

# THE CRITICAL ROLE OF ACCESSIBLE DATA FOR BIM-BASED AUTOMATED RULE CHECKING SYSTEMS

Wawan Solihin<sup>1</sup>, Johannes Dimiyadi<sup>2</sup>, Yong-Cheol Lee<sup>3</sup>, Charles Eastman<sup>4</sup>, and Robert Amor<sup>5</sup>

**Abstract:** This paper proposes a concept of an accessible BIM database that supports integration with geometry enabling simplified and efficient queries of the IFC-based building model. The simplified schema, BIMRL, is shown to be significantly effective for the purpose of an implementation of an automated BIM-based rule checking system. The schema has been shown to successfully work in both traditional RDBMS and the NoSQL graph database. It complements a missing piece in the current research of automated rule checking, which mostly focuses on the formulation and representation of computable rules involving logic, checking algorithm, and parameterization. Even though these present approaches have largely assumed that data is available and easily accessible from a building model, this assumption is typically infeasible in a real-world implementation. Building rules require not only base data explicitly available in the model but also higher level semantic concepts that typically involve multiple relationships and spatial operations, which cannot be captured explicitly in the model. Without addressing this issue, a rule checking system will severely underperform and will be filled with opaque algorithms that act as black-boxes.

**Keywords:** BIM, automated rule checking, IFC, BIMRL, Big Data.

## 1 INTRODUCTION

Research into automated rule checking systems has had continuous interests since Fenves pioneered the concept in the 1960s with a decision table (Fenves 1966). Since then, various research efforts for enhancing an automated rule-checking process in the broad disciplines have come and gone. Despite these efforts, the AEC-FM industries have made little progress towards a reliable, robust, and executable automated rule checking system. Even with the wider adoption of BIM in these industries today, there is only one commercial off-the-shelf product available, i.e. Solibri Model Checker (SMC) (Solibri). Besides SMC, and there are a few options that require much higher level of customization of the rule-checking system. EDM ModelServer offers a toolkit that allows the user to write a script for a rule using its ruleSchema directly on the IFC express model (EPM 2014). Revit Model Review offers a proprietary tool for the user to define rules on Revit models. FORNAX provides a toolkit to develop rules using C++ APIs to run on its proprietary internet server

---

<sup>1</sup> Senior Software Architect, Invicara Pte. Ltd., Singapore, [wawan.solihin@gatech.edu](mailto:wawan.solihin@gatech.edu)

<sup>2</sup> Post-doctoral Researcher, Department of Computer Science, University of Auckland, New Zealand, [j.dimiyadi@auckland.ac.nz](mailto:j.dimiyadi@auckland.ac.nz)

<sup>3</sup> Assistant Professor, Bert S. Turner Department of Construction Management, Louisiana State University, Baton Rouge, LA, USA, [ycheol@lsu.edu](mailto:ycheol@lsu.edu)

<sup>4</sup> Professor, School of Design, Georgia Institute of Technology, Atlanta, USA, [eastman@design.gatech.edu](mailto:ceastman@design.gatech.edu)

<sup>5</sup> Professor, Department of Computer Science, University of Auckland, New Zealand, [trebor@cs.auckland.ac.nz](mailto:rtrebor@cs.auckland.ac.nz)

(NovaCityNets 2016). And very recently a new product from a startup company Invicara Inc., BIM Assure (Invicara 2016), offers a tool to define rules using user-friendly templates to run on its cloud service. Its current capability supports Revit and checks on properties and some relationships of a building model. With the exception of SMC, none of the existing tools deals with the geometry and spatial operations that frequently adorn the BIM-based rules, especially in building codes (Solihin and Eastman 2015).

This paper looks into recent research in automated code checking and discusses the missing piece that is largely overlooked, i.e. the building data and its accessibility. Particularly, this paper argues that the role of data is critical if we would like to achieve a practical and workable automated rule checking system. We propose a simplified schema approach called BIMRL to address this current gap. It has been implemented using both relational database and a NoSQL multi-mode and graph database.

## 2 CURRENT RESEARCH IN AUTOMATED BIM BASED RULE CHECKING

Research into automated rule checking has intensified in recent years. The CIB W78 conference in 2015 featured a special track dedicated to this topic, followed by a special issue in the Journal of Information Technology in Construction (ITcon) in November 2016 (ITcon 2016). A recent effort within the buildingSMART Regulatory Room also focuses on a commercially implementable approach to the automated code checking. Given that the subject of automated rule checking encompasses a wide range of topics, the targeted scope of proposed checking approaches is often unclear. Claiming a solution for an automated rule checking system may be too broad ignoring other parts of the subject matter that may be as complex. A simple classification to identify the scope of a particular approach to the automated rule checking system from its implementation has been proposed (Solihin and Eastman 2015). To recognize various approaches to the subject of automated rule checking, the authors introduce a map and place those approaches into the map (Figure 1).

