Reasoning in Common Data Environments

Re-thinking CDEs to enhance collaboration in BIM processes

Davide Simeone[1], Stefano Cursi[2], Ugo Maria Coraglia[3], Antonio Fioravanti[4]
[1,2,3,4]Sapienza University of Rome, Department of Civil, Construction and Environmental Engineering
[1,3,4]{davide.simeone|ugomaria.coraglia|antonio.fioravanti}@uniroma1.it
[2]stefanocursi@hotmail.it

In BIM processes, the concept of Common Data Environment - CDE - has often been depicted as a key element for successful collaboration and information sharing among different actors but, in current practice, acts as a mere documentation repository ineffective for true collaborative purposes. Therefore, the idea of CDE seems to be overrated on the one hand and unexploited on the other, while effective collaboration is still far from being decisively supported. To overcome this lack, the present research focuses on the definition of a new generation of CDEs, enhanced with an information level for knowledge integration provided by different information carriers such as models and datasets. The paper discusses its development through a graph database platform and dedicated methodologies for data retrieval and query, to verify coherence and consistency of information among different models.

Keywords: *Collaboration in AEC processes, Common Data Environment, Graph Databases, Building Information Modeling, Queries and data retrieval*

INTRODUCTION

In current BIM-enabled AEC processes, Common Data Environment - CDE- has been considered as a key element for successful collaboration and information sharing among different actors and stakeholders. The UK government strategy implies its adoption within the BIM Level 2 scope and some administrations are progressively mandating its implementation in AEC projects. At present, despite a large amount of CAAD research in the field of Collaborative AEC Design process, the most spread structure of CDE is the one proposed in the PAS 1192 (2013) and ISO 19650 (2018) norms, composed of different areas (WIP, Shared, Published, Archived) where project documentation is made selectively accessible to different stakeholders. Relying on this structure, some software industries have proposed their platforms, usually cloud-based, such as ProjectWise (Bentley), Aconex and, more recently, Autodesk BIM 360. As we look at CDEs how it is described in these norms and it has been implemented a big ambiguity, at least from the CAAD research perspective, arises: are we dealing with data or are current CDEs acting as documentation, although shared, repositories? And, as a corollary, is this CDE structure effective for collaboration purposes? At present, full potentials of the idea of CDE seem to be unexploited, while effective collaboration among AEC processes actors is still far from being decisively supported by CDEs. In the context, this research investigates new methodologies for data, information and knowledge unified formalization in CDE, based on graph databases, to improve coherence, reliability and accessibility of information stored and shared, as well as improving quality of the integration of models, documents and other information carriers that contributes to the central core of data acting as a reference. It essentially focuses on the addition of an information level in current CDEs and discusses the use of graph databases as a way to successfully implement and manage it.

STATE OF THE ART

In the eCAADe scope and, more widely, in the CAAD and BIM world, much research has been focusing on the integration of Building Information Modeling, information and knowledge ontologies (Cursi, 2017) and linked data as a way to drive, standardize and extend content and semantics of AEC informative models. Potentials of this approach have been demonstrated by research works like the one from Beetz (2005), Pauwels (2015) with a progressive extension of the application to built heritage field (Simeone, 2019) (Di Mascio, 2013). In this field, the development of IFC-OWL ontology has represented an important shift, introducing a reference ontology for AEC entities formalization for interoperability purpose and as an extendable base for domain-specific semantic enrichment of building information models. Similarly, some research has focused on the adoption of information ontologies as an approach to improve knowledge sharing and collaboration between different actors and disciplines (Jelokhani-Niarak, 2018) (Zhong, 2018). This is particularly relevant for this research since it targets one of the main intended features of CDEs: collaboration through comprehensive, coherent and interrelated knowledge (Carrara et al., 2017). If compared to this research direction, research in the CDE field has been quite static. The concept of CDE was created in the IT field as a centralized, unique set of data for any kind of project that requires collaboration and mutual accessibility to data. Later, the concept of CDE has shifted to the AEC field to define the single source of truth for design development, shaped accordingly to major steps in the delivery of an AEC project. On one side, its basic structure and functioning rely on the original framework provided by BS PAS 1192 norms, while research has mainly investigated procedural aspects of CDEs with only partial attention to methods and tools to enhance and ensure the quality, interoperability and richness of the stored content. On the other side, AEC software houses have spent much effort on CDE development, often to improve integration with their proprietary authoring software or to provide scalable, cloud-based solutions for AEC projects. As a result, CDEs are currently used as documents storage environments, with additional metadata progressively introduced to solve documents-controlling common issues such as accessibility, revisioning, file codification, issuing dates, etc. In the last years, the development of tools such as Autodesk Forge has made possible to extract and elaborate stored information directly in the cloud environment, also to cross-link and compare different models (Yan, 2017), (Preidel, 2016). Nevertheless, this approach can be only considered a palliative that relies on the extraction of data, while a few efforts have been made regarding shaping and structuring data, information and knowledge - modelled in the bim model or adjunct in a database - to ensure horizontal reasoning in the CDE. This lack in current CDEs is progressively emerging in the AEC practice, and linked data methodology represents a potential solution that needs to be further investigated (Werbrouck, 2019).

