Semantic Enrichment of Object Associations Across Federated BIM Semantic Graphs in a Common Data Environment

B. Ouyang, Z. Wang, R. Sacks
Seskin Virtual Construction Lab, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa, Israel

ABSTRACT: Explicit association semantics across federated BIM models facilitate the interoperability between design systems and open opportunities for intelligent applications. However, the automated inference and generation of such semantics remains a key challenge. Graph representations of BIM models have shown potential in supporting semantic enrichment and conveying formal building semantics across domains. In this paper, we present our implementation of a semantically-enriched Common Data Environment – the *Graph-based Core-extension Data Framework* – composed of a semantic graph layer and an object geometry extension layer, hosting the heterogeneous information from BIM models. A novel rule-based inferencing approach, the *TIOC algorithm*, was implemented on the federated building representations to infer and establish inter-domain topological and correspondence relationships, interconnecting disjoint domain specific graphs with meaningful formal semantics. This study is the first known attempt to implement a semantic enrichment algorithm for implicit object associations in the context of multidisciplinary models.

1 INTRODUCTION

Digitalization in the Architecture, Engineering and Construction (AEC) industry has opened a new field of research that focuses on developing computerized solutions to real-world problems that do not arise with traditional ways of conveying building information. Building Information Modelling (BIM), a key aspect of digitalization in AEC, adds enormous value in planning, design, construction, maintenance, and other processes throughout the entire building lifecycle. Through the past decade, graph theory and algorithms from the fields of mathematics and computer science have matured to the point where graph technologies have great potential to represent and operate on building information from BIM models, enabling advanced applications beyond what current BIM can offer.

A key use-case of graphs in BIM is the knowledge graph representations of building models in the context of the Semantic Web for efficient information exchange. The Semantic Web, devised by Berners-Lee et al. (2001), is an extension of the World Wide Web that aims to make information on the web not only human-readable but also machine-understandable. Linked Data is a concept within the Semantic Web (Berners-Lee, 2006), in which structured web data expressed in Resource Description Framework (RDF) are interlinked with each other, enabling semantics from different domains to be related. These technologies have been recognized as a promising foundation for representing multidisciplinary building information on the web, making them transparent and accessible to every project stakeholder.

To address the holistic picture of multi-domain collaboration in AEC, where data representations are not limited to RDF graphs, investigations have been made into federated Common Data Environments (CDE) that act as hubs to host heterogeneous information from different domains (ISO, 2018). RDF graphs and supplementary data in their native formats can be inter-linked and shared on a CDE, to which every project participant has access. Beyond its original intent of sharing information, we propose to further explore the potential of the data made available in CDEs for semantic enrichment, especially considering that data from multiple disciplines can now be utilized for a single enrichment task. Semantic enrichment is the process that adds meaningful semantics to a digital building model by applying artificial intelligence (AI) techniques (Belsky et al., 2016). This paper aims to demonstrate the feasibility of automating the association semantic enrichment task on federated BIM models hosted in a Linked Data CDE.

We first give an overview of the course of development of graph representations and applications in the context of BIM. Next, we present our implementation of a CDE in which multi-disciplinary building semantics in RDF are stored alongside with the object geometries to cater for the subsequent semantic enrichment task. The procedure and reasoning of the entire semantic enrichment process on graphs is explained in detail in section 4. Finally, the paper concludes with a discussion about the possible applications that could be built on top of the enriched model representations, as well as steps for upcoming research to perfect the presented pipeline.

2 RELATED WORK

2.1 *Semantic Enrichment on BIM Graphs*

Existing BIM software can only process explicit information, and the lack of semantics leads to interoperability issues (Sacks et al., 2018). Two types of semantic information in BIM were classified by Xue et al. (2018), including information for an individual object and relationships between objects. Bloch & Sacks (2020) identified four basic semantic enrichment tasks, including classification, calculation, association, and creation, which could also be branched into 1) enrichment of information for individual objects or aggregations of objects, and 2) enrichment of object relationships.

For the examples of semantic enrichment on BIM graphs, a pioneering work from Collins et al. (2021) applied Graph Convolutional Neural Networks on object geometry graphs to classify their object types. In this study, each graph represents a BIM element instead of a BIM model. Shortly after, Wang et al. (2021, 2022) constructed the first BIM graph dataset that registered apartment layouts, and proposed an improved graph neural network algorithm, *SAGE-E*, to predict the unlabeled room types. The results showed better prediction accuracy compared to experiments that adopted traditional machine learning approaches on similar tasks (Bloch & Sacks 2018).

However, these works encounter the problem of generalizability and scalability, as the graph datasets are mostly created from the BIM models according to rules defined by the authors, and the products are only tailored to the specific use-cases. More importantly, previous semantic enrichment research are limited to BIM object classifications that relate to single BIM objects, while the enrichment of object relationships (the association task) remains largely unexplored.

On the other hand, researchers realized the need for such association semantics for enabling BIM intelligent applications. Törmä (2013) presented the idea that, with individual BIM objects from different partial models linked, intelligent functions like change management, cross-model information access, status monitoring, etc. can be realized. Similarly, Pauwels (2014) argued that when models are linked with inter-domain semantics, information exchange and management can be improved to facilitate decision-making in the BIM lifecycle. Both studies identified the need for inter-domain semantics, but their results are conceptual, with no validation of implementations to enrich such association relationships automatically.

