blue.

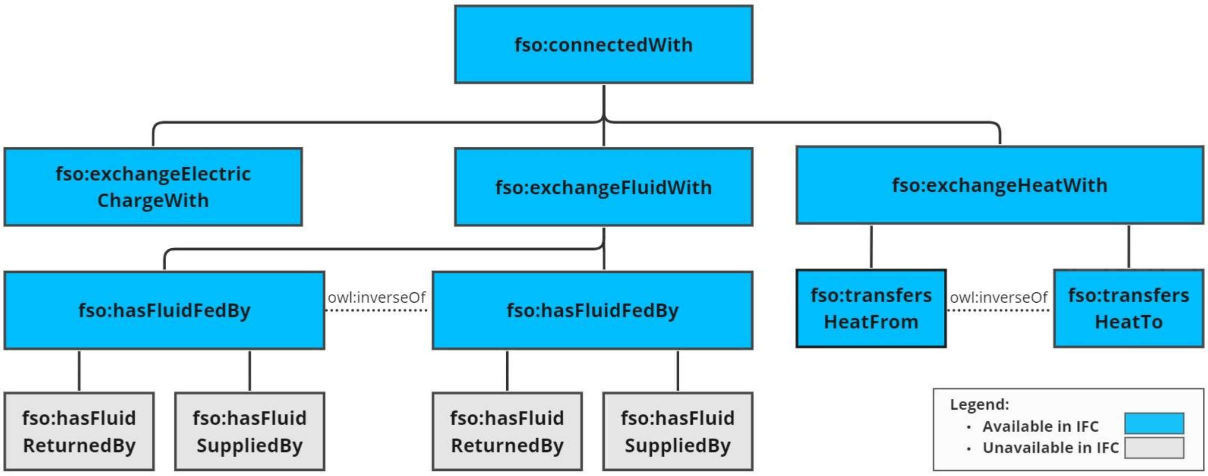

Figure 2 Relation hierarchy tree between FSO components with IFC alignment. The diagram is inspired by (Kukkonen *et al.*, 2022)

The reason for the limited accessibility of the connection between components is based on the IFC structure, which does not provide standardised information about the system classification and its characteristics. Therefore, the analysis extends the relations between components relay on the flow direction and ifcPorts centre points. To perform the analysis efficiently, it is crucial to deeply understand the relations and features provided by the class, being an inseparable element of IFC models.

3.2. Port features

The presented methodology focuses on unconnected ports assuming that the rest of the connected elements are properly connected using the IFC schema. The analysis is based on the IFC 2x3 schema because the IFC-LBD converter is not fully compatible with IFC4 and newer schemas (Rasmussen, 2023). After converting from a model to its semantic equivalent, they are extracted from an IFC file. Based on properties and their relations, its representation includes the following properties:

- Id
- IfcGuid
- ModelId
- Representation:
  - Location
  - ConnectorNormal
  - ParentId
  - PortCharacteristic

By default, the FlowNormal and Location properties are expressed in the local coordinate system, referring to the recursively nested IFCLOCALPLACEMENT and IFCAXIS2PLACEMENT3D instances. Therefore, the first step is retrieving the placement information in the form of the transformation matrix used to convert local ifcPort characteristic description into global.

Except for the port's geometrical properties, their representation requires additional information about the PortCharacteristic. Based on the IFC 2x3 provided by Liebich et al. (Liebich, 2009) expression of this parameter is limited to the following options:

- SOURCE (provides fluid to a port)
- SINK (receives fluid from a port)
- SOURCEANDSINK (port which receives or provides fluid from/to a port)
- NOTDEFINED (port without explicitly defined flow direction)

3.3. Connectivity analysis

All the unconnected ports were converted, retrieved, and compared between models to discover new connections. They were tested under three rules. The first is checking their Euler distance between connection location points P1 and P2 defined by three-dimensional coordinates x, y, z. The calculated Euclidean distance between elements must be less than a fixed number using equation (1). In the analysis, 25 millimetres was selected as a representative value, a diameter equivalent to the thinnest element in a model.