INFORMATION GRAPH DATABASES FOR NEXT-GEN COMMON DATA ENVIRONMENTS

In CDEs, two main levels can be usually depicted: 1) the information carrier level, and 2) the data/information level. The first one comprises the different kinds of documents, models, reports, calculation notes, etc. that support data formalization and representation and that are usually produced, uploaded, edited, and accessed by different actors. The second level, instead, refers to all elementary data and information, formalized in the information carriers following discipline-specific methodologies and representation structures. By analyzing current CDEs applications, it is clear how CDE platforms mainly focus on the information carrier level, in some cases specifying a simplified set of horizontal relationships among different documents to indicate which of them are related or "linked". No coherence and actual correspondence among data stored in those information carriers are declared, checked, and ensured, and, therefore, CDEs usually result in heterogenous, incoherent, and unreliable information storage. While information ontologies have shown great potential in abstracting entities and relationships, at present data and information are currently managed in BIM software through relational databases that make the entire data structure too rigid and difficult to manage for AEC design process, in particular when those databases are structured based on proprietary software requirements. This issue has been discussed in the IT world, where an alternative approach, based on no-SQL databases, and in particular graph databases, are progressively being adopted, in particular for knowledge formalization and creation of *data lakes* that deal with complex *data patterns* rather than *single data*. In this context, this research proposes and assesses the use of graph databases as a way to access and make homogenous, flexible and queryable the data provided by different information carriers in the CDE, in an integrated way. Among the different advantages of the adoption of graph databases over RDMS (Relational Databases Management System), three are particularly relevant for our scope: 1) focus on relationships, that allows users to depict them from each model and to create new relationships between different entities of different datasets; 2) flexibility, that allows the same data model to be applied to the different dataset stored in a CDE; 3) easy evolution on time, that can adapt to the evolution of the project models better than any RDMS, usually static after the formalization of the conceptual model.

THE ENHANCED COMMON DATA ENVIRONMENT SCHEMA

In this proposed framework for a new generation of CDEs, this information level developed through graph databases is mainly an environment where information, derived from different information carriers through a process of data extraction, selection, and translation, is represented in a homogenous way, structured and compared. To provide this required homogeneity, the proposed Information Level has been conceived and implemented relying on the use of Information ontologies, while a general ontology structure, based on previous research developed by this research group, serves as an ontology template for re-formalization of data extracted by the different Information Carriers (Simeone, 2019). As described in figure 1, the passage from the heterogeneous formalization of the information carriers to the homogeneous and common Information level requires two crucial passages:

1. Data selection and extraction;
2. Data re-formalization and re-structuring in graph databases connectable and queryable.

In the proposed system, the first step depends on different methodologies following the variety of information carriers and to the availability of specific applications and tools. In our experiments, we relied on Revit DBLink application for the extraction of data from models and Dynamo for extraction of specific datasets (Simeone, 2017), while for other information carriers, such as reports or calculation notes, we con-

