<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Comparative analysis of the applicability of BIM query languages for energy analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Ying, H; Lee, SH; Lu, Q</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>The 2016 CIB W78 Conference, Brisbane, Australia, 31 October-2 November 2016.</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2016</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/241033">http://hdl.handle.net/10722/241033</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Comparative analysis of the applicability of BIM query languages for energy analysis

Huaquan Ying, u3004315@connect.hku.hk
The University of Hong Kong, Hong Kong
Sanghoon Lee, sanghoon.lee@hku.hk
The University of Hong Kong, Hong Kong
Qiuchen Lu, u3003601@connect.hku.hk
The University of Hong Kong, Hong Kong

Abstract
A range of query languages have been used or developed to query partial information from Building Information Model (BIM)-based databases and files in recent decades. This paper aims to investigate the applicability of existing BIM query languages to extract necessary information from BIMs for energy analysis. A total of 16 query languages categorized into two groups, namely programming or generic query language, and domain specific query language, are summarized through extensive literature review. The key requirements of BIM data query for energy analysis are also developed, which include MVD based query support, custom query support, and easiness to construct queries. Taking these requirements as the criteria, the applicability of the 16 query languages is compared and analyzed. This paper then proposes some suggestions for developing effective and efficient building information query mechanisms for energy analysis.

Keywords: Building information model (BIM), query languages, partial information extraction, energy analysis

1 Introduction
BIM, which provides a digital representation of building physical and functional characteristics (National BIM Standard 2016), is an emerging technology to facilitate information exchange among stakeholders in the Architecture, Engineering and Construction (AEC) industry. Industry Foundation Classes (IFC) has been developed as an open and international standard for representing and exchanging building information. Currently, a complete BIM in a file format is generally used during the project life cycle (Zhang & Issa 2013). However, researchers have reported that large sizes of BIM or IFC files would make it difficult to exchange a model as a whole file (Adachi 2003; Won & Lee 2011; East et al 2013). In addition, individual stakeholders in different use-case scenarios usually have an interest in certain information stored in the model (Mazairac & Beetz 2013; Zhang & Issa 2013). These make it necessary to exchange partial information of a complete BIM (Mazairac & Beetz 2013).

To this end, a range of approaches have been developed. From the end user's perspective, they can be categorized into 2 groups: 1) query language based approaches, which extract partial model information by using existing programming languages (e.g. Java), generic query languages (e.g. Structure Query Language (SQL)), or developing domain specific languages (e.g. BIMQL); 2) filter based approaches, which rely on specific filters to fetch data from a BIM. The filters may be a standalone tool or a function component integrated in BIM authoring tools or downstream applications.

Building energy analysis is conducted widely to quantitatively evaluate and justify building design and operations decisions (O'Donnell et al 2013). Various efforts have been
undertaken to automate the process of BIM based building energy analysis (Hitchcock & Wong 2011). In particular, efforts have been made to adopt filter based approaches to extract and process necessary information from a BIM. For example, Lawrence Berkeley National Laboratory (LBNL) developed a tool called Space Boundary Tool (SBT) that can be used to extract building geometry information from an IFC model and then transform it to an EnergyPlus input file in the IDF format (O’Donnell et al 2013). Although a range of BIM query languages have been developed for partial BIM query, there are few efforts to explore applications of these languages in querying certain BIM information for energy analysis (Jiang et al 2015).

This paper mainly focuses on comparatively analyzing the applicability of BIM query languages for extracting the necessary information required for energy analysis. Extensive literature review on existing BIM query languages are conducted and summarized in the section 2. The details of the requirements of BIM data query for energy analysis are presented in the section 3. Based on these requirements, an applicability of the reviewed BIM query languages for energy analysis is analyzed and then some suggestions for building information query mechanisms for energy analysis are proposed.

2 Existing BIM query languages review
A range of query languages has been used or developed to extract partial information from a BIM in recent decades. In this section, an overview of 16 query languages is presented from the perspectives of their specific objectives and implementation platforms. These query languages can be categorized into two groups, namely programming or generic query languages and domain specific query languages. Table 1 shows the classification of these query languages.