2.2 *Knowledge Graphs in the AEC Industry*

Törmä (2013) and Pauwels (2014) introduced the idea that knowledge graph representations may serve as the technical foundation for expressing association semantics between concepts and objects across domains. Development of the ifcOWL ontology, a Web Ontology Language representation of the IFC EXPRESS schema, was a milestone achievement in this field (Pauwels & Terkaj 2016). Recognizing that the excessive complexity inherent in the full IFC-RDF representation prevents any practical application (Pauwels & Roxin 2017), subsequent researchers have aimed to develop better knowledge graph representations from IFC models that are simpler, more extensible, and more modular (Pauwels et al., 2022). These works, led by the W3C LBD community group, contributed to the development of modularized ontologies, including BOT (M. H. Rasmussen et al., 2020) as the core, plus a series of purpose-specific extensions. The dedicated IFCtoLBD converter allows conversion of IFC models into RDF Abox graphs structured according to the ontologies defined under the LBD ecosystem (Oraskari, 2022).

An RDF graph may not be the most appropriate data storage format for all types building information. Nevertheless, research in this field opens a new way for building information to be communicated across domains transparently. Semantics can now be expressed within and across domains in a formal language, along with ontologies that schematize the way knowledge is expressed.

2.3 *Common Data Environment*

Much work has also been done to investigate data exchange in the complex real-world scenarios of multidisciplinary collaboration. Here, given the federated model approach, heterogeneous information packets in various forms are exchanged among multiple project participants, and the need for a central digital platform with regulated collaborative procedures is well recognized. To address this issue, the concept of Common Data Environment (CDE) was devised as a "common digital project space which provides well-defined access areas for the project stake-holders combined with clear status definitions and a robust workflow description for sharing and approval processes" (Preidel et al., 2015) and the procedures of managing CDEs were formalized in ISO 19650 (ISO, 2018).

Recently, an interesting research thrust demonstrated the possibility of a Linked Data-based CDE, bringing together the benefits of both technologies (Malcolm et al. 2021, Werbrouck et al. 2019, 2022). Such an implementation would allow heterogeneous building information in different forms to be linked to the RDF layer that captures core building semantics, and this can be realized at object-level granularity. Multiple mentions of semantic enrichment appear in these studies. In the most recent work (Werbrouck et al., 2022), an exemplifying semantic enrichment task was provided as part of the case study, demonstrating the user-initiated manual documentation process of a newly discovered incidence of damage to an object, illustrating how the information was made available to all through linking with other models. Although one might argue whether that particular example can be regarded as a semantic enrichment task according to its canonical definition (Belsky et al., 2016), which stipulates automation of the process, it undoubtedly focused attention on the potential of such an integrated data repository in facilitating the generation and representation of new semantics.

To date, information in a CDE is mostly shared and utilized 'as-is'. Applications that use CDEs are rather predictable and straightforward compared to those demonstrated with the application-specific graphs. Current research focuses on improving the framework for data representation and sharing on the theoretical level, but without human intervention, information richness can go only as far as what the original BIM model contains.

2.4 *Gaps in Knowledge*

Research related to the association of BIM objects in the context of semantic enrichment is scarce. There is reason to believe that, with the abundant multidisciplinary data made available by a CDE, AI applications of various forms can be deployed to facilitate the automated reasoning and inference of new knowledge, especially when it comes to the implicit semantics between objects across federated models. Enriched information can be explicitly expressed and appropriately shared between stakeholders using Linked Data approaches to provide access to every project participant.

3 CDE FOR SEMANTIC ENRICHMENT

RDF graphs facilitate the expression of associations between BIM entities, but as mentioned before, it is not the most convenient format for storing all types of building related data. For example, buildings could have sensors that collect data through time, which suggest the need for dedicated time-series management systems (Jensen et al., 2017). Exact geometries are another exemplifying type of data that cannot be efficiently represented in RDF graphs (Pauwels & Roxin, 2017). Moreover, even for data that can be efficiently represented as graphs, they are not equally important in terms of the frequency with which they will be accessed. Storing them all in graphs would

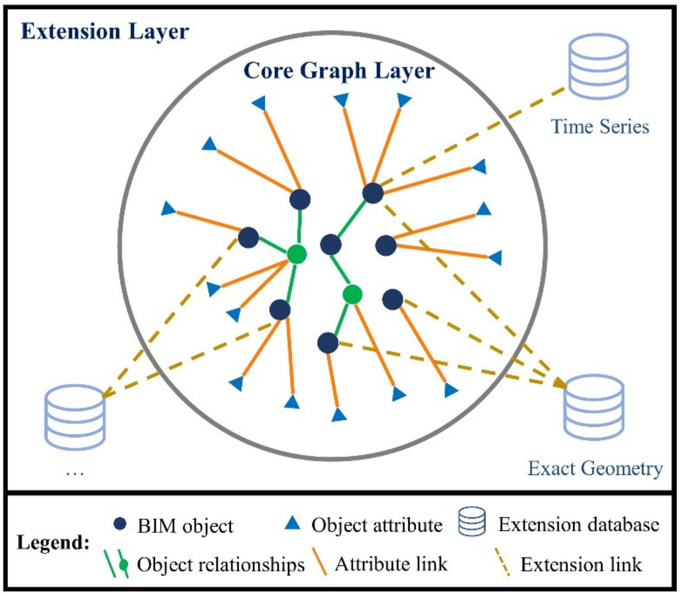

Figure 1. Graph-based Core-extension Data Framework.

increase the size of the graphs and potentially impair retrieval speed. Extracting those data can help to keep the core graph concise. A possible solution is to save non-graph data and less essential data independently outside the graph but with virtual links to connect them to the core graph.