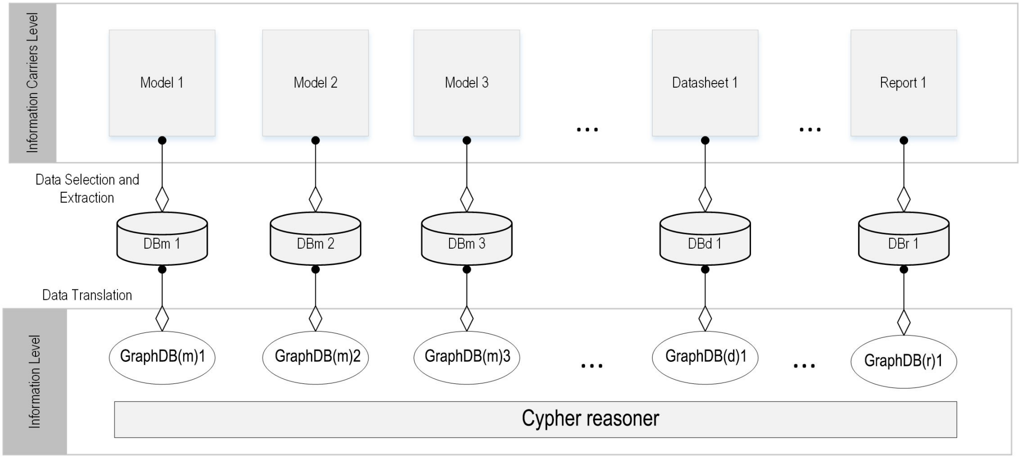

Figure 1
The proposed new CDEs schema composed of the Information Carriers level and the Information Level.

sidered useful using synthesis datasheets linked to actual documents. Because of the wide number of information stored, this extraction activity is coupled with a filtering feature that selects only specific data required by the information level. Although this process could be potentially applied to all information formalized, limiting the amount of information formalized is functional to the necessity of having a lean system to ensure qualitative coherence of information while delegating more strict and accurate data coherence and consistency verification (i.e. clash detection) to other applications outside the Common Data Environment. The second passage is instead based on the re-organization of the extracted and selected data in a Common graph database that ensures information to be formalized in a way to allow check, comparison, validation among different datasets (Fig.1).

Each model template requires a specific translation method depending on the BIM authoring tool and its database structure. This aspect only partially hinders the process because of the simplicity of the translation methods and their re-usability in similar AEC processes. This two steps process translates information heterogeneously provided in different information carriers into a single, homogenous knowledge base, hence allowing reasoning operations as well as information comparison and general coherence checking. The use of common representation standards, such as the use of IFC owl template (Pauwels, 2015), and agreed formalization methodologies, such as schedules templates and synthesis data structures, deeply affects efficiency and speed of the process. To operate on the Information Level, a Cypher-based reasoner has been integrated into the proposed system. The reasoner allows users to develop and execute reasoning queries to verify and compare data stored in the information level. For instance, it allows verifying if the number of spans of a bridge as design in a BIM model is coherent with the data stored in its calculation report, alerting the design team in case of incoherencies and discrepancies. Specific sets of reasoning rules can be used for specific features of the design to be checked directly in the CDE, improving actors' awareness of the state of the information shared in the collaboration platform.

THE IMPLEMENTATION OF THE INFORMATION LEVEL THROUGH GRAPH DATABASES

To manage, query and capitalize the data introduced in the CDE by different information carriers, this research developed the information level relying on the use of the graph database technology, implemented through the Neo4J platform (Francis, 2018). In this way, it is possible to have a homogenous way of formalization of information extracted from or connected to the different models provided. Each model provides a set of information that can be organized in a network of entities, relationships, and properties that can be retrieved and used in simple or complex queries. In the case of BIM models, usually organized through the family->type->instance structure, the graph databases nodes are the instances of the model (the ones usually defined by an ID in the models, such as a specific pillar, a column, etc.) while the labels provide the family/type classification. Relationships among nodes represent the constraints and the link between entities in the BIM Model, such as a floor constrained to a specific level (fig. 2 and fig. 3).

In the case of other models, such as planning schedules, Quantity Take Off models, or 4D models, it is usually possible to directly access to modelling entities (i.e. tasks, take off items, etc.) and their connecting relationships, recreating graph databases. A bit more difficult is the extraction of data from other information carriers such as reports, etc. In that case, it is necessary to have at least the relevant data sets to translate them in a different data graph. At first, the graph databases elaborated from the different models reside in the information level as independent, coherent graphs but, since they are formalized homogeneously, it is possible to construct queries and algorithms to 1) run across in a transversal direction between different graphs, 2) access and verify the information and, if conditions are verified, generate new relationships between nodes/entities in different graphs. For instance, if the same building component (i.e. a beam) is represented in different models and there is a way to recognize the correspondence (such as the same ID, or the reference to the same structural grid point) it is possible to create a relationship corresponding to and start connecting the two graphs.