$$\mathrm{d} = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2 - (z_1 - z_2)^2} \tag{1}$$

If the condition is valid, the next step is checking the angle between two port normals – a and b. The proposed methodology assumes that the angle between connector normals must be less than one cartesian degree using equation (2). One cartesian degree is a standard tolerance of a deviation for the majority of MEP elements.

$$\theta = \cos^{-1}\left(\frac{a \cdot b}{|a|\,|b|}\right) \tag{2}$$

The last step is the definition of the connection order. This definition is crucial in ensuring that the fluid flows in the correct direction. Such flow can be defined only in six variants based on the PortCharacteristic property, as presented in

Table 1. Provided comparison analysis cannot use more specific information like port geometrical properties or a domain because such knowledge is not available in the IFC 2x3 (Liebich, 2009).

Table 1 Flow direction predictability based on the ports features

PortCharacteristic	SOURCE	SINK	SOURCEANDSINK	NOTDEFINED
SOURCE	X	✓	✓	X
SINK	✓	X	✓	X
SOURCEANDSINK	✓	✓	X	X
NOTDEFINED	X	X	X	X

4. Results

4.1. Case study

The following section presents the implementation of the proposed methodology. Using Autodesk Revit 2021, three IFC models were generated to reflect the three domains: ventilation, sanitary, heating and cooling, as presented in Figure 3.

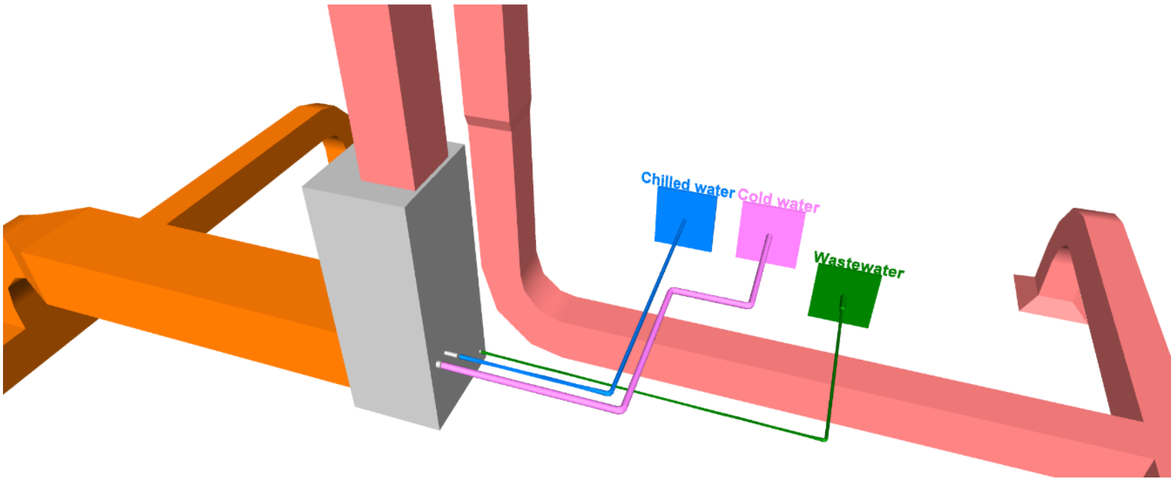

Figure 3 Federated model of three IFC models

All models are generated from Autodesk Revit to IFC separately using a standard IFC 2x3 Coordination View 2.0 skimmer, one of the standardised approaches of ISO 19650. IFC files contain elements and systems presented in Table 2.