For these reasons, we present in Figure 1 the Graph-based Core-extension Data Framework as our implementation of a CDE. The core LBD graph layer includes BIM elements, object relationships and direct object attributes in RDF format. The extension layer stores BIM-related resources in open schemas without format limitations. Virtual links between the graph and extension layer (yellow dashed lines) are generated and stored as attributes to the BIM element nodes. For example, the exact geometry of a wall compiled to a PLY file will be stored in the extension layer. The path or URL through which the file can be accessed is stored as an attribute of the wall node and serves as a virtual link.

The new data framework provides a flexible basis for the complete and efficient storage of all kinds of building information. Moreover, the availability of data makes it ideal for deployment of semantic enrichment applications.

4 ASSOCIATION SEMANTIC ENRICHMENT

This section explains and details the semantic enrichment procedures shown in Figure 2. The pipeline has three stages. Section 4.1 deals with data preparation, during which each IFC partial model is compiled into a raw subgraph and stored in the core graph layer as the basis for the subsequent enrichment. Section 4.2 discusses the construction of the CDE, where object geometries within each model are supplemented to their respective raw graphs to complete the representations. Section 4.3 considers the inference and

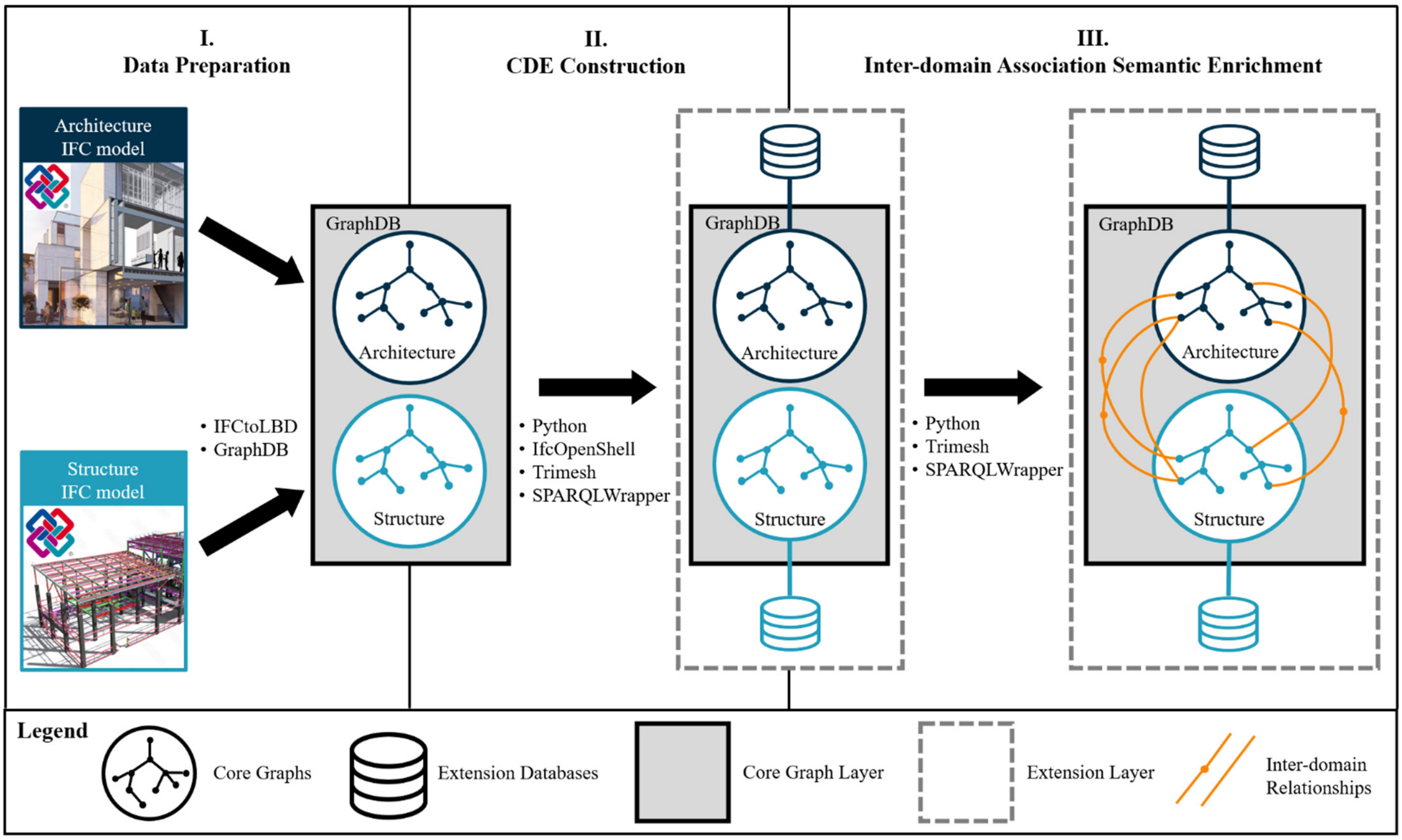

Figure 2. Pipeline of the semantic enrichment workflow on graphs.

enrichment of inter-domain association semantics, a step that adds relationships between objects across disciplines, making design intents explicit. In the last section, we present a case study as a validation of this workflow.