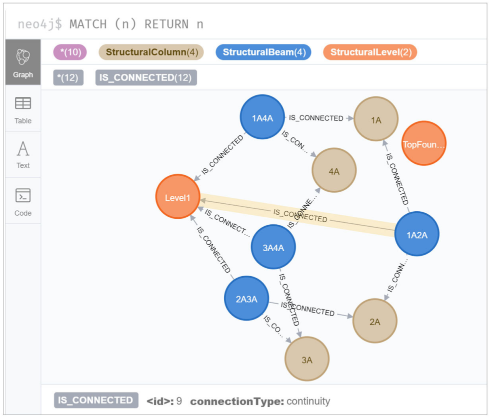

Figure 2
A portion of the tested graph database that shows entities (nodes) and relationships. Labels represent family/type classes as per usual BIM elements schema.

"n"
{"xPosition":5.5,"xDimension":0.5,"yPosition":0.0,"yDimension":0.7,"Height":3.0,"reinforcementRate":270,"gridReference":"3A","id":1573,"information_source":"StrModel_054"}
{"xPosition":0.0,"xDimension":0.5,"yPosition":0.0,"yDimension":0.7,"Height":3.0,"reinforcementRate":270,"gridReference":"4A","id":1574,"information_source":"StrModel_054"}
{"xDimension":0.3,"Length":5.5,"Height":1,"id":1600,"reinforcementRate":240,"gridReference":"2A3A","information_source":"StrModel_054"}
{"xPosition":5.5,"xDimension":0.5,"yPosition":6.0,"yDimension":0.7,"Height":3.0,"reinforcementRate":270,"gridReference":"1A","id":1572,"information_source":"StrModel_054"}
{"xPosition":0.0,"xDimension":0.5,"yPosition":6.0,"yDimension":0.7,"Height":3.0,"reinforcementRate":270,"gridReference":"2A","id":1571,"information_source":"StrModel_054"}
{"xDimension":0.3,"Length":5.5,"Height":0.7,"reinforcementRate":240,"gridReference":"1A2A","id":1601,"information_source":"StrModel_054"}
{"xDimension":0.3,"Length":6.0,"Height":1,"id":1602,"reinforcementRate":240,"gridReference":"1A4A","information_source":"StrModel_054"}
{"xDimension":0.3,"Length":6.0,"Height":0.8,"reinforcementRate":240,"gridReference":"3A4A","id":1603,"information_source":"StrModel_054"}
{"Elevation":3.0,"LevelDef":"Level1","id":"L01"}
{"Elevation":0.0,"LevelDef":"TopFoundation","id":"L02"}

Figure 3
The same portion of the graph database in fig.2, expressed as a table to show the formalization of properties of entities as extracted by the BIM models.

```
neo4j$ MATCH (x:StructuralBeam), (y:StructuralColumn) where (x.gridReference
       contains y.gridReference) and (x.information_source =
       y.information_source) CREATE (x)-
       [r:IS_CONNECTED{connectionType:'continuity'}]→(y) RETURN x, y, r
```

Figure 4
An example of a constructed query in Cypher language, that retrieves structural elements ("MATCH" part) and generates new relationships based on the structural grid ("CREATE" part).

For this research, we decided to rely on the use of Cypher, a declarative query language that enables both queries and data formalization, specifically designed to operate in graph databases environment and interact with nodes, relationships, properties, and patterns. In this proposed evolution of CDEs, the creation and control of 'horizontal queries' between graph databases represent a new task in the current AEC information management processes, that can impact the current role of the project information manager. In this task, the set of queries necessary to connect different data models stored in the CDE is project-dependent because it has to be tailored in accordance to the scope of each model, the formalization structure of its data, and the objectives of the query itself. If compared to other query languages (i.e. the SWRL - Semantic Web Rule Language), Cypher supports this process of queries set customization using its syntax based on natural language and its very linear logic. As in the example shown in figure 4, the use of keywords such as MATCH and CREATE allows structuring complex queries to recognize patterns in the graph databases and generate new data, providing a query template that can be adapted to similar necessities in other projects. The development and refining of query sets for AEC projects is an open topic that can introduce a new standardization level in the current practice.

