

Building performance optimisation: A hybrid architecture for the integration of contextual information and time-series data

Shushan Hu ^{a,b}, Edward Corry ^{a,b}, Edward Curry ^c, William J.N. Turner ^b, James O'Donnell ^{a,*}

^a School of Mechanical and Materials Engineering, University College Dublin, Belfield, Dublin 4, Ireland

^b UCD Energy Institute, University College Dublin, Belfield, Dublin 4, Ireland

^c School of Computer Science and INSIGHT Research Centre, University College Dublin, Belfield, Dublin 4, Ireland

ARTICLE INFO

Article history:

Received 5 June 2015

Received in revised form 6 April 2016

Accepted 22 May 2016

Available online 18 July 2016

Keywords:

Building performance optimisation

Data interoperability

Linked data

BMS

Time-series data

Relational database

ABSTRACT

Buildings tend to not operate as intended, and a pronounced gap often exists between measured and predicted environmental and energy performance. Although the causes of this 'performance gap' are multi-faceted, issues surrounding data integration are key contributory factors. The distributed nature of the Architecture, Engineering and Construction (AEC) industry presents many challenges to the effective capture, integration and assessment of building performance data. Not all building data can be described semantically, nor is it feasible to create adapters between many different software tools. Similarly, not all building contextual data can easily be captured in a single product-centric model.

This paper presents a new solution to the problem based upon a hybrid architecture that links data which is retained in its original format. The architecture links existing and efficient relational databases storing time-series data and semantically-described building contextual data. The main contribution of this work is an original RDF syntax structure and ontology to represent existing database schema information, and a new mechanism that automatically prepares data streams for processing by rule-based performance definitions. Two test cases evaluate the concept by 1) applying the hybrid architecture to building performance data from an actual building, and 2) evaluating the efficiency of the architecture against a purely RDF-based solution that also stores all of the time-series data in RDF for a virtual building. The hybrid architecture also avoids the duplication of time-series data and overcomes some of the differences found in database schemas and database platforms.

© 2016 Published by Elsevier B.V.

1. Introduction

Buildings are responsible for 40% of EU energy consumption and 36% of the EU's CO₂ emissions. Improved environmental and energy performance of buildings is therefore a key EU objective [1]. A recognised and multi-faceted performance gap exists between measured and predicted building performance. De Wilde attributed components of the gap to three broad stages of the building lifecycle: design, construction (including handover) and operation [2].

One key contributing factor to the performance gap is the accuracy of physical measurement of building performance information. One cannot manage what one does not measure, and existing measurement practises can be inadequate or difficult to implement effectively [3,4]. The poor transfer of performance information [2] throughout the building lifecycle [5] and poor interoperability and data integration between toolsets in the domain can lead to inadequate levels of performance assessment [6].

Schein [7] and Granderson [8] have provided robust and effective methods of building performance assessment, while approaches to data integration such as Building Information Modelling (BIM) and the Internet of Things (IOT) broadly promise a resolution to the data integration problem. The limiting factor restricting optimal performance assessment remains the use of integrated performance data.

The AEC industry presents some unique challenges to data integration due to the distributed nature of the industry and the presence of many interacting but poorly linked domains. A direct consequence of this is that building managers do not have access to the data and information they need in order to optimally manage buildings [9].

1.1. Computer science approaches to data integration

When considered more closely, there are three main computer science approaches which might be applied to the data integration issue; the common data model, the adapter approach, and semantically-described data. Each approach has significant drawbacks when applied in the AEC industry.

Proprietary software suites use an underlying data model to describe data exchange in specific domains for instance [10,11]. The common

* Corresponding author.

E-mail address: james.odonnell@ucd.ie (J. O'Donnell).

data model approach uses a central underlying data model as a data hub through agreement between tool vendors. The data hub is then interpreted by domain-specific applications. In response, the buildingSMART alliance provides the open Industry Foundation Class (IFC) data model to describe data exchange in specified domains in an open manner [12], though significant challenges remain relating to the description of diverse building data domains, and the broader acceptance and adoption of the standard in the wider BIM community.

Rather than attempting to convert all data to a particular data model, the adapter approach uses adapters to integrate different data models, representing different AEC domains [13]. Previous efforts such as the HESMOS and Cooperate projects [14–16] used such an approach, enabling data transfer between different native data formats. However, significant data loss is inherent in such an approach [17]. Also, in an industry as diverse as AEC, it is not feasible to develop and maintain adapters for all cases of data transfer between model and toolsets in the industry [18].

The semantic web approach uses ontologies to provide a way to add knowledge to unstructured data. These ontologies are typically based on Web Ontology Language (OWL) [19], which uses the Resource Description Framework (RDF) to describe the relationships between specific objects using the subject-object-predicate structure [20]. This approach has proven to be quite successful in integrating cross-domain data sources in the AEC industry [11,21–23], by allowing the description of building context data in a homogeneous fashion [11,22,24]. For example, the Sensor Ontology [25] was developed to describe sensor information including placement and measurement contexts.