4.1 *Data Preparation*

The semantic enrichment pipeline begins with the federated architectural and structural BIM models of a given project, assuming they are well coordinated and aligned. Correct placement of the model objects in global coordinates is critical to the subsequent workflow. The IFCtoLBD converter (Oraskari, 2022) converts IFC models exported from BIM authoring tools into LBD graphs in RDF formats. Initializing the core graph layer involves loading these LBD-based subgraphs as model representations for each domain into the graph database, which in our case is GraphDB (2021). Upon successful establishment of the database, all remaining workflows use direct communication of the local Python environment with the database's endpoint. The *SPARQLWrapper* (2022) library was used to establish this connection, allowing query and update requests to be sent to the database in the form of SPARQL scripts.

The storage of domain subgraphs in the database follows the Named Graphs data model. As an extension of the original RDF data model, the Named Graphs approach gives each RDF subgraph a unique identifier, allowing properties like metadata to be assigned on the graph level rather than node level only. In our use case, such metadata might include information like model creator, creation date, authoring software, accessibility, etc., all of which are essential for data management in the collaborative design environment. Moreover, this approach aggregates the subgraphs elegantly by allowing information from one subgraph to be accessible from the others through referencing the unique identifiers, while at the same time preserving the federation structure of the subgraphs so that each can be maintained and manipulated separately. This is an important feature for our subsequent semantic enrichment since the enriched relationships can be stored and managed externally to the domain subgraphs. We call this the 'CBIM' (Cloud BIM) Graph.

4.2 *CDE Construction*

As mentioned, geometry is one of the key features of building objects for determining their relationships with other objects, yet this information is lost during the conversion to LBD graphs. Once the graph database is set up, the first step towards global semantic enrichment is to supplement the missing geometry information for each subgraph in a CDE approach.

The GUID of every building object in each subgraph is extracted for locating the corresponding IFC object in the original IFC models. Then, using the *geom* module from the *IfcOpenShell* library (2022), the geometric representation of each object, with its placement in the global coordinate system, can be processed and converted into a raw triangulated mesh, on which various analyses can be performed with third-party mesh processing libraries. Among many other features, *trimesh* (Dawson-Haggerty et al., 2019) provides a toolset for exporting such in-memory mesh geometries to standard PLY files and for generating their axis-aligned bounding boxes. We encode the defining parameters of the bounding box and the directory of the exported PLY file as strings and attach them to the building object node as new properties. The *bot:hasSimple3DModel* and *fog:asPly* relations are datatype properties adopted from the existing BOT Ontology (M. H. Rasmussen et al., 2020) and FOG Ontology (Bonduel et al., 2019) suitable for linking the simplified and exact geometry encodings. The reason for storing the dual representations derives from considerations of computation efficiency, and will be discussed later.

4.3 *Inter-domain Association Semantic Enrichment*

Following the preliminary enrichment of object geometries comes the actual enrichment of inter-domain relationships across the subgraphs, utilizing all information available from the graph database and the extension repositories. The enrichment process involves the sequential generation of three types of cross-domain relationships from the CBIM ontology (as proposed in (Sacks et al., 2022)): the correspondence of building storeys, the spatial relationships between building elements, and their inferred correspondence. Here, we define corresponding objects (or aggregations of objects) as objects that are functionally equivalent in a physical building. As such, correspondence should exist between cross-domain objects that give identical functionality, or between intra-domain objects (or aggregations of objects) if they are mutually exclusive alternatives. In this paper, the degree to which two objects overlap geometrically in the 3D space is regarded as the primary factor in evaluating their correspondence.

Here, we propose a rule-based algorithm named the *Topology-based Inter-domain Object Correspondence* test algorithm (*TIOC* algorithm), summarized in Figure 3. The core idea is the following. For every input object pair, a computation of the clear distance between their bounding box encodings is performed. If they are in proximity, defined as a pre-set distance threshold, a *CBIM:RelSpatial* node that registers the computed topological properties between the two bounding boxes is generated. Since overlap of the bounding boxes is a necessary but not a sufficient condition for the exact geometries to be in contact, the algorithm then filters for object pairs that have a zero clear distance registered and performs a Boolean intersection check (a function available in *trimesh*) on their PLY exact geometries retrieved from the