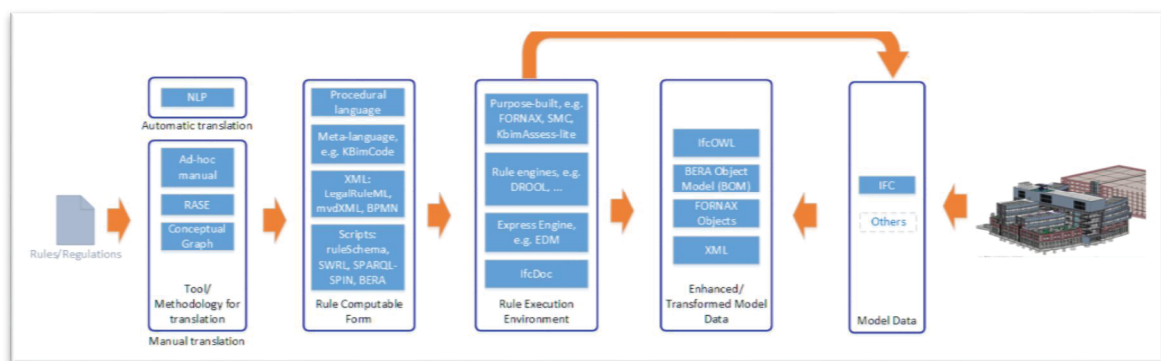


Figure 1: Components of automated rule checking systems

The map in Figure 1 illustrates several recent efforts according to the scope of what they focus on solving. Starting from the left, there are several efforts to translate rules into computable forms. All of them require largely manual translation, except research on Natural Language Processing (NLP) that attempts to fully automate the translation (Zhang and El-Gohary 2016). The manual approaches include ad-hoc free-form specifications with manual documentation such as those used in CORENET ePlanCheck (Solihin and Eastman 2016a), markups on legal texts using specific tags as adopted by tools such as RASE (Hjelseth and Nisbet 2011), or the use of a more formal form of documentation to capture

the rule requirements using the Conceptual Graph (Solihin and Eastman 2016b). In all approaches, expert knowledge is still required in one form or another especially in the manual approaches. Even in the automatic approach, expert knowledge is required to provide training samples, and the "gold standard" for validation (Zhang and El-Gohary 2016). The final product of the translations step is the form of computable representation that can take various forms of semantic models using XML such as LegalRuleML (Governatori et al. 2016), BPMN (Dimiyadi et al 2016), and mvdXML for IFC validation (Chipman 2013), meta-languages such as in KBim Code that can be "compiled" into a procedural language (Park et al 2016), standard query or rule language such as SWRL and SPARQL-SPIN that make use of semantic web (Zhang and Beetz 2016), or BERA (Lee 2011). All these methods are generally linked closely with the execution environment, which may be proprietary or custom developed for the method, or based on existing rule engines such as DROOL, Jena, or EXPRESS rule schema.

All the above tools eventually have to work with the BIM data. Virtually all of them currently work with IFC as the data source. Some work directly with the IFC file and build own internal data model such as FORNAX (Solihin et al 2004) and BERA Object Model (BOM) (Lee 2011), others use an intermediate toolkit to parse the IFC data and, in some cases, manage the resource usage; some may use some forms of database such as the open source BIMserver (Beetz et al 2010), or the commercial EDM ModelServer. One approach that is gaining popularity is the use of the semantic web (Pauwels et al 2016). To support the semantic web, an intermediate data representation for RDF is required, which is when IfcOWL fits in (Van Leeuwen et al 2009). IfcOWL converts IFC's EXPRESS schema and data into the RDF triplets that can be operated upon by SWRL, SPARQL, and SPIN.

Having the ability to do all of these, there is an impression that we now have the complete pieces to achieve the holy grail of the automated rule checking system. However, the role of data seems to be largely ignored and this is a missing piece that needs to be addressed.