The semantic web approach can be used to overcome some of the issues associated with product-centric data models such as IFC, and ontology matching approaches [26] can be used to provide greater meaning across domain-specific ontologies. The semantic web approach can be used very effectively to link contextual building data, including building geometry, material properties, as-built construction details and HVAC specifications. Time-series data can be managed very efficiently in existing database structures.

This paper presents a flexible, *as-needed* solution to information sharing in the AEC domain, whereby data is retained in the most *appropriate* format and shared as required. This hybrid approach may provide an inexpensive and effective solution to the data integration issue.

The well-recognised performance gap between measured and simulated data is a multi-faceted problem with root causes spread throughout the building lifecycle [2]. Different aspects of the performance gap have been considered in isolation, including issues around assumption, approximation, and simplification of both measurement and modelling [27]. Additionally the performance gap associated with user behaviour and occupancy patterns within buildings is difficult to qualify for a number of complex and related reasons [28]. Notably, the manual linkage of person and room information has a number of practical and accuracy constraints that permeate the building life cycle [29]. This paper is concerned specifically with enhancing the interoperability between measured and simulated performance data encountered during building operation [2].

The hybrid approach enables enhanced building performance assessment through a hybrid data architecture solution, integrating existing and available data sources in buildings using both semantic web technologies and more traditional database solutions. A new mechanism is defined to automatically prepare data streams for processing by rule-based performance definitions.

Building on previous work by Corry et al. [8] and O'Donnell et al. [9,11,21], the objective of this paper is to show how building context data can be mapped to related performance data. The proposed approach keeps time-series data within its original database, avoiding duplication while benefitting from the high efficiency of mature database platforms, especially for structured fixed data [30,31]. This approach defines an RDF syntax structure and vocabulary to represent

database schemas based on semantic web technology. It also provides a framework for the access of time-series data in databases based on building geometry contextual data and SSN contextual data.

Section 2 describes the technologies used in this work. Section 3 introduces the hybrid architecture used to integrate AEC performance data with building context data. Section 4 details the syntax of the mapping description, the connection module and the mapping module. Section 5 describes two test cases for evaluation and verification of the new approach. Finally, Section 6, the conclusion, presents the main outcomes from this work.

2. Case for a hybrid solution

Augenbroe advocates a rigorous use of building performance indicators to ensure compliance between project specification and performance [32]. This type of rigour is hampered by the nature of the data capture and storage systems used currently in the AEC industry.

Due to the fragmented nature of the AEC industry, many domain-specific data models exist. The rigorous use of performance indicators requires the integration of data retained in these formats, in some manner.

The performance framework using the scenario modelling method [9] followed on from previous performance metric/indicator work by Hitchcock and Augenbroe [33–35]. Additionally, BuildingEQ developed practical measurement and metric sets for buildings [36]. Corry et al. addressed some of the limiting factors of this framework through the formalisation of a performance assessment ontology for buildings [37]. The value of the performance indicator approach is enhanced through recent research developments [11,21,22,38]. Each of these efforts identified how the successful implementation of a performance indicator approach is dependent on access to reliable, integrated, building information.

Choosing one method for the integration of such diverse data is neither feasible, nor appropriate.

2.1. Data sharing for performance evaluation

Cross-domain data sharing is not commonplace in the industry. For example, linking occupancy patterns to building operation, as illustrated in Fig. 1, would enable traditionally separate information sources to be combined and potentially offer new insights into building operation. In this example, tracking the number of members in a given part of a sports centre can aid with staff scheduling and more focused HVAC operational strategies. Similarly, linking design intent to building operation or linking simulation to building operation requires an integrated approach to data management that is not always evident.

This paper strongly advocates leaving building data in the most appropriate platform and format, only using and linking such data on an *as-needed* basis. The paper presents a novel architecture to allow this to take place.

Contextual information about a building can add to the effectiveness of performance data. Context information may include building geometry and HVAC descriptions as obtained from Building Information Model (BIM) formats, including IFC [39,40] and SimModel [41,42]. Sensor definitions described using the Semantic Sensor Network (SSN) ontology [43] and other soft building information [11] based on the RDF format contribute to a richer set of context information. This type of information can be described and integrated very effectively using the semantic web approach.

Time-series data mainly describes continuous records from deployed sensors and metres in buildings [44]. There are some relationships between context information and time-series data. For example, a sensor ID described using the SSN ontology may also be referred to in time-series data generated from the sensor (see Fig. 2).

Traditionally, semantically-described contextual data might be linked to time-series data through an Application Programming Interface (API),