Algorithm 1: Topology-based Inter-domain Object Correpondence test

Input: Empty Enrichment Graph for inter-domain relationships: *e_graph*
Storey objects from the Architectural and Structural Subgraph: *a_storeys* = {*a_storey*}, *s_storeys* = {*s_storey*} where *a_storey* and *s_storey* carry the elevation information *elev;*
Building objects from the Architectural and Structural Subgraph: *a_objs* = {*a_obj*}, *s_objs* = {*s_obj*}, where *a_obj* and *s_obj* carries the bounding box encoding *bbox*, the exact geometry encoding *geom*, the storey information *stor,* and the type information *type;*
Predefined elevation tolerence: *ε_elev;*
Predefined offset tolerence: *ε_offset;*
Predefined volume overlap tolerence: *ε_overlap;*
Predefined list of corrsponding class sets: *cclass_list;*

Output: Final Enrichment Graph: *e_graph*

```
# Find corresponding storeys across domain graphs
foreach a_storey in a_storeys do
    foreach s_storey in s_storeys do
        if |a_storey.elev − s_storey.elev| ≤ ε_elev then
            e_graph.AddTriple(a_storey, "equivalentTo", s_storey)
            e_graph.AddTriple(s_storey, "equivalentTo", a_storey)
        end if
    end foreach
end foreach
# Generate topological relationships across domain graphs
foreach a_obj in a_objs do
    foreach s_obj in s_objs do
        # Verbal description of topology
        bbox_topo ⟵ ComputeTopology(a_obj.bbox, s_obj.bbox)
        # Bounding box offset
        bbox_offset ⟵ ComputeOffset(a_obj.bbox, s_obj.bbox)
        if e_graph.HasTriple(a_obj.stor, "equivalentTo", s_obj.stor)
        and bbox_offset ≤ ε_offset then
            relSpatial ⟵ InitializeNode()
            e_graph.AddTriple(rel_spatial, "hasSubject", a_obj)
            e_graph.AddTriple(rel_spatial, "hasObject", s_obj)
            e_graph.AddTriple(rel_spatial, "hasTopology", bbox_topo)
            e_graph.AddTriple(rel_spatial, "hasOffset", bbox_offset)
            if bbox_offset == 0 then
                # Physical contact status
                is_contact ⟵ TestContact(a_obj.geom, s_obj.geom)
                if is_contact == TRUE then
                    e_graph.AddTriple(rel_spatial, "inContact", TRUE)
                else
                    e_graph.AddTriple(rel_spatial, "inContact", FALSE)
                end if
            else
                e_graph.AddTriple(rel_spatial, "inContact", FALSE)
            end if
        end if
    end foreach
end foreach
# Generate correspondsTo relationships across domain graph objects
foreach a_obj in a_objs do
    foreach s_obj in s_objs do
        if e_graph.HasTriple(rel_spatial, "hasSubject", a_obj)
        and e_graph.HasTriple(rel_spatial, "hasObject", s_obj)
        and e_graph.HasTriple(relSpatial, "inContact", TRUE)
        and {arc_obj.type, str_obj.type} ∈ cclass_list then
            # Architectural object volume
            a_vol ⟵ ComputeVol(a_obj.geom)
            # Structural object volume
            s_vol ⟵ ComputeVol(s_obj.geom)
            # Overlapping volume
            o_vol ⟵ ComputeOverlapVol(a_obj.geom, s_obj.geom)
            if overlap_vol / a_vol ≥ ε_overlap
            or overlap_vol / s_vol ≥ ε_overlap then
                e_graph.AddTriple(a_obj, "correspondsTo", s_obj)
                e_graph.AddTriple(s_obj, "correspondsTo", a_obj)
            end if
        end if
    end foreach
end foreach
return e_graph
```

Figure 3. The TIOC algorithm.

extension layer. The results, specifying whether the objects are indeed in contact, are recorded to the *CBIM:RelSpatial* node. If true, the procedure checks in parallel the geometries of the two entities and their overlapping part, as well as other features extractable from the core graphs, against a set of predefined rules that encapsulate expert knowledge for correspondence judgement. If all the conditions are satisfied, an explicit *CBIM:correspondsTo* relationship is established between the two object nodes.
For sake of computational efficiency, we implemented a preliminary mapping of the building stories across models based on the storey elevations, assuming they have one-to-one correspondence. Potential nuances in the elevations of the architectural and structural storeys can be taken care of by wisely setting the tolerance, which is another example of expert knowledge encapsulation. In this way, the search space can be significantly reduced by specifying that only objects from matching storeys are to be checked for spatial and correspondence relationships.

4.4 *Case Study*

Given the step-by-step explanation of the technical details above, the enrichment pipeline was validated using a set of federated BIM models retrieved from Autodesk's sample Revit model repository. The architectural model of a technical school and its corresponding structural model were used. The Revit models were each exported as IFC files, converted to LBD RDF graphs, loaded into GraphDB, supplemented with object geometries, then semantically enriched.
To illustrate the semantic enrichment process, we zoom into one of the columns located on the 3rd floor, at the south-west corner of the building (see Figure 4). The same column appears in both the architectural and the structural models, and when overlaying the two models together, the columns overlap in most of their volumes. When comparing the models side-by-side, any professional engineer can infer that the two column objects represent the same physical column in the real-world building, and hence should be considered as corresponding to one another. Figure 5 demonstrates how this kind of implicit association information was made explicit after the three-stage enrichment process.
First, the equivalence between the architectural and the structural storeys was deduced and established during the pair-wise exhaustive search over the storey objects from different subgraphs. Following that, an exhaustive search was performed over the building objects belonging to each pair of equivalent storeys, during which new *CBIM:RelSpatial* objects were generated whenever the bounding boxes of the pairs of building objects were found to be in close proximity (relative to a predefined threshold). In the example shown, the columns have overlapping bounding