- Programming or generic query languages
  Some existing programming languages (e.g. Java) or generic query languages (e.g. SQL, SPARQL, Language Integrated Query (LINQ)) have been used for partial BIM model query. An open source BIM server (i.e. the BIMserver by BIMserver.org) provides an interface for users to construct partial IFC model query using Java code (Beetz et al 2010). As the most well-established database query language, SQL also has been explored to use for the partial IFC model or other EXPRESS model (e.g. CIS/2 model) query (You et al 2004; Lee et al 2014), SPARQL is a Resource Description Framework (RDF) query language, which is able to retrieve and manipulate data stored in RDF format. Liu et al (2016) developed a SPARQL query enabled Revit plug-in for construction quantity take-off (QTO) information query. Daum & Borrmann (2013) leveraged the query capability of LINQ for spatial-temporal query on IFC models. XQuery was created as a standard language for querying XML (W3C, 2014). By extending XQuery with custom spatial query predicates, Nepal et al (2012) used it to query construction-specific spatial information from ontology-augmented BIMs.

- Domain specific query languages
  The domain specific query languages in this paper refer to the languages that are developed to facilitate information exchange in a specific domain by partial information query. Six domain specific query languages, including PMQL (Partial Model Query Language), ProMQL (Product Model Query Language), EDMespressX, BIMQL, 3D spatial query language and NL (natural language), are implemented on building model repositories (e.g. IMSvr, the BIMserver by BIMserver.org and EDM Model Server). The remaining five languages, including BERA (Building Environment Rule and Analysis) Language,QL4BIM, EQL (Express Query Language), MMQL (Multi-Model Query Language), and VBQL (Visual BIM Query Language), are implemented on a specific standalone application.

  From the domain aspect, these languages can be further divided into two categories: general domain query languages (i.e. PMQL, ProMQL, EDMespressX, BIMQL and EQL), which support general query function for IFC or generic EXPRESS models; high-level domain query languages (i.e. 3D spatial query language, BERA Language, MMQL, NL, QL4BIM and VBQL), which focus on specific information query from BIM models. For

Proc. of the 33rd CIB W78 Conference 2016, Oct. 31st – Nov. 2nd 2016, Brisbane, Australia
example, the 3D spatial query language (Borrmann & Rank 2009b, c; Borrmann et al 2009) and QL4BIM (Daum et al 2015) were developed for spatial analysis and partial IFC model extraction based on spatial constrain, MMQL (Fuchs et al 2014) aimed at construction-specific information query, and BERA Language was developed for the purposes of building circulation and spatial analysis (Lee et al 2015).

### Table 1 Classification of existing BIM query languages

<table>
<thead>
<tr>
<th>Query language</th>
<th>Objective</th>
<th>Implementation platform</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Programming or generic query language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINQ</td>
<td>Spatial-temporal query on IFC models</td>
<td>Standalone application</td>
<td>Daum &amp; Borrmann (2013)</td>
</tr>
<tr>
<td>SQL</td>
<td>Partial CIS/2 model query</td>
<td>GTCIS2SQL (Relational database)</td>
<td>You et al (2004)</td>
</tr>
<tr>
<td></td>
<td>Partial IFC model query</td>
<td>OR-IFC server (CUBRID based IFC server)</td>
<td>Lee et al (2014)</td>
</tr>
<tr>
<td>BIMQL</td>
<td>Partial IFC model query</td>
<td>BIMserver.org (Oracle Berkeley DB based IFC server)</td>
<td>Mazairac &amp; Beetz (2013)</td>
</tr>
<tr>
<td><strong>Domain specific query language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BERA Language</td>
<td>Building circulation and spatial analysis on IFC models</td>
<td>Solibri Model Checker Plug-in</td>
<td>Lee et al (2015)</td>
</tr>
<tr>
<td>MMQL</td>
<td>Construction-specific information query on multi-models including IFC models</td>
<td>Standalone application</td>
<td>Fuchs &amp; Scherer (2014)</td>
</tr>
<tr>
<td>NL</td>
<td>Construction-specific information query on Cloud BIM</td>
<td>Standalone application</td>
<td>Lin et al (2016)</td>
</tr>
<tr>
<td>QL4BIM</td>
<td>Spatial information query on IFC and CityGML models in an integrated context</td>
<td>Standalone application</td>
<td>Daum et al (2015)</td>
</tr>
<tr>
<td>VBQL</td>
<td>Construction-specific information query on IFC models</td>
<td>Standalone application</td>
<td>Wölfing et al (2014)</td>
</tr>
</tbody>
</table>

Generally, domain specific query languages require users to spend extra efforts learning how to use the languages. To lower the threshold for normal users, VBQL (Wülfing & Scherer 2014) are implemented with a well-defined GUI in which a range of query requirements are predefined and query patterns are fixed. Users only need select a query from provided options based on the intent.