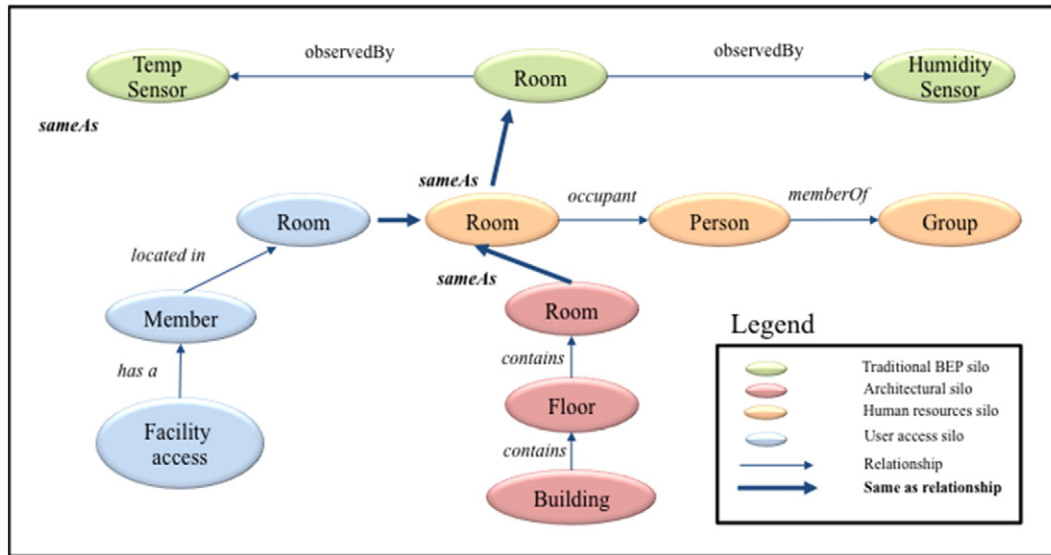


Fig. 1. Semantic web based integration of traditionally separate silos of information: BMS data from the BEP silo; user access records; BIM for the Architectural silo and human resources systems [11].

using Structured Query Language (SQL) queries. However, API coding incurs an additional effort overhead. These APIs directly translate SPARQL queries (the language used to query RDF files) into SQL queries [45]. This approach is imperfect as some data is retained in RDF format [21] for the purpose of linking [46] and inferencing [47]. In addition, the format of time-series data in databases is well established and it is not efficient to leverage a complex algorithm for the translation of SPARQL into SQL to query such data.

Another approach is to extract time-series data from a relational database and transform it into RDF format. Much work has been done to represent database schemas [48] and transform data in databases into RDF data [49–51]. Although such data can be linked to other existing RDF data silos, several issues arise:

1. Time-series data from the original BMS database is duplicated in an RDF store, leading to inefficiencies;
2. RDF is not effective at representing fixed-structure data and more space is consumed for its storage [31];

3. Triple stores (frameworks used for storing and querying RDF data) for RDF data imply less efficient lookups compared with relational databases [52].

The key contribution of this work is the integration of time-series data with contextual building data in a manner that does not require significant reinterpretation and conversion. This enables a faster run-time performance of the hybrid system. Based on this work, building performance can be more comprehensively assessed with integrated data sources, using building performance assessment methods [53].

3. System architecture overview

The hybrid approach proposed in this paper aims to overcome the current disconnect between context information and time-series data to enable in-depth building performance analysis. This hybrid approach illustrates how to effectively integrate time-series data stored in relational databases with other building contextual information in order to enable rule-based building performance evaluations. Design of this software system aligns with an approach taken by a relevant research project, namely the Linked dataspace for Energy Intelligence (LEI) [38]. In doing so, the hybrid architecture uses four distinct layers that clearly distinguish between: 1) applications for energy management, 2) support services for accessing linked data, 3) linked cross-domain context data and 4) data sources (see Fig. 3). When combined, these layers deliver a robust, flexible, loosely-coupled architecture for building energy management. Each layer is described in detail between Sections 3.1 and 3.4.

3.1. Application Layer

The Application Layer provides software interfaces to end-users and is easily extensible with additional tools. Applications come in many forms with the intention of enabling effective decision support. For example, dashboards graphically present data in a simple and accessible format, while building energy analysis tools are used to gain a deeper understanding of the energy consumption patterns within a building.

3.2. Support Services Layer

The Support Services Layer enables access to linked data formats by providing an access infrastructure to the linked data (Fig. 4).

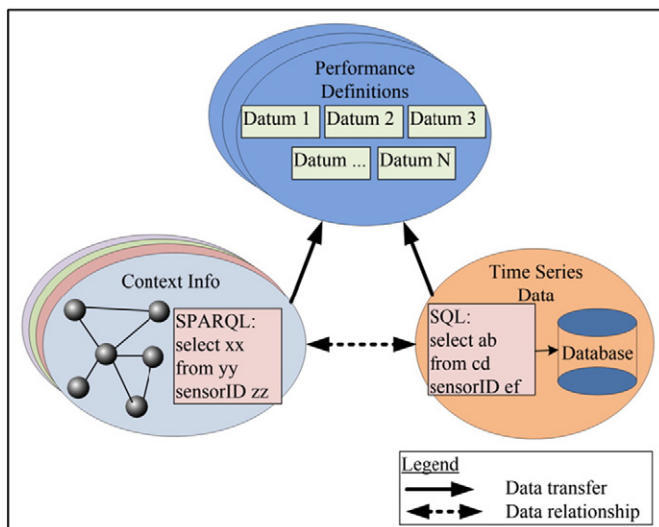


Fig. 2. Integrating diverse building data silos in an appropriate manner can enhance building performance assessment.