### 3 THE MISSING PIECE OF THE RULE CHECKING SYSTEM - THE ACCESSIBLE DATA

Experience from developing the CORENET ePlanCheck automated code checking system in Singapore in early 2000 shows that it is equally important to have efficient access to the building data as it is having good, transparent, and maintainable computable rules. Efficient access to the building data is important since rule checking potentially has to iteratively go through almost the entire data set with a large number of rules. Even though current efforts of automated rule checking described in the previous section primarily use IFC data, it is actually not suitable due to its highly structured data model and its complexity. Relying on the raw IFC data will result in poor performance (Ghang et al 2014). Even a small delay in one rule may easily snowball into a significant delay in the real world usage where large numbers of rules are usually executed. BIM Assure (Invicara 2016), for example, is taking this approach seriously and has achieved highly efficient access to the data. From this perspective, IFC-based model servers and even IfcOWL suffer from the same issue. The SimpleBIM approach to simplify the IFC structure in generating the RDF graph is an effort to overcome this issue (Pauwels and Roxin 2016). But, the approach may not be adequate to solve the issue. Test results using SPARQL-SPIN on a Jena TDB triple store IfcOWL file show that relatively simple queries may take up to 28 seconds for a 14.7 MB IFC file. Given that a rule-checking process evaluates the entire

data of an IFC file iteratively, the time taken to evaluate many complex queries on a typical size of building model that ranges in a few hundred megabytes may take a considerable time.

Another critical issue that has been absent in the discussion so far is the support for geometry and spatial operations on the BIM data. Studies on building codes show that the majority of rules involve geometry and spatial operations. There are very few rules that can be done without the geometry besides the basic validation checks (Solihin and Eastman 2015). Even IFC MVD validation checks are not complete without the geometry validations (Solihin et al 2015). In most cases, the support of geometry is done using purpose-built geometry engine for rule checking as done in FORNAX (Solihin et al 2004) and DesignCheck (Ding et al 2006). Borrmann et al. propose a standardized query language extension to support geometry and spatial operations (Borrmann et al 2009). We are yet to see this approach implemented in a real world application.

Examples from the BuildingSMART Regulatory Room's project in identifying potential technology for the automated rule checking system for real world implementation have not addressed this issue completely. Many rules that require relationships or spatial operations of building data cannot be pre-computed and pre-assigned in advance because of a potentially infinite number of possible configurations. For example, a compliant means of egress cannot always be pre-determined because its minimum requirement may vary with other means of egress as well as the dynamic environmental changes within the building. These requirements can be shown in the following example (excerpt from IBC 2009):

1014.2 Egress through intervening spaces.  
Egress through intervening spaces shall comply with this section.  
Exceptions:  
2. Means of egress are not prohibited through stockrooms in Group M occupancies when all of the following are met:  
2.2 Not more than 50 percent of the exit access is through the stockroom;

The above rule allows means of egress to pass through stockrooms, but only for 50% of all means of egress. This means that certain paths passing through stockrooms are allowed only if there are already other means that do not go through stockrooms. To compute the selection of the sequence of which path should be considered first will be more complex than to identify all possible paths through or not through stockrooms first and to compare the counts afterward.

While in theory, it is possible to develop an optimized and accessible BIM data using various technology, it is ultimately a question of economy of scale. If each of the automated rule checking systems has to build such features complete with optimization techniques required, it will be a serious prohibition to realize a workable solution for automated rule checking system. It will bring us to the point as remote as ever to the practical solution.

## 4 BIMRL - A SIMPLIFIED SCHEMA TO ACCESS BIM DATA

Looking at the issue from the building data perspective, one can learn from the database world and the support for geospatial data, which is mainly 2-dimensional currently. Ideas to extend the 2D support to 3D support has been introduced (Zlatanova 2006). Oracle Spatial and PostgreSQL (with PostGIS extension) has an ability to store 3D polyhedron data and they have limited spatial functions on the 3D data. However, they may not be sufficient for use in automated rule checking for BIM due to the highly relational building



data. Our research on this issue has led us to look into a data warehouse approach for high-performance queries to the otherwise complex queries against the typical highly normalized production data in the relational database world. This approach allows us to create a version of the same data using a simplified star schema, which defines one central fact table and multiple dimensional tables that are only one step away, thus minimizing complex queries and allowing highly optimized queries to be computed (Solihin et al 2016a). We propose an equivalent simplified schema that is star-like for IFC data, called BIMRL simplified schema (Figure 2). The schema allows almost lossless IFC data transformation and utilizing the established relational database system, it allows highly efficient queries to be performed even on real-world sized models. The schema also supports geometry and builds spatial indexes that integrate the queries to the BIM data with the geometry and spatial operations. Now it is possible to define a single query to obtain data efficiently and with the spatial operations integrated. Below is an example of such query:

The above query searches for data outlet terminals and projection screens inside a

```
select unique c.elementid, c.elementtype, a.ifctype, a.name, a.tag, a.predefinedtype, c.name, c.modelid
from bimrl_type a, bimrl_element c, bimrl_element d, bimrl_spatialindex e, bimrl_spatialindex f
where a.elementid(+) = c.typeid and d.elementtype='IFCSPACE' and d.name='106'
and e.elementid=c.elementid and f.elementid=d.elementid and e.cellid=f.cellid
and (upper(c.name) like 'DATA OUTLET%' or upper(c.name) like 'PROJECTION SCREEN%');
```

space number 106. This query uses only the spatial index. In the full implementation of BIMRL for rule checking system, it supports two-step evaluation for the precise geometry comparison. This query takes 0.67 seconds on an 87MB IFC model (Solihin et al 2016b).

Dealing with geometry is challenging because of the complexity and multi-layered hierarchy of the 3D data. BIMRL persists the geometry data and its spatial indexes to allow optimization to be done at the database level, easing the burden from the application. It uses Octree decomposition commonly used in game and graphics application for the spatial approximation of the geometry shape. Each geometry will have a set of Octree cells dynamically decomposed up to the lowest level of subdivision set by the system. Each of these cells is encoded into a one-dimensional string so that the database can index the cells using an ordinary b-tree index commonly used in databases. Additionally, during the spatial indexing, a master tree tracks for overlapping cells found in geometries that may overlap such as a Space and objects contained in it. In order to minimize search time, the system resolves the overlapping cells into uniform cell decomposition. The details of the algorithm can be found in (Solihin et al 2016b). BIMRL also represents the geometry using a concept of multiple redundant representations. It keeps the geometry data in 1) a boundary representations that is derived from the triangulated mesh forming a set of planar boundary faces of the geometry, 2) triangulated polyhedra/mesh that represents the original geometry shapes, 3) an Octree approximation of the geometry that serves as the spatial indexes, 4) a X-Y plane aligned oriented bounding box, and 5) an axis-aligned bounding box (Figure 3). Each of different representations can be used in different operations depending on the exact needs. This approach allows very fast access to the geometry without the need to go through a geometry engine in most cases. (Solihin et al 2016) describes the concept in details.