3 Requirements of BIM data query for energy analysis
This section develops a set of key requirements for BIM data query languages when it is used to extract necessary information required for building energy analysis from a BIM. These requirements are presented from the view of information exchange requirements in the process of BIM based building energy modelling as well as the threshold for users to construct queries.

The inputs for energy simulations include building geometry, construction and material thermal properties, HVAC systems, internal loads, weather conditions, operating strategies and specific simulation parameters (Maile et al. 2007). The information of the first four items can be defined by IFC specifications, and thus in theory these pieces of information could be extracted from an IFC model (Hitchcock & Wong 2011; GSA 2015). Although in practice leveraging BIM data for energy analysis depends on the specific workflow of BIM based energy analysis as well as tools used, ideally the BIM query languages should be able to support the query of all the relevant information (i.e. building geometry, construction and material thermal properties, HVAC system and internal loads) that can be obtained from a BIM.

3.1 MVD based query support
A Model-View-Definition (MVD), which defines a subset of the IFC schema that is required by specific information exchange requirements (Davis et al 2012), can be used to clearly define the exact information set required for energy analysis. To date, several open MVDs related to building performance analysis, including energy analysis, have been or are being developed, some of which can be accessed in the Internet such as buildingSMARTalliance (https://www.nibs.org/?page=bsa_infoexchange) and IFC Solutions Factory (http://www.blis-project.org/IAI-MVD/).

As mentioned earlier, the information that can be extracted from a BIM for energy analysis is complicated, aggregated and various. It would be difficult and time-consuming for typical energy modelers to query all these large and aggregated information by writing lengthy and sequential query statements. It would be also a challenge for them to understand the information that need to be queried on a data model level (e.g. IFC model). It is effective if a query language would be able to take a MVD for energy analysis as an input to extract necessary data from a BIM. It would be also efficient as a set of query statements can be embedded and reused.

3.2 Custom query support
The data query language used for energy analysis shall be flexible and, more specifically, shall support custom queries. The complementary function of custom query allows users to construct their own queries tailored to meet specific intentions, in addition to the MVD based query. In practice, the scope and the level of detail (LoD) of energy modeling vary through different stages, which result in the difference of the LoD of information that needs to be extracted from a BIM for energy analysis (GSA 2015). For example, in the preliminary concept design phase, the goal of energy modelling may be to evaluate the impacts of site location and building orientation. Then querying building geometry information from a BIM is enough (use assumptions for other inputs). Actually, at this stage the IFC model can only provide building geometry information. However, in the Design Development phase, almost all the design decision would be made and the majority of the information in the four categories (i.e. building geometry, construction and material thermal properties, HVAC systems, internal loads) is needed and can be queried from a BIM (GSA 2015).
3.3 Easiness to construct queries

High-level threshold of a query language would be a huge obstacle to its wide adoption (Wülfing et al 2014; Mazairac & Beetz 2013). The threshold level of a query language can be evaluated based on the knowledge required for users to construct queries using that language. According to the survey of the languages listed in this paper, the specific knowledge can be divided into three categories: knowledge related to query language (language’s syntax and grammar), knowledge related to underlying data model (e.g. IFC), and knowledge related to specific domains (e.g. database, mapping rules from IFC model to a database). We believe that these categories of knowledge should be hidden from end users such as energy modelers or analysts as much as possible. Ideally, it would be effective and efficient if a query language only requires users to learn simple syntax and grammar before constructing a query.

4 Applicability analysis of existing query languages

In this section, an applicability analysis of existing query languages for energy analysis is presented. A total of 16 query languages to be analyzed are summarized through extensive literature review, as shown in Table 1. The criteria for applicability analysis are developed with the identified requirements. Based on the analysis results, several suggestions for developing effective and efficient BIM data query mechanisms for energy analysis are proposed.

4.1 Criteria for applicability analysis

The criteria used for the applicability analysis include the following three points:

- MVD based query support: Check to see if each query language can query the fundamental building information of building geometry, construction and material thermal properties, mechanical systems, internal loads and others that are defined in a MVD.
- Custom query support: Check to see if each query language can support to develop custom queries so that the user can query a set of information that are needed for the specific analysis intent.
- Knowledge requirements for users: Analyze whether the query language requires users to obtain knowledge related to query language, knowledge related to underlying data model, and knowledge related to specific domains.

4.2 Results of applicability analysis

16 query languages are analyzed based on the established criteria. Table 2 shows the applicability analysis results.