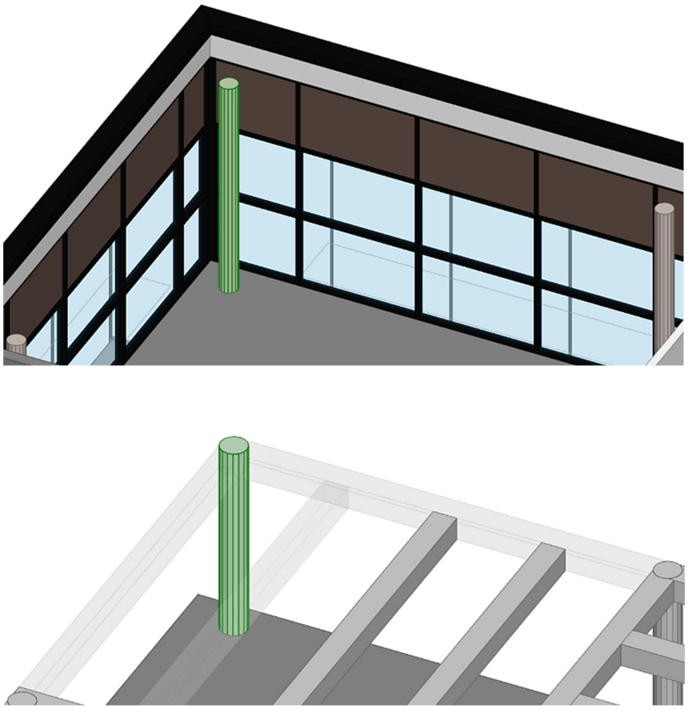

Figure 4. Corresponding columns from the architectural BIM model (above) and the structural BIM model (below).

boxes in all three projections X, Y, and Z, indicating that they may be in physical contact. This triggered a computation on the exact geometries retrieved from the extension layer, which revealed that the two columns are not only in contact, but overlap to a great extent. Upon further testing the two column instances against predefined rulesets that use their non-geometric properties, such as object type, name, and their class identifiers, the algorithm deduces that the columns indeed correspond to one another. The enriched information is supplemented back to the graph.
Snippet 1 shows a fragment of the TriG file exported from GraphDB that depicts the part of the graph relevant to the described column example. The supplemented geometry information is stored in the

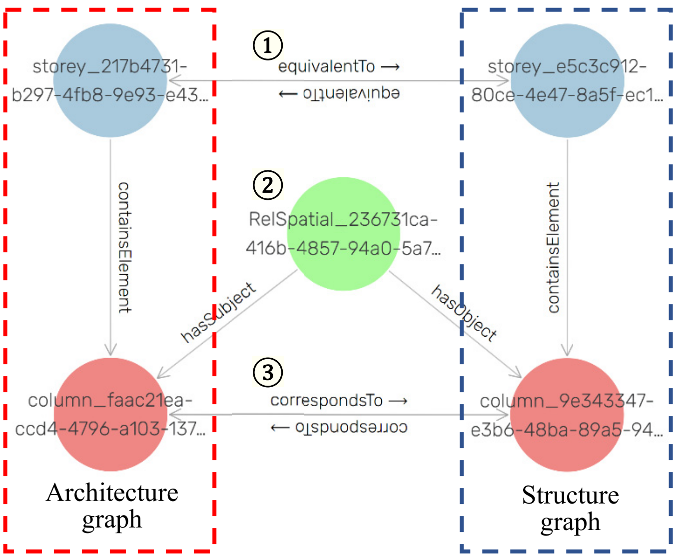

Figure 5. Association semantic enrichment results viewed in GraphDB. The enrichment proceeded in the indicated sequence.