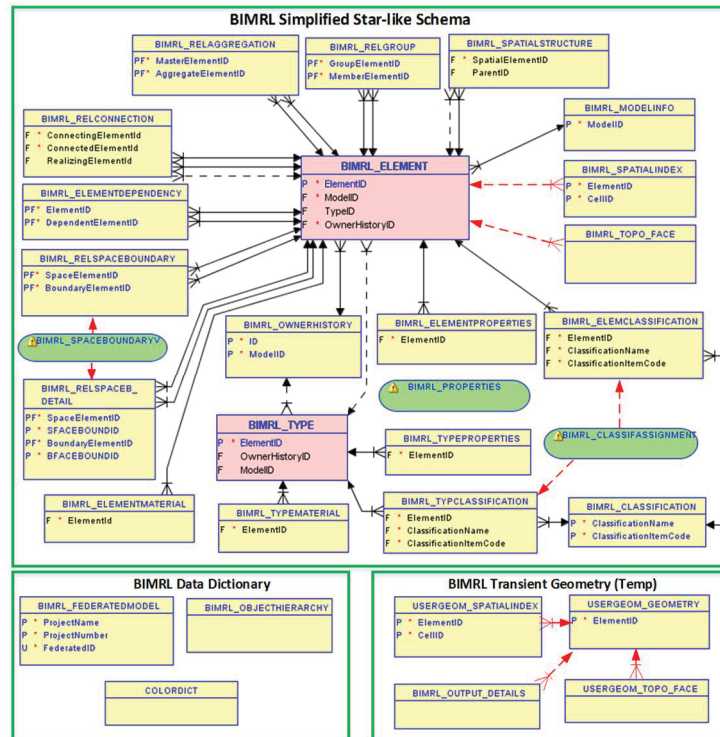


Figure 2: BIMRL simplified schema

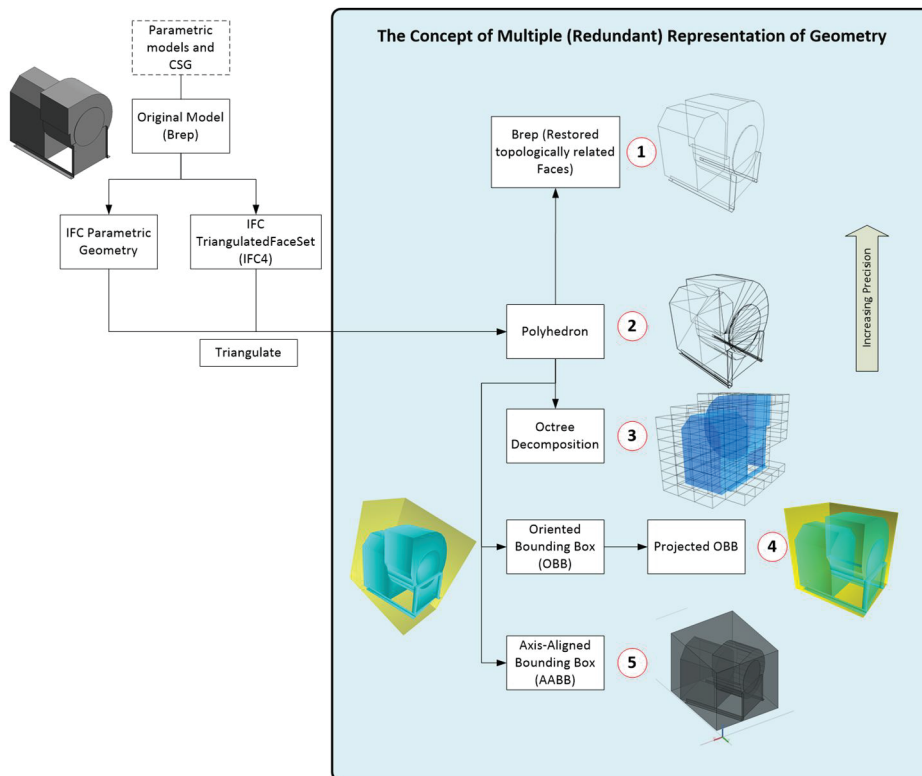


Figure 3: The concept of multiple representations of geometry in BIMRL

## 5 THE IMPLEMENTATION OF BIMRL

The BIMRL simplified schema has been implemented using Oracle RDBMS with Spatial extension to support the 3D geometries storage inside the database (Solihin 2015). The nearly identical schema has also been adapted to work with a NoSQL multi-mode and graph database (Solihin and Eastman 2016c). The effectiveness of the schema has been demonstrated with various real-world models (Solihin 2015).

BIMRL also adds two important built-in features, i.e. a space connectivity graph in the database, and the ability to construct temporary geometry for use in advanced spatial-based evaluations. The built-in graph makes it fast and easy to compute graph traversal without the need to re-compute it all the time. It even allows dynamic change to the edges such as temporarily increasing the cost to an undesirable path (Solihin 2015).

## 6 CONCLUSIONS

This paper discussed the critical role of accessible data for analysis of BIM data. The separation of data and its optimization is crucial for rule checking development so that a rule checking system and associated research studies can primarily deal with the rule logic and not with data accessibility and processing performance. In addition, this approach makes it possible for various optimizations to be done without affecting the rule logic.

A simplified model also plays an essential role to enable the efficient access to the data. Its simple adaptation to the graph database seems to offer a superior option to the RDF graph since the simplified schema allows constant time ( $O(1)$ ) to access the linked information instead of a search through the relevant collections of the RDF triplets. It is foreseen that a certain level of consolidation will occur where various specializations in the boxes shown in Figure 1 will work together to offer a practical and workable rule checking system for BIM in future.