For data extraction from BIM Models, and their connection to the graph databases in the information level, different solutions can be conceived relying on approaches and tools already available in the AEC sector. In the case of Autodesk Revit models, we utilized two different methodologies that rely on the use of the Revit DBLink and the use of Dynamo algorithms for data extraction in tables. We found the DBLink approach useful in the case of necessity of a bulk translation in a graph of the entire model, although some issues arise in terms of manageability and conversion from the relational database to the graph one. The use of dynamo scripts, instead, allows for a selective extraction that can be tailored following the query necessities of each project. Both these methods provide a bidirectional connection between the models and the dataset elaborated in the CDE information level. Current Visual programming approaches are a relevant contribution to this process, allowing to select, filter, and export datasets in formats such as .csv that allow graph databases editor to access data and automatically reconstruct the network of entities. Another opportunity, at present only available in the exporting direction, is represented by the use of the IFC schema. In this case, the use of the IFC-OWL ontology (Pauwels, 2015) is a reference for the implementation of standardized procedures for its conversion in a graph database, and the construction of queries accessing the IFC data (fig. 5).

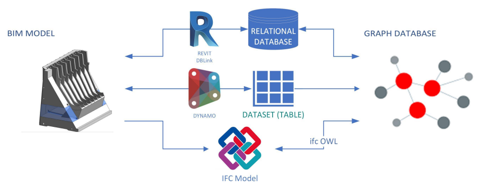

Figure 5
Interoperability between BIM Models and Graph databases in the CDE Information Level, based on different data exchange methodologies (RDMS, Dynamo, IFC schema).

POTENTIALS AND OPEN PROBLEMS OF GRAPH DATABASES APPROACH TO COMMON DATA ENVIRONMENTS

In our experiments on the adoption of graph databases as a new way of formalizing and make available information in AEC Common Data Environments, some potentials and limits have emerged, not specifically related to technical aspects but rather to the introduction of this approach within already consolidated workflows and protocols. The processes described in this paper underlay the more general information management protocols and practices that drive current AEC projects, essentially dealing with how to formalize information and knowledge. From an end-user perspective, it is not possible to define a unique workflow but we rather suggest that some additional actions have to be carried out, while sharing documents and models, to favour the development of a coherent and efficient Single Source of Truth as necessary in AEC projects. The data translation from models - or other information carriers - to graph databases is an activity not yet considered in current AEC collaboration processes, that requires efforts, competencies and time, especially if performed cyclically as in architectural design. Heterogeneity of models is a hindering factor that requires the development of customized algorithms and protocols for data extraction, impacting the adoption of this new approach to the entire project dataset. Another obstacle is represented to the necessity of new figures in the process, able to manage and query this new CDEs: the graphs themselves can easily become too complex, and database experience is, therefore, necessary to control that complexity and ensure efficacy in data retrieval. At the same time, we assessed relevant improvements on the quality, consistency and coherence of the information formalized through graph databases, as well as in its understanding by the different actors involved in the process. Simplified interfaces, for instance, allow the actors to explore the data graphs reaching the needed information and being supported in its interpretation by visualizing the connected information. We also noticed that the effectiveness of this approach is higher if not applied to the single elements that compose the building but rather operating at a higher level of representation where decisions are made on systems and general aspects of the building, while highly detailed, discipline-specific issues are usually more manageable through dedicated tools. This approach also provides a relevant improvement in terms of queries capabilities: horizontal queries allow to move across the different data formalized in the different models, while the Cypher language allows to perform also complex systems of queries and to elaborate a higher level of semantics, usually not accessible by direct operations on the single models.