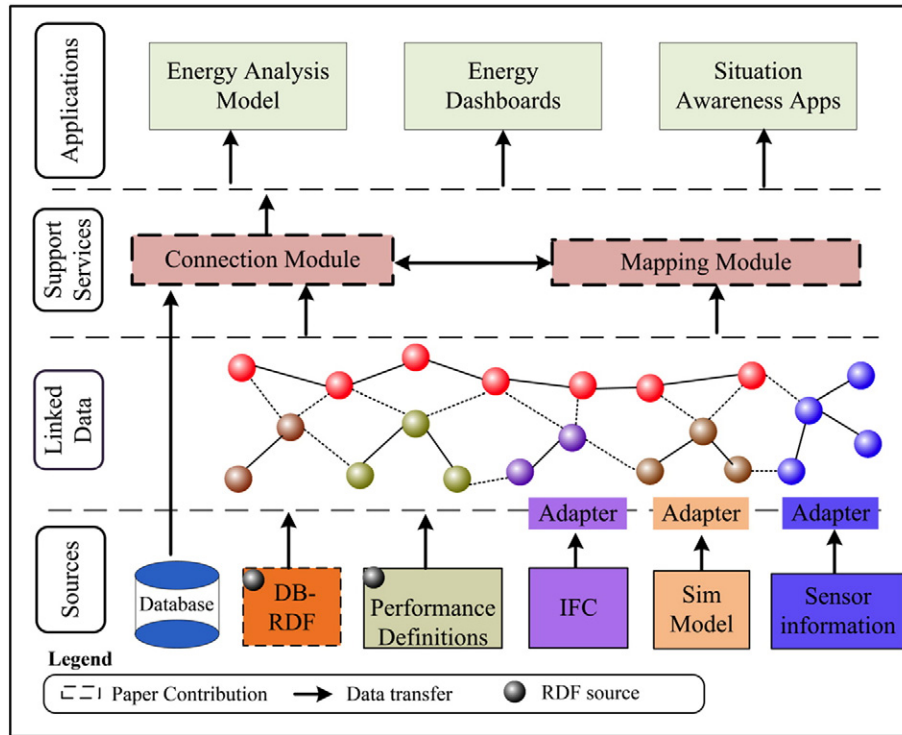


Fig. 3. Hybrid architecture used to integrate contextual and time-series building performance data for access by specific applications.

In the context of this paper, the ‘connection module’ and ‘mapping module’ are key new developments. The connection module links RDF contextual building data to relational database platforms such as Mysql, DB2, and Oracle (Section 4.2). The mapping module on the other hand obtains time-series data from linked BMS data-bases using schema information stored in the *DB-RDF* file and other contextual building data (Section 4.3). The mapping module furthermore identifies data needed by rule-based performance definitions such as specific performance metrics and associated measurements. The novelty of the mapping module lies in the manner in which these two types of query are integrated:

1) SPARQL queries on RDF contextual building data, and 2) SQL queries on time-series data stored in relational databases. SQL queries are sent to the connection module that accordingly returns time-series data to the mapping module.

3.3. Linked Data Layer

The Linked Data Layer contains the RDF format contextual data sources after they have been linked using relationships between entities. Two types of relationships exist:

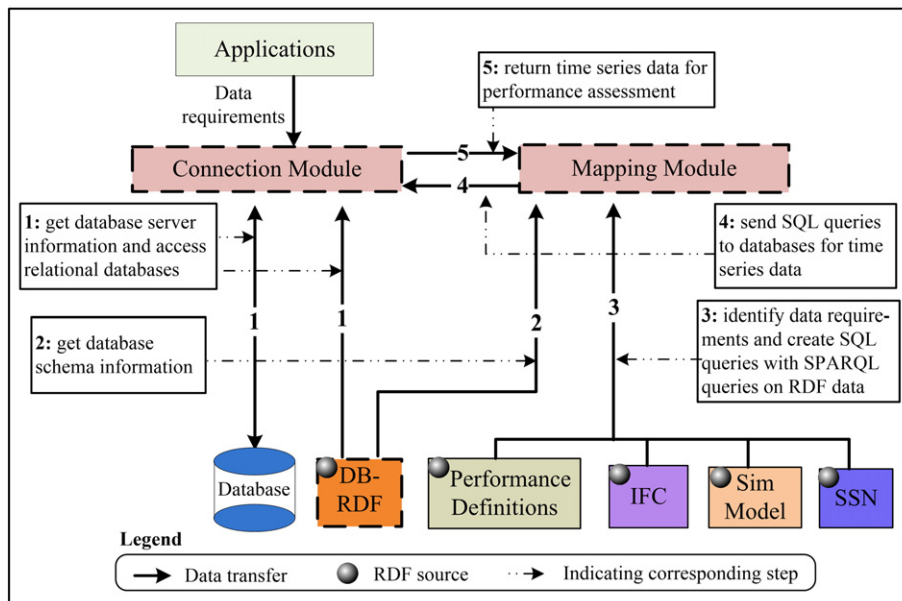


Fig. 4. Workflow of the architecture enabling integration of time-series data in relational databases.

1. General relationships between different entities within the same data source;
2. 'Same As' relationships between entities in different data sources (Fig. 1).