Table 2 Elements and systems included in IFC Models

Model	Ventilation	Piping	Heating&Cooling
Elements	IfcFlowSegment, IfcFlowFitting, IfcFlowTerminal, IfcBuildingStorey, IfcEnergyConversionDevice	IfcFlowSegment, IfcFlowFitting, IfcFlowTerminal, IfcBuildingStorey, IfcFlowMovingDevice, IfcFlowTreatmentDevice	IfcFlowSegment, IfcFlowFitting, IfcFlowTerminal, IfcBuildingStorey, IfcEnergyConversionDevice
Systems	Supply Air Return Air	Water supply, Wastewater	Chilled water supply

4.2. Semantic alignment

In accordance with the methodology outlined in Section 3, the IFC-LBD was utilised to convert each individual IFC model into an RDF data structure using the N-Quads form (Rasmussen, 2023). The outputted representation of three IFC files was then persisted in a triple store reflecting the files' structure using FSO and BOT ontologies, providing a total of 1468 nodes. After applying the connectivity analysis presented in the Methodology section, the database was enriched by six additional predicates marked in red ovals in Figure 4.

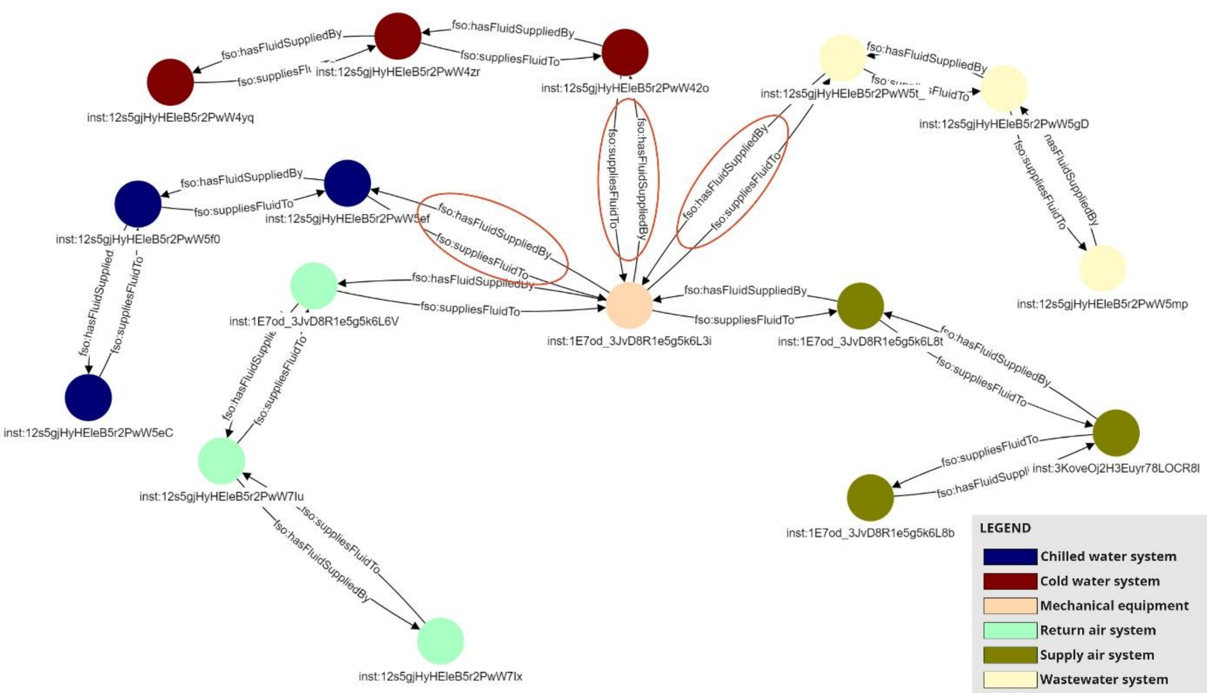

Figure 4 Intersection between AHU and other installation systems

Additionally, generated predicated provided new options for querying the database and link systems modelled in domain-specific IFC models with other systems by unconnected ports. Such knowledge extension was used to query the database to retrieve all devices impacting an element under investigation.