CONCLUSIONS AND FUTURE PROSPECTS

The research described in this paper aims at enhancing the current Common Data Environment for BIM-oriented processes, by integrating the CDE repository structure with a knowledge-based system, developed through graph databases. Its scope is to extend CDEs usual repository function with the possibility to perform qualitative checking and evaluation of the data stored in the different models and information carriers, immediately detecting potential incoherence and improve design integration. The conceptual framework of the presented system relies on a semi-automated process of data extraction, selection and translation into a homogeneous set of graph databases that allows reasoning operation and data accessibility. As shown by first tests in simulated AEC processes, the proposed system is potentially able to enhance the efficacy of a crucial element of BIM processes - the Common Data Environment - by transforming it from a repository of not-coherent documentation to an integrated data environment oriented to actual collaboration. Also, the introduction of CDE information level developed through graph databases allows for enhanced flexibility in terms of queries and verifications that can be performed by connecting and moving across different models datasets, without constraining the users to rely only on pre-defined metadata controls as in current CDEs. The presented approach, although still explorative, opens the research to further developments and even integration with other research streams that could be beneficial for the entire sector. In particular, research is still needed to make the entire workflow more efficient by reducing the variability of stored data and the re-usability of queries. The use of ontology-based schemas such the IFC-OWL could ensure a better homogeneity of the formalized data and, at the same time, allowing for a standardized library of queries. Another aspect, not often discussed in our field, is related to the integration of information sources different from models, that are currently attached as external documents: further research should focus on re-thinking such way of communicating a project, formalizing information with different technologies and ensuring automation in data elaboration. Another research step to be considered is the use of machine learning techniques for queries and data retrieval, further enhancing the efficacy of the system. To conclude, in the scope of current research AEC processes research, this work contributes to the current quest for a new generation of collaborative digital environments, able not only to make data and information accessible but also to integrated them, ensuring their coherence and their comprehension to all the actors involved in the process.

REFERENCES

Beetz, J, Van Leeuwen, J and De Vries, B. 2005 'An ontology web language notation of the industry foundation classes', *Conference on Information Technology in Construction, , CIB-W78*, Dresden, p. 193–198

BS, PAS 1192:2 2013 (eds) 2013, - *Specification for information management for the capital/delivery phase of construction projects using building information modelling*, BSI, London

Carrara, G, Fioravanti, A, Loffreda, G and Trento, A 2017, *Knowledge collaboration design. Theory techniques and applications for collaboration in architecture*, Gangemi Editore, Rome

Cursi, S, Simeone, D and Coraglia, UM 2017 'An ontology-based platform for BIM semantic enrichment', *ShoCK – Sharing of Computable Knowledge - Proceedings of the 35th eCAADe Conference*, Rome, pp. 649-656

Francis, N, Green, A, Guagliardo, P, Libkin, L and Lindaake, T 2018 'Cypher: An Evolving Query Language for Property Graphs', *SIGMOD'18 Proceedings of the 2018 International Conference on Management of Data, Jun 2018*, Houston, p. 1433

ISO, 19650-1:2018 2018, *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*, ISO-International Organization for Standardization

Jelokhani-Niaraki, M 2018, 'Knowledge sharing in Web-based collaborative multicriteria spatial decision analysis: An ontology-based multi-agent approach', *Computers, Environment and Urban Systems*, 72, pp. 104-123

Di Mascio, D and Pauwels, P 2013 'Improving the knowledge and management of the historical built environment with BIM and ontologies. The case study of the book tower', *CONVR2013 (eProceedings of the 13th International Conference on Construction Applications of Virtual Reality)*, London, pp. 427-436

Pauwels, P and Terkaj, w 2015, 'EXPRESS to OWL for Construction Industry: Towards a Recommendable and Usable ifcOWL Ontology', *Automation in Construction*, 63, pp. 100-133

Preidel, C, Borrmann, A, Oberender, CH and Tretheway, M 2016 'Seamless integration of common data environment access into BIM authoring applications: The BIM integration framework', *E-work and E-business in architecture, engineering and construction*

Simeone, D and Cursi, S 2017 'A Platform for Enriching BIM Representation Through Semantic Web Technologies', *LC3 2017: Volume I - Proceedings of the Joint Conference on Computing in Construction (JC3)*, Heraklion, pp. 423-430

Simeone, D, Cursi, S and Acierno, M 2019, 'BIM semantic-enrichment for built heritage representation', *Automation in Construction*, 97, pp. 122-137

Werbrouck, J, Pauwels, P, Beetz, J and van Berlo, L 2019 'Towards a Decentralised Common Data Environment using Linked Building Data and the Solid Ecosystem', *Proceedings of the 36th CIB W78 Conference on Information Technology in Construction*

Yan, W 2018 'WP-BIM: Web-based Parametric BIM Towards Online Collaborative Design and Optimization', *ShoCK! - Sharing Computational Knowledge! - Proceedings of the 35th eCAADe Conference*, Rome, pp. 527-534

Zhong, B, Ganab, C, Luo, H and Xing, x 2018, 'Ontology-based framework for building environmental monitoring and compliance checking under BIM environment', *Building and Environment*, 141, pp. 127-142