For example, the entity named 'Room' in the BEP silo is the 'Same As' the entity named 'Room' in the IFC silo. These relationships are described through defining RDF syntax structure and vocabulary.

Another key contribution of this work is a new RDF syntax structure and vocabulary, called *DB-RDF*, which describes database schemas (e.g. database name, table name etc.) and server information (e.g. database platform, IP information etc.) in a common format. The server information is used to automatically connect various database platforms. The database schema information is leveraged to create SQL queries, based on other contextual building data, to obtain time-series data from databases (Fig. 4). *DB-RDF* is described in detail in Section 4.1.

3.4. Sources Layer

The Sources Layer contains data and information in native formats, i.e. time-series building performance data in a database and contextual building data stored in BIM, RDF and other formats. Applications that work in their native formats offer greater flexibility in terms of available functionality and remove the need for extensive data transformation between models. If necessary, the contextual building data, originally stored in IFC, SimModel, and SSN is converted to RDF format using a set of existing adapters [40,54,55].

4. Detailed system specifications

The new hybrid approach uses existing technologies, such as the RDF standard and SQL queries, and requires three additional elements to integrate previously disconnected contextual data and time-series data contained within relational databases:

1. *DB-RDF* — an RDF syntax structure that describes database schemas and database server information (Section 4.1);
2. A connection module that connects various database platforms based on the *DB-RDF* file. These connections enable querying of time-series data from relational databases (Section 4.2);
3. A mapping module that integrates contextual data and time-series data used for building performance definition. The integration mechanism uses SPARQL queries and SQL queries (Section 4.3).

4.1. DB-RDF — a database schema representation in RDF

Different databases are designed with different schemas and are deployed on different database platforms. In order to access different databases the schema and platform information are represented as instances of RDF(S) or OWL ontologies, which are subsequently stored in a *DB-RDF* file (Fig. 5).

The *DB-RDF* file structure contains seven nodes. The root node stores all database platform-relevant properties as well as the database schema. Table nodes store time-series data and comprise separate column nodes. Additional nodes store store rules for formatting values stored in database fields if required.

BMSs usually use time or location variables to index time-series data from sensors. Specifically designed rules match tables based on time or location variables. Each rule parameter is stored in a parameter node construct and other constructs hold translation rules that convert date or location parameters between different formats, e.g. the first month of a year can be represented with 01, 1, Jan., January, etc.

4.2. A connection module to access BMS databases

Many mature relational database platforms efficiently store BMS time-series data. However, there are some structural differences between the databases that prevent them from being accessed in a consistent way. The connection module addresses this issue by using established connection pool technology [56] to efficiently manage the database connections (Fig. 6).

The connection module loads server information from the *DB-RDF* file, including drivers for various platforms such as Mysql, SQL Server and Oracle. Server information is also used to compose a specific URL pointing to a database deployed on a particular server. For example, if the database (named 'record') in a building is deployed on the Mysql platform and a local server (IP: 127.0.0.1, Port: 3306), a specific string (jdbc:mysql://127.0.0.1:3306/record) is automatically generated to access the database.

The connection module waits for SQL queries from the mapping module and maintains the status of connections if an SQL query is not successfully executed. The connection module also creates a connection pool to manage connections and statements related to connected databases. Other modules can easily obtain a connection or a statement. This mechanism efficiently creates connections and statements, removing the need for a significant time and resource overhead. The connection module also achieves a loosely coupled and safer system by only allowing the connection pool to create and manage connections and statements.

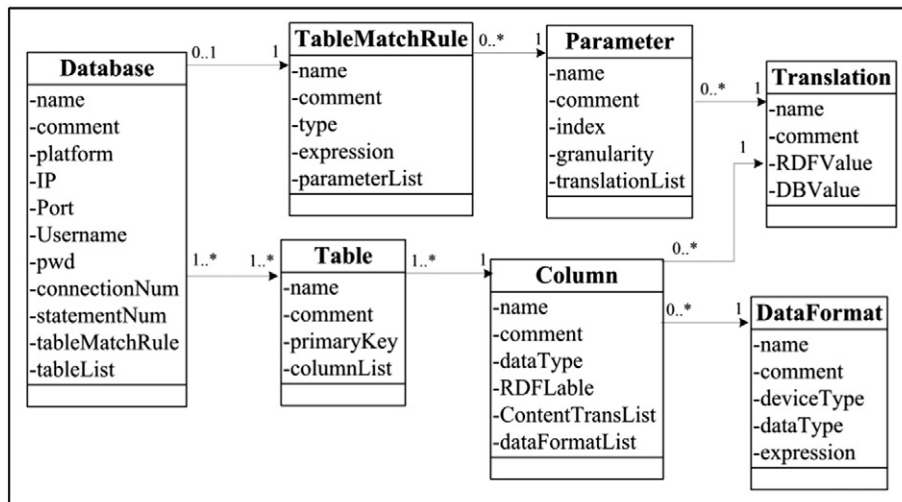


Fig. 5. The structure of the *DB-RDF* file used to automatically connect to database servers and compose SQL queries.