4.3. Usage of interdisciplinary connectivity

Additional predicates might be used differently depending on a task and required knowledge. Therefore two examples of practical usage of the framework by reaching elements stored in various IFC models were presented. The first use case is a query returning all IfcEnergyConversion devices affecting the selected "startElement" component by supplying a fluid. Such elements are found in the database using the SPARQL query presented in Listing 1.

```
PREFIX fso: <https://w3id.org/fso#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?device
WHERE {
  BIND(fso:EnergyConversionDevice AS ?DESIRE_DEVICE_TYPE)
  BIND(<http://www.theproject.org/ventilation/1E7od_3JvD8R1e5g5k6L3i> AS ?startElement)

  ?startElement fso:hasFluidSuppliedBy* ?DEVICE_SUPPLYING_FLUID .
  ?DEVICE_SUPPLYING_FLUID rdf:type ?DESIRE_DEVICE_TYPE .
}
```

Listing 1 SPARQL query returning a distinctive ifcEnergyDevices supplying a fluid to an element

Another example returns all elements at the path from a selected device to an ifcTreatmentDevice to which fluid it supplied. Based on the model example, the AHU instance generates wastewater and returns it to an IfcFlowTreatmentDevice element in the piping model. To find all segments and fittings between these two devices, the query presented in Listing 2 is used.

```
PREFIX fso: <https://w3id.org/fso#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

CONSTRUCT {
  ?component ?predicate ?connectedComponent
}
WHERE {
  BIND(fso:TreatmentDevice AS ?DESIRE_DEVICE_TYPE)
  BIND(<http://www.theproject.org/ventilation/1E7od_3JvD8R1e5g5k6L3i> AS
?SELECTED_DEVICE)
  VALUES ?predicate { fso:suppliesFluidTo fso:hasFluidSuppliedBy }
  ?SELECTED_DEVICE fso:suppliesFluidTo* ?elementConnected .
  ?elementConnected rdf:type ?DESIRE_DEVICE_TYPE .
  ?DESIRE_DEVICE rdf:type ?DESIRE_DEVICE_TYPE .
  ?component fso:suppliesFluidTo* ?DESIRE_DEVICE ;
             fso:hasFluidSuppliedBy* ?SELECTED_DEVICE .
  ?component ?predicate ?connectedComponent .
  FILTER (?component != ?SELECTED_DEVICE && ?component != ?DESIRE_DEVICE)
}
```

Listing 2 SPARQL query returning an ifcFlowTreatmentDevice receiving a fluid from a selected element

5. Discussion

The proposed methodology presented a conceptual framework for full semantic connectivity between domain-specific MEP models. The case study demonstrates that such an approach might provide a fully integrated environment explaining the connectivity between MEP components. Such integration reduces the difficulty of digital twin development because it allows different building elements to share knowledge about their state and understand their dependencies. Another benefit of concept application is the effortless knowledge extraction of data for a Building Management System (BMS) system. Based on relations and links between interdisciplinary elements, automation engineers might plan systems more efficiently and feasibly discover relations between multiple components, impacting a certain level, zone or space.

The presented framework solves the interoperability issue and lack of connectivity between elements in various IFC models. Based on the case study, different IFC models were connected, including the correct information about the flow direction. Nevertheless, the methodology has limitations caused by the constraints of the IFC schema. To make analysis more versatile and trustworthy, the IFC file could include additional characteristics of every connector. The first missing connector property is the geometrical characteristic of a shape and size. The lack of the property impedes the analysis and cannot ensure that two elements can be connected and maintain the same capacity and flow speed. Another useful feature would be the property of a port function, describing what type of system can be connected to a certain ifcPort. Additional features would add more value to the analysis, especially for complex elements, where belonging to only one system can not be determined. Therefore the current state of the proposed concept requires prudence and well-developed models.

The concept provides additional relation between IFC components and knowledge integration, which is beneficial for effortless and efficient re-usage of already defined knowledge converting it into informational basic for digital twin. The state of other dependence elements is crucial for asset management systems applying the DT concept.

6. Conclusion and further work

The proposed framework successfully implements a connectivity algorithm between MEP elements modelled in different models. The concept is easy to replicate and is a potential solution for extensions of the information stored in IFC files.