architecture and the structure named graphs respectively, in the form of new datatype properties to the building object entities. On the other hand, the new association semantics are stored in the CBIM Graph, including the topological relationships between the building objects, as well as their inferred correspondence relationships that explicate the design intents.

```
https://example.org/domains/architecture

arc:storey_217b4731-b297-4fb8-9e93-e43ee72f3cb9 a bot:Storey ;
    bot:containsElement
        arc:column_faac21ea-ccd4-4796-a103-137652987219 ;

arc:column_faac21ea-ccd4-4796-a103-137652987219 a beo:Column,
        beo:Column-COLUMN,
        bot:Element ;
    fog:asPly
        "arc_geom/faac21ea-ccd4-4796-a103-137652987219.ply" ;
    bot:hasSimple3DModel
        "-9352.0, -22785.0, 7600.0, 300.0, 300.0, 3800.0" .
    ...
```

```
https://example.org/domains/structure

str:storey_e5c3c912-80ce-4e47-8a5f-ec197cf1b30f a bot:Storey ;
    bot:containsElement
        str:column_9e343347-e3b6-48ba-89a5-94b31bdfd30e,

str:column_9e343347-e3b6-48ba-89a5-94b31bdfd30e a beo:Column,
        beo:Column-COLUMN,
        bot:Element ;
    fog:asPly
        "str_geom/9e343347-e3b6-48ba-89a5-94b31bdfd30e.ply" ;
    bot:hasSimple3DModel
        "-9427.0, -22860.0, 7600.0, 450.0, 450.0, 3400.0" .
    ...
```

```
https://example.org/domains/CBIM

arc:storey_217b4731-b297-4fb8-9e93-e43ee72f3cb9
    CBIM:equivalentTo
        str:storey_e5c3c912-80ce-4e47-8a5f-ec197cf1b30f .

CBIM:RelSpatial_37de1833-79c5-4118-9044-cf9b44fc83a8 a
        CBIM:RelSpatial ;
    CBIM:hasSubject
        arc:column_faac21ea-ccd4-4796-a103-137652987219 ;
    CBIM:hasObject
        str:column_9e343347-e3b6-48ba-89a5-94b31bdfd30e ;
    CBIM:topology
        "CONTAINED_IN, CONTAINED_IN, CONTAIN" ;
    CBIM:offset "0, 0, 0" ;
    CBIM:inContact true .

arc:column_faac21ea-ccd4-4796-a103-137652987219
    CBIM:correspondsTo
         str:column_9e343347-e3b6-48ba-89a5-94b31bdfd30e .
```

Snippet 1. TriG fragment for the case study graph.

5 DISCUSSION

5.1 *Applications of the Enriched Semantics*

In an early application, Le & Jeong (2016) demonstrated that linking data from different sources can support decision-making in highway management when used in combination with graph query and reasoning techniques. Moreover, Törmä (2013) claimed that cross-domain change maintenance can be achieved on the linked graphs, which to-date has not been implemented. One of our ongoing experiments is to test the feasibility of such a function on the linked multidisciplinary building semantic graphs. Specifically, we envisage that when one participant makes a change to an element in a specific partial model, this information would be propagated through the established *CBIM:correspondsTo* relationships to the related BIM entities in other domains. Accordingly, these elements should be updated in the corresponding federated graphs, then in their native BIM software. Naturally, this would require discipline-specific bi-directional communication between the graph database and the BIM authoring tools.

5.2 *Richer Semantics*

In this research, we have demonstrated the feasibility of the proposed semantic enrichment pipeline with the *CBIM:correspondsTo* relationship, which is a relatively simple semantic between any pair of objects. More thorough research is needed to discover the pool of inter-domain object semantics and to develop appropriate methods to instantiate them. For example, to implement the change maintenance function mentioned above, one may realize the need for explicit constraint relationships between the objects to make their intended interactive behaviours under the dynamically changing design environment explicit. Whenever a violation of a constraint is detected as a result of a design change from one party, the CBIM system will be triggered and react accordingly to alert other users and to restore integrity.

The CBIM ontology can also be extended by other ontologies to unlock more functionality. For example, incorporating the OPM ontology (M. Rasmussen et al., 2018) would enable the tracking of design properties through time, bringing in features like design alternative management and version control that are essential for a fully functional multi-disciplinary design coordination system.

5.3 *Connectors between BIM editors and CDE*

As no BIM editors to date support the direct export of graph models natively, the current workaround solution is to convert the BIM models first to IFC files and then to LBD graphs. Manual operations are still an integral part of the workflow, making it sub-optimal. Some third-party plugins that export models from Revit to LBD graphs within the software interface do exist in academia (Pauwels, 2020; Tchouanguem Djuedja et al., 2021). However, such tools 1) ignore the non-semantic information (like geometry) contained in the model, and 2) export information in batches, which cannot capture the dynamically changing behaviour within the design models.

A better approach would be to devise connectors that directly connect BIM editors to the vendor-neutral graph CDE, so that both semantic and non-semantic information from the native models can be composed into their appropriate representations and linked in the CDE. All these should happen in real time, automatically. This too is the subject of ongoing work.

6 CONCLUSIONS

This paper presents a novel algorithmic approach to enrich the association semantics among federated BIM models hosted under a Linked-Data based Common Data Environment. A case study showcasing the dual representation of a column in federated design models was used to validate the presented approach. The algorithm successfully captured and established the correspondence relationship between the column entities from distinct model representations, making the association explicit to all project participants. This is the first known study that demonstrates the feasibility of automatically associating multidisciplinary building models at object levels.

7 ACKNOWLEDGEMENT

This research project is receiving funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860555 for the CBIM project.

REFERENCES

Belsky, M., Sacks, R., & Brilakis, I. (2016). Semantic Enrichment for Building Information Modeling. *Computer-Aided Civil and Infrastructure Engineering*, *31*(4), 261–274.

Berners-Lee, T. (2006). *Linked Data—Design Issues*. https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43.

Bloch, T., & Sacks, R. (2018). Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models. *Automation in Construction*, *91*, 256–272.

Bloch, T., & Sacks, R. (2020). Clustering information types for semantic enrichment of building information models to support automated code compliance checking. *Journal of Computing in Civil Engineering*, *34*(6), 04020040.