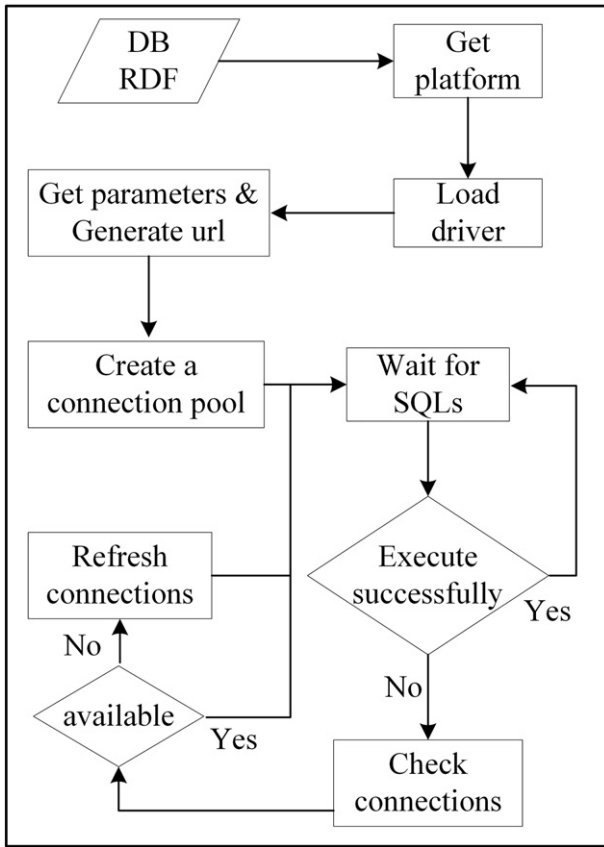


Fig. 6. Workflow of the connection module, which creates and manages connections to relational databases.

4.3. A mapping module for performance definitions and native data sources

The newly developed mapping module integrates time-series performance data with contextual building data. Fig. 7 presents an example use case for the mapping module and this example carries through this section. The mapping module generates a table-index-tree through assuming that records of each room within the building in question are stored in a table. The building node (e.g. SLLS) is the root node of the tree and has two child nodes, one for each storey (storey 0 and storey 1) which are in turn connected to relevant room nodes (e.g. Pharmacy, Pool Hall and Theatre). Each room node connects to a table node containing three columns that store parameters id, time and value respectively (Fig. 7B).

For a given performance definition from the sources layer (e.g. relative humidity quantification between an upper and lower bound), the mapping module extracts the target contexts, i.e. building objects, and data streams required by performance indicators. In Fig. 7B the 'Pool Hall' is the target building object.

Furthermore, the mapping module constructs a list of relevant sensors (one humidity sensor in this case) by generating SPARQL queries to identify particular building objects and sensor information (Fig. 8). Subsequently, the mapping module uses the sensor list and time period information to automatically generate SQL queries for time-series data based on the schema information stored within the DB-RDF file. It is important to note that a target building object may contain child objects, and multiple sensors of the same type (e.g. humidity sensors) might be deployed in large rooms such as a 'Pool Hall' to track a single performance indicator. Accordingly, the mapping module preserves a sufficient level of detail by containing a specific data structure to store time-series data and related properties.

A key feature of this hybrid architecture is the link between SPARQL and SQL queries (Fig. 9). Once the performance analysis context and the device types have been identified, then SQL queries extract the relevant time series data from the relational database. The mapping module automatically generates SQL queries by integrating device information and database schema information contained within DB-RDF documents. In order to increase query efficiency for cases where match rules exist in database tables, target building object information and time period information are analysed by accessing specifically appointed database tables (Fig. 9).

The hybrid architecture is now tested for building performance evaluation and query efficiency perspectives.

5. Testing performance assessment using the hybrid approach

Two use-cases were created to evaluate the hybrid architecture concept. These cases are not intended to serve as an exhaustive exploration of the viability of these data sources as indicators of building performance, but as an illustration of how time-series data sources in BMSs can be accessed and linked to other building context data. In particular, we look to qualify if additional engineering insight is available when compared against equivalent data available from single sources such as a BMS.

In the first case, we integrate time-series data from the UCD sports centre with context data from an original BIM source, together with output from a Building Energy Performance Simulation (BEPS) model. This integration enables the evaluation of measured building operation against previously defined, rule-based performance criteria and quantifies the performance of a significant yet unmeasured heating coil. This experiment is designed to show how time-series data and rudimentary

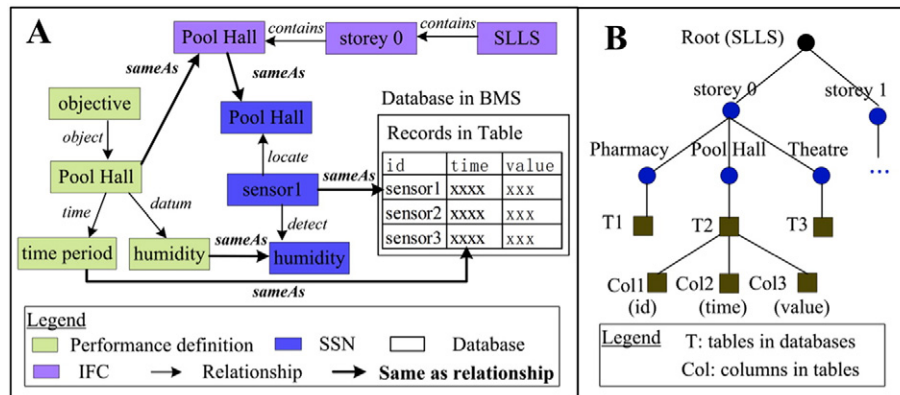


Fig. 7. A) Linked data example for a building with a pool hall containing a humidity sensor with records stored in a time-series database and B) structure of the corresponding table index tree.