Further work will combine the proposed concept with the framework merging rooms and floor representation. Combining these two approaches might provide a close system to create a basis for the BIM to Digital Twin approach, enhancing multidisciplinary interoperability and feasible data exchange.

References

Beetz, J., van Leeuwen, J. and de Vries, B. (2009) 'IfcOWL: A case of transforming EXPRESS schemas into ontologies', *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 23(1), pp. 89–101. doi: 10.1017/S0890060409000122.

Boje, C. *et al.* (2020) 'Towards a semantic Construction Digital Twin: Directions for future research', *Automation in Construction*, 114(November 2019), p. 103179. doi: 10.1016/j.autcon.2020.103179.

Borrmann, André *et al.* (2018) 'Building Information Modeling: Why? What? How?', in *Building Information Modeling*. Cham: Springer International Publishing, pp. 1–24. doi: 10.1007/978-3-319-92862-3_1.

Borrmann, Andre *et al.* (2018) *Building Information Modeling*, *Building Information Modeling: Technology Foundations and Industry Practice*, Springer International Publishing. doi: 10.1007/978-3-319-92862-3.

Eastman, C. (1974) 'An Outline of the Building Descripition System, Institute of Physical Planning', *Carnegie-Mellon University*, pp. 1–23.

Eastman, C. (1975) 'The use of computers instead of drawings in building design', *AIA journal*, 63, pp. 46–50.

Hosamo, H. H. *et al.* (2022) 'A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics', *Energy and Buildings*, 261, p. 111988. doi: 10.1016/j.enbuild.2022.111988.

Hu, S. *et al.* (2021) 'Building energy performance assessment using linked data and cross-domain semantic reasoning', *Automation in Construction*, 124(February), p. 103580. doi: 10.1016/j.autcon.2021.103580.

Hu, Z.-Z. *et al.* (2018) 'BIM-based integrated delivery technologies for intelligent MEP management in the operation and maintenance phase', *Advances in Engineering Software*, 115, pp. 1–16. doi: 10.1016/j.advengsoft.2017.08.007.

Kukkonen, V. *et al.* (2022) 'An ontology to support flow system descriptions from design to operation of buildings', *Automation in Construction*, 134(February), p. 104067. doi: 10.1016/j.autcon.2021.104067.

Laakso, M. and Kiviniemi, A. (2012) 'The IFC standard - A review of history, development, and standardisation', *Electronic Journal of Information Technology in Construction*, 17(May 2012), pp. 134–161.

Lai, H. and Deng, X. (2018) 'Interoperability Analysis of IFC-based data exchange between heterogeneous BIM software', *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT*, 24(7), pp. 537–555. doi: 10.3846/jcem.2018.6132.

Liebich, T. (2009) *IFC 2x Edition 3*. Version 2., *buildingSMART International*. Version 2. Edited by T. Liebich.

Pauen, N. *et al.* (2021) 'TUBES System Ontology: Digitalisation of building service systems', *CEUR Workshop Proceedings*, 3081, pp. 43–54.

Rasmussen, M. H. *et al.* (2019) 'Managing interrelated project information in AEC Knowledge Graphs', *Automation in Construction*, 108(January), p. 102956. doi: 10.1016/j.autcon.2019.102956.

Rasmussen, M. H. *et al.* (2020) 'BOT: The building topology ontology of the W3C linked building data group', *Semantic Web*. Edited by K. Janowicz, 12(1), pp. 143–161. doi: 10.3233/SW-200385.

Rasmussen, M. H. (2023) 'IFC-LBD'. Available at: https://github.com/LBD-Hackers/IFC-LBD/ (Accessed: 9 February 2023).

Sacks, R. *et al.* (2018) *BIM Handbook*, *BIM Handbook*. Hoboken, New Jersey: John Wiley & Sons, Inc. doi: 10.1002/9781119287568.

Xiao, Y.-Q., Li, S.-W. and Hu, Z.-Z. (2019) 'Automatically Generating a MEP Logic Chain from Building Information Models with Identification Rules', *Applied Sciences*, 9(11), p. 2204. doi: 10.3390/app9112204.