**MVD based query support**

6 query languages (i.e. Java, SQL, BIMQL, EDMexpressX, PMQL, and ProMQL) can be used to query general elements (e.g. IFC entities and their attributes or properties) defined in IFC data models. In addition, EQL is also applicable to IFC model query theoretically (Mazairac & Beetz 2013), although it is designed as a query language for generic EXPRESS models (Koonce et al 1998). The four types of information (i.e. building geometry, construction and material thermal properties, HVAC systems, and internal loads) required for energy analysis can be explicitly defined in the IFC schema and can be stored in specific IFC entities and their attributes. For example, the information of building geometry is stored in IFC physical entities such as IfcWallStandardCase, IfcSlab and IfcWindow, etc.; the information of construction and materials can be found in IfcMaterial, IfcMaterialLayer, and IfcMaterialLayerSet, etc.; HVAC systems can be defined by the IFC entities in IfcHvacDomain and internal loads by IfcPropertySet. Therefore, all the 7 languages that support the generic query of IFC model on IFC entities and attributes level can be used to query all the four types of information.
Table 2 Comparison between existing BIM query languages based on the established criteria

<table>
<thead>
<tr>
<th>Query language</th>
<th>MVD based query support</th>
<th>Custom query support</th>
<th>Knowledge requirements for users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Query language</td>
</tr>
<tr>
<td>Programming or generic query language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Java</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>LINQ</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>SPARQL</td>
<td>√</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>SQL</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>XQuery</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIMQL</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDMexpressX</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>PMQL</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProMQL</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>3D spatial query language</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>EQL</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>BERA Language</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMQL</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>NL</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QL4BIM</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>VBQL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The remaining 9 query languages, including 3 generic query languages (i.e. LINQ, SPARQL, and XQuery) and 6 domain specific query languages (i.e. 3D spatial query language, BERA Language, MMQL, NL, QL4BIM, and VBQL) cannot support the query of all the four types of information required for energy analysis. The reasons are summarized as the following three aspects:

1) The following languages, BERA Language, LINQ, 3D spatial query language, MMQL, QL4BIM, XQuery, and VBQL, are developed for specific information query and their query scopes are predefined with specific languages’ grammar. For example, BERA Language, which is developed specifically for building circulation and spatial analysis in early design stage, focuses on the query and analysis of space objects and spatial relationships (Lee et al 2015). The reference of attributes and properties through a [Entity].[attribute/property] pattern (e.g. “Space.GUID”) are hard-coded into the grammar of BERA Language (Mazairac & Beetz 2013).

2) SPARQL was used by Liu et al (2016) to query construction QTO information from a construction ontology-augmented BIM where only the information of building geometry and construction and materials can be queried.

3) NL based Cloud BIM query approach proposed by Lin et al (2016) relies on specific and predefined mapping of natural languages to IFC entities or properties. Currently, a specific underlying mapping mechanism for energy analysis related information query has not been developed.

The function of query languages that allow users to predefine, store and reuse query statements of specific view can be useful when the languages are used to query complicated...
and aggregated information like MVDs. This functionality can save user’s time by leveraging the predefined query statements. In addition to general-purpose programming languages, PMQL has the functionality that allows users to predefine ‘view definition’ data that is required for the implementation with the specific view (Adachi et al 2003). The predefined PMQL statements can be stored in IFC Model Server and reused in a new PMQL query. BIMQL also has the similar function. On the top of BIMserver.org, BIMQL query statements can be predefine, stored and reused. However, the predefinition of MVD concepts using PMQL or BIMQL requires users to have an in-depth understanding of the MVD concepts and then write corresponding query code manually.

An effort to automate the generation of MVD based query code has been made by Jiang et al (2015). They developed a tool called QueryGenerator which can be used to semi-automatically generate Java code corresponding to a MVD query. The Java code can be implemented on BIMserver.org to produce the query results. However, currently they only showed the tool’s application on a simple building model for airflow analysis.

In summary, all the 16 languages reviewed in this paper currently do not have dedicated constructs which can make the languages use a MVD or MVD concept as an input directly to query information from a BIM.

**Custom query support**

15 query languages support custom query while one language, namely VBQL, do not. The 15 query languages give users enough flexibility to construct custom queries for a wide range of purposes although some of them may focus on specific information query as discussed earlier. Although VBQL provide several options for users to select for different query requirements, the scope of the options is limited and users can only query the information that is predefined.