## 7 REFERENCES

- Beetz, J., Van Berlo, L., De Laat, R. & Van Den HE, P. (2010). BIMserver. org—An open source IFC model server. Proceedings of the CIP W78 conference, 2010.
- Borrmann, A., Schaufstetter, S. & Rank, E. (2009). Implementing metric operators of a spatial query language for 3D building models: octree and B-Rep approaches. *Journal of Computing in Civil Engineering*, 23, 34-46.
- Chipman T, L. T., Wiese T. (2013). mvdXML. buildingSMART Model Support Group.
- Dimyadi, J., Pauwels, P. & Amor, R. 2016. Modelling and accessing regulatory knowledge for computer-assisted compliance audit. *ITcon journal*, 21, 317-336.
- Ding, L., Drogemuller, R., Rosenmant, M., Marchant, D. & Gero, J. (2006). Automating code checking for building designs-DesignCheck. Online at: <http://ro.uow.edu.au/engpapers/4842>
- Fenves, S. J. (1966). Tabular decision logic for structural design. Proceedings of the American Society of Civil Engineers, 92, 473-490.
- Ghang, L., Jiyong, J., Jongsun, W., Chiyon, C., Seok-Joon, Y., Sungil, H. & Hoonsig, K. (2014). Query Performance of the IFC Model Server Using an Object-Relational Database Approach and a Traditional Relational Database Approach. *Journal of Computing in Civil Engineering*, 28, 210-222.
- Governatori, G., Hashmi, M., Lam, H.-P., Villata, S. & Palmirani, M. (2016). Semantic Business Process Regulatory Compliance Checking Using LegalRuleML. Knowledge

- Engineering and Knowledge Management: 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, 2016. Springer, 746-761.
- Hjelseth, E. & Nisbet, N. (2011). Capturing normative constraints by use of the semantic mark-up (RASE) methodology. CIB W78 2011 28th International Conference- Applications of IT in the AEC Industry, 2011.
- ICC 2009. IBC 2009. <http://shop.iccsafe.org/codes/2009-international-codes/2009-international-building-code-tab-combo.html>.
- Invicara. 2016. BIM Assure [Online]. <http://bimassure.com>. [Accessed 23-Nov-2016 2016].
- ITcon. 2016. CIB W78 2015 Special track on Compliance Checking [Online].
- Lee, J. K. (2011). *Building environment rule and analysis (BERA) language and its application for evaluating building circulation and spatial program*. Published Ph.D. Dissertation, Georgia Institute of Technology.
- NovaCITYNETS. FORNAX [Online]. <http://www.novacitynets.com/fornax/index.htm>. [Accessed 23-Nov-2016 2016].
- Park, S., Lee, Y.-C. & Lee, J.-K. (2016). Definition of a domain-specific language for Koran building act sentences as an explicit computable form. *ITcon*, 21, 422-433.
- Pauwels, P. & Roxin, A. (2016). SimpleBIM: From full ifcOWL graphs to simplified building graphs. 11th ECPPM, 2016. CRC Press, 11-18.
- Pauwels, P. Zhang, S. & Lee, Y.-C. (2016). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, 73, 145-165.
- Solibri. Solibri Model Checker [Online]. <https://www.solibri.com/products/solibri-model-checker> [Accessed 03/01/2014 2014].
- Solihin, W. & Eastman, C. (2015). Classification of rules for automated BIM rule checking development. *Automation in Construction*, 53, 69-82.
- Solihin, W. & Eastman, C. (2016a). A knowledge representation approach in BIM rule requirement analysis using the conceptual graph. *ITCon*, 21, 370-401.
- Solihin, W. & Eastman, C. (2016b). A Simplified BIM Model Server on a Big Data Platform. CIB W78 2015. Brisbane, Australia.
- Solihin, W., Eastman, C. & Lee, Y.-C. (2015). Toward robust and quantifiable automated IFC quality validation. *Advanced Engineering Informatics*, 29, 739-756.
- Solihin, W., Eastman, C., Lee, Y. & Yang, D. (2016). A Simplified Relational Database Schema for Transformation of BIM Data into a Query-Efficient and Spatially Enabled Database. Submitted for publication. Draft available at: <https://tinyurl.com/hmj8mr>
- Solihin, W. & Eastman, C. M. (2016c). Multiple Representation Approach to Achieve High-Performance Spatial Queries of 3D BIM Data Using a Relational Database. Submitted for publication. Draft available at: <https://tinyurl.com/hyepct5>
- Solihin, W., Shaikh, N., Rong, X. & Poh, L. K. (2004). Beyond Interoperability of Building Models: A Case for Code Compliance Checking. BP-CAD Workshop. Carnegie Mellon University. Available online at <http://tinyurl.com/zaz8lbh>
- Van Leeuwen, J., De Vries, B. & Beetz, J. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, 23, 89-101.
- Zhang, C. & Beetz, J. (2016). Querying linked building data using SPARQL with functional extensions. Online at ResearchGate: <http://tinyurl.com/hhbvwed>.
- Zlatanova, S. (2006). *3D geometries in spatial DBMS. Innovations in 3D geo information systems*. Springer.