Bonduel, M., Wagner, A., Pauwels, P., Vergauwen, M., & Klein, R. (2019). Including widespread geometry formats in semantic graphs using RDF literals. *Proceedings of the 2019 European Conference for Computing in Construction*, 341–350.

Collins, F. C., Braun, A., Ringsquandl, M., Hall, D. M., & Borrmann, A. (2021). Assessing ifc classes with means of geometric deep learning on different graph encodings. *Proc. of the 2021 European Conference on Computing in Construction*.

Dawson-Haggerty, M. (2022). *Trimesh* [Python]. https://github.com/mikedh/trimesh

GraphDB. (2021). *GraphDB Semantic Graph Database*. https://graphdb.ontotext.com/

IfcOpenShell. (2022). [C++]. IfcOpenShell. https://github.com/IfcOpenShell/IfcOpenShell

ISO. (2018). *ISO 19650-1, 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*.

Jensen, S. K., Pedersen, T. B., & Thomsen, C. (2017). Time Series Management Systems: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, *29*(11), 2581–2600.

Le, T., & Jeong, H. D. (2016). Interlinking life-cycle data spaces to support decision making in highway asset management. *Automation in Construction*, *64*, 54–64.

Malcolm, A., Werbrouck, J., & Pauwels, P. (2021). LBD server: Visualising Building Graphs in web-based environments using semantic graphs and glTF-models. In *Formal Methods in Architecture* (pp. 287–293). Springer.

Oraskari, J. (2022). *IFCtoLBD* [Python]. https://github.com/jyrkioraskari/IFCtoLBD

Pauwels, P. (2014). Supporting decision-making in the building life-cycle using linked building data. *Buildings*, 4(3), 549-579.

Pauwels, P. (2020). *Revit LBD Exporter* [C#]. https://github.com/pipauwel/revit-bot-exporter

Pauwels, P., Costin, A., & Rasmussen, M. H. (2022). Knowledge Graphs and Linked Data for the Built Environment. In M. Bolpagni, R. Gavina, & D. Ribeiro (Eds.), *Industry 4.0 for the Built Environment* (Vol. 20, pp. 157–183). Springer International Publishing.

Pauwels, P., & Roxin, A. (2017). SimpleBIM: From full ifcOWL graphs to simplified building graphs. *EWork and eBusiness in Architecture, Engineering and Construction*.

Pauwels, P., & Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*, *63*, 100–133.

Preidel, C., Borrmann, A., Oberender, C., & Tretheway, M. (2015). *Seamless Integration of Common Data Environment Access into BIM Authoring Applications: The BIM Integration Framework*. 9.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., & Pauwels, P. (2020). BOT: The building topology ontology of the W3C linked building data group. *Semantic Web*, *12*(1), 143–161.

Rasmussen, M., Lefrançois, M., Bonduel, M., Hviid, C., & Karlshøj, J. (2018). OPM: An ontology for describing properties that evolve over time. *6th Linked Data in Architecture and Construction Workshop*.

Sacks, R., Eastman, C. M., Lee, G., & Teicholz, P. (2018). *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors and Facility Managers* (3rd ed.). John Wiley and Sons.

Sacks, R., Wang, Z., Ouyang, B., Utkucu, D., & Chen, S. (2022). Toward Artificially Intelligent Cloud-Based Building Information Modeling for Collaborative Multidisciplinary Design. *Advanced Engineering Informatics*, 53, 101711.

SPARQLWrapper. (2022). [Python]. RDFLib. https://github.com/RDFLib/sparqlwrapper

Tchouanguem Djuedja, J. F., Abanda, F. H., Kamsu-Foguem, B., Pauwels, P., Magniont, C., & Karray, M. H. (2021). An integrated Linked Building Data system: AEC industry case. *Advances in Engineering Software*, *152*, 102930.

Törmä, S. (2013). Semantic linking of building information models. *Proceedings - 2013 IEEE 7th International Conference on Semantic Computing, ICSC 2013*, 412–419.

Wang, Z., Sacks, R., & Yeung, T. (2022). Exploring graph neural networks for semantic enrichment: Room type classification. *Automation in Construction*, *134*, 104039.

Wang, Z., Yeung, T., Sacks, R., & Su, Z. (2021). Room Type Classification for Semantic Enrichment of Building Information Modeling Using Graph Neural Networks. *Proc. of the Conference CIB W78*, *2021*, 11–15.

Werbrouck, J., Pauwels, P., Beetz, J., & Mannens, E. (2022). *LBDserver-a Federated Ecosystem for Heterogeneous Linked Building Data*. [Submitted for publication].

Werbrouck, J., Pauwels, P., Beetz, J., & van Berlo, L. (2019). Towards a decentralised common data environment using linked building data and the solid ecosystem. *36th CIB W78 2019 Conference*, 113–123.

Xue, F., Lu, W., & Chen, K. (2018). Automatic Generation of Semantically Rich As-Built Building Information Models Using 2D Images: A Derivative-Free Optimization Approach: Automatic generation of semantically rich as-built building information models using 2D images. *Computer-Aided Civil and Infrastructure Engineering*, *33*(11), 926–942.