**Knowledge requirement for users**

9 out of 16 languages (i.e. Java, LINQ, BIMQL, EDMexpressX, PMQL, ProMQL, EQL, MMQL and QL4BIM) require users to obtain not only the specific knowledge on the query languages themselves but also on the underlying data models (e.g. IFC data model). It is worth to note that, among the 9 query languages, Java is a programming language while other query languages are SQL-like. This means that Java requires a higher level of threshold for users, compared with others. Other than the knowledge on languages and underlying data models, SQL also needs users to understand specific domain knowledge such as database and the mapping rules of IFC model to database (You et al 2004; Lee et al 2014). 3D spatial query language is implemented based on ISO standard SQL: 1999 by providing spatial types and operators in SQL (Borrmann & Rank 2009b, c). Similar to SQL, it also requires users to obtain the knowledge related to the relevant mapping rules. Although the NL-based query is constructed with natural language, the terms in the query sentence should be the “effective” keywords that have been predefined. Users need to be familiar with these mapping rules in advance. By mapping the BIM data model to a new semantic model, the complicated underlying data model is hidden from users in SPARQL (Liu et al 2016), XQuery (Nepal et al 2012) and BERA Language. Therefore, users only need to learn the languages themselves. To make XQuery based quires easier for practitioners, Nepal et al (2013) further developed a series of form-based query specification templates which assist users to specify queries without knowing the language of XQuery. VBQL does not have such knowledge requirements for users to construct queries, while users are required to know some implementation instructions.

**Summary**

The above analysis demonstrates that the query languages listed in this paper do not provide an efficient and effective query mechanism to fulfill the three requirements. There are two major limitations in using these query languages for energy analysis:

1) 7 query languages (i.e. Java, SQL, BIMQL, EDMexpressX, EQL, PMQL, and ProMQL) can support the query of all the four types of information (i.e. building geometry, construction
and material thermal properties, HVAC systems, and internal loads) required for energy analysis. They also allow users to construct custom queries. However, the 7 query languages currently have not been implemented for energy analysis-related MVDs. Users need to query the large and aggregated set of building information required for energy analysis by writing complicated query statements manually. Although the development is challenging as well as time-consuming, it is still implementable. The COBie import/export plugin for the open source BIMserver.org platform is a good example (https://github.com/opensourceBIM/COBie-plugins).

2) All the 7 query languages not only require users to learn the languages’ syntax and grammar but also need users to have a basic understanding of the underlying data model. For example, users who adopt BIMQL to construct queries need have an in-depth knowledge of the IFC data model (e.g. entities expression and hierarchy). It would largely reduce the burden of general users who have not expertise on the IFC data model if these are hidden.

4.3 Suggestions of the BIM data query mechanism for energy analysis

Based on the applicability analysis of the 16 query languages, we propose two suggestions to develop effective and efficient BIM query mechanisms for energy analysis.

**Extend BIMQL with the function of MVD based query**

The 7 query languages (i.e. Java, SQL, BIMQL, EDMexpressX, EQL, PMQL, and ProMQL) in this paper potentially support for general and/or custom queries of the information required for energy analysis. One major limitation in using the query languages is that they have not been successfully implemented with energy analysis-related MVDs. Developing specific operators for query languages like BIMQL that can parse MVD concepts can be a promising solution. With this function, users can reference MVD or a MVD concept as a normal variable instead of writing corresponding query statements. BIMQL is preferable to other languages listed in this paper as it has demonstrated serval advantages. BIMQL supports to query all the four types of information required for energy analysis, and can be implemented on an open source BIM server. Although other domain specific languages such as PMQL and EDMexpressX also have the potential to implement energy analysis related MVDs query, all of them are not open source and could be costly for custom development in a specific platform.

**Automate the generation of MVD based query code for energy analysis**

The idea inspired by Jiang et al (2015) is developing a query code generator that can take MVDs or MVD concepts as an input and automatically output the corresponding query code in BIMQL. This would be greatly beneficial to reduce users’ burden as both the query code writing process and the IFC data model are hidden from end users. The application can be developed as a plug-in and integrated into the environment of BIMserver.org.

5 Conclusion

In this paper, we first conducted extensive literature review of a total of 16 BIM query languages categorized into two groups, namely programming or generic query language and domain specific query language. The key requirements of BIM data query for energy analysis were developed, including: MVD based query support, custom query support, and easiness to construct queries. By taking the requirements as the criteria, we then made a comparative analysis of all the 16 query languages. The analysis results showed that the query languages listed in this paper have not been successfully used to extract the whole set of the information required for energy analysis efficiently and effectively as the query languages lack the functions needed, especially for MVD based query support. Furthermore, the languages still have a high-level threshold for normal users. Therefore, we recommended two suggestions to address the limitations: extend BIMQL with the function of MVD based query and automate the generation of MVD based query code for energy analysis.
References