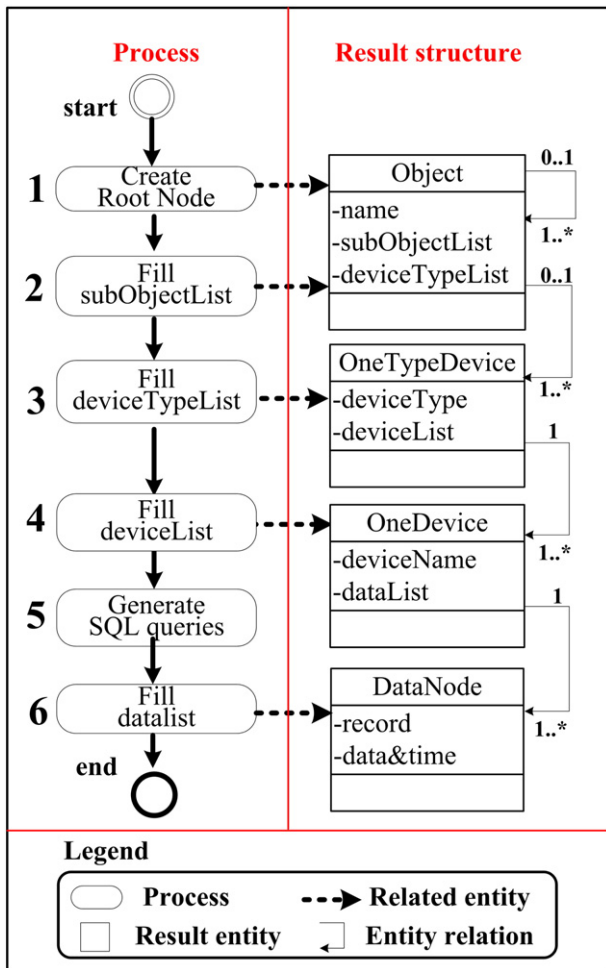


Fig. 8. Workflow and the result structure of the mapping module preparing data stream for performance definition.

calendar information can be integrated with semantic building data to allow a greater degree of understanding of building performance.

The second use-case focuses on the comparison of query times for data stored in two different environments. The first is the hybrid architecture approach presented in this paper and the second stores the same time-series data but in linked data format.

5.1. Test case 1: integrating context and time-series data

The UCD Student Learning, Leisure and Sports (SLLS) Complex is an 11,000 m² facility, spread over three-storeys. This evaluation focuses on the 50 m × 25 m swimming pool area and associated AHU. The building uses a Unitron BMS, manufactured by Cylon Controls Ltd., Ireland, for all building controls and ancillary equipment.

In order to demonstrate the applicability of the hybrid approach, this case uses a linked data environment to integrate a number of previously disconnected data sources (Fig. 10).

The hybrid-architecture links the following information:

1. Time-series data from Cylon sensors and metres as well as output from the EnergyPlus (E+) model.
2. A DB-RDF file representing database schema information and server information of the BMS.
3. EnergyPlus model of the SLLS facility originally modelled in the Simergy interface.
4. Selected rule-based performance definition are set and stored in an RDF file (Fig. 11).

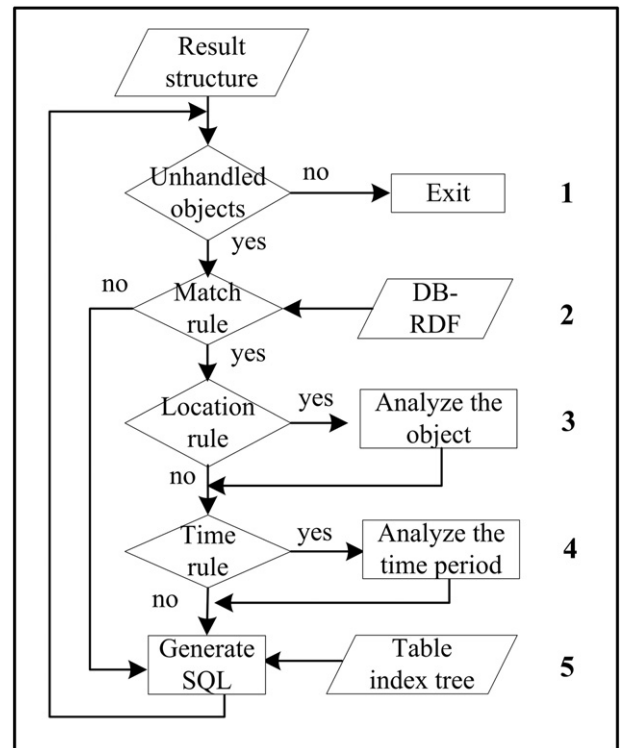


Fig. 9. The workflow for the composition of SQL queries to extract time-series data in databases.

5. Building geometry information, based on a Building Information Model (BIM) of the building, which is transformed from an IFC file into an IFC-RDF file [40]. The source provides object information about the building, which is used to link sensor information for each object.
6. In order to provide context data for the test case, we converted sensor information, stored in an Excel file, into a RDF file based on rules defined in Table 1.

This test case uses a rule-based performance assessment mechanism called the scenario modelling method [9], which can leverage raw or unprocessed data during all phases of the building life cycle. Scenario models present different aspects of building performance in parallel and allow for a more complete or holistic perspective on local and global performance.

For this experiment, a scenario model reflects some key concerns in the area of comfort and energy consumption in the building (Fig. 11) by examining building function and energy consumption of the Pool Hall AHUs simultaneously. In this case, the hybrid architecture uses the Pool Hall instance from the IFC-RDF source as the analysis context.

The objective is to compare traditional analysis based on data from the Cylon system (Fig. 12a) against an equivalent analysis that also uses available data from the outputs of a building energy simulation model (Fig. 12b). The intention is to examine the capabilities of the hybrid architecture in terms of presenting meaningful additional engineering information, through which building managers can make informed decisions.

The key issue for this test case is that the heating coils are not adequately metered. The quantification of their energy consumption requires additional equipment. The first performance objective details the air temperature within the swimming pool hall zone. Each objective uses a functional metric, which simply returns the current air temperature value for the zone at a given time. The second performance objective focuses on the energy consumption of the heating coils within AHU1 and AHU2 that serve the pool hall zone. The two objectives, taken together, constitute a scenario model (Fig. 11).

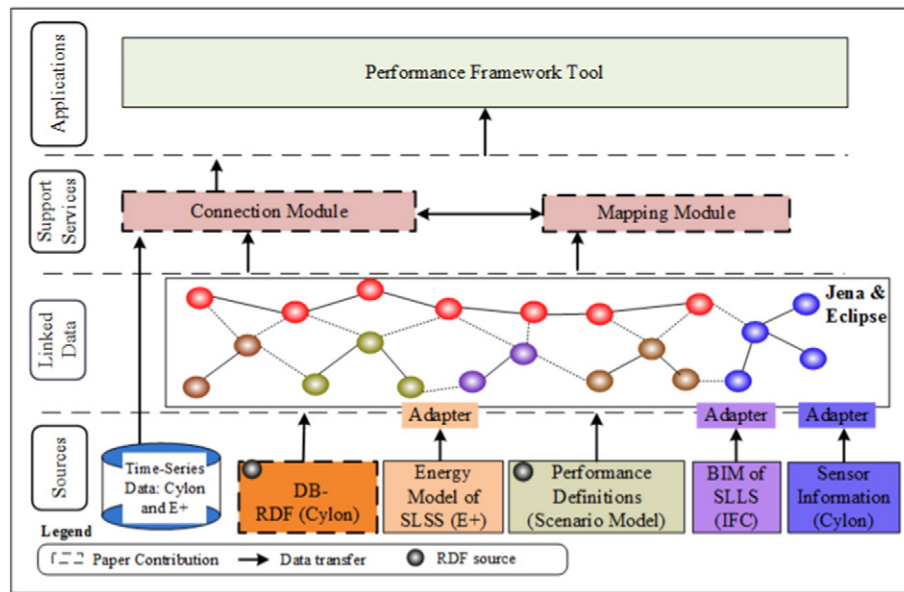


Fig. 10. Implementation diagram for test case 1 highlight the sources of information and the technologies that comprise the hybrid-architecture.

When the model is considered over a 24-h period, a clear picture emerges of the comfort conditions and energy consumption in the zone. Such processed and precise information is enormously beneficial for building managers, who typically have limited resources available to process large volumes of disconnected information. A closer look at the temperature objective indicates that the zone conditions adhere with the specified functional intent.

The temperature of the space remains at 30 °C while staying within relative humidity bounds of 40–60%. With regard to associated energy consumption, heating coil energy output is relative high before 11 am, compared with the rest of the day. The cause for which is that the outside air dry-bulb temperature is lower up to 11 am and as result more energy is used for heating.

By coupling the outputs of a calibrated energy simulation model with available data from the BMS, a building manager can examine

one of the largest energy consumers in the facility without additional workload, cost or effort. This work is the first step in optimising the operation of the AHUs that service the pool hall area.

5.2. Test case 2: time efficiency evaluation of hybrid architecture and pure RDF data sources

The second test case examines the scalability of the hybrid architecture approach, particularly in terms of query efficiency for large volumes of time series data. This test case compares query efficiencies for data stored in standard relational databases and accessed by the hybrid architecture against time-series data stored in pure RDF format.

The approach taken used a significant volume of hypothetical data for a theoretical modern office building with a number of different sensors in fifteen rooms (details shown in Table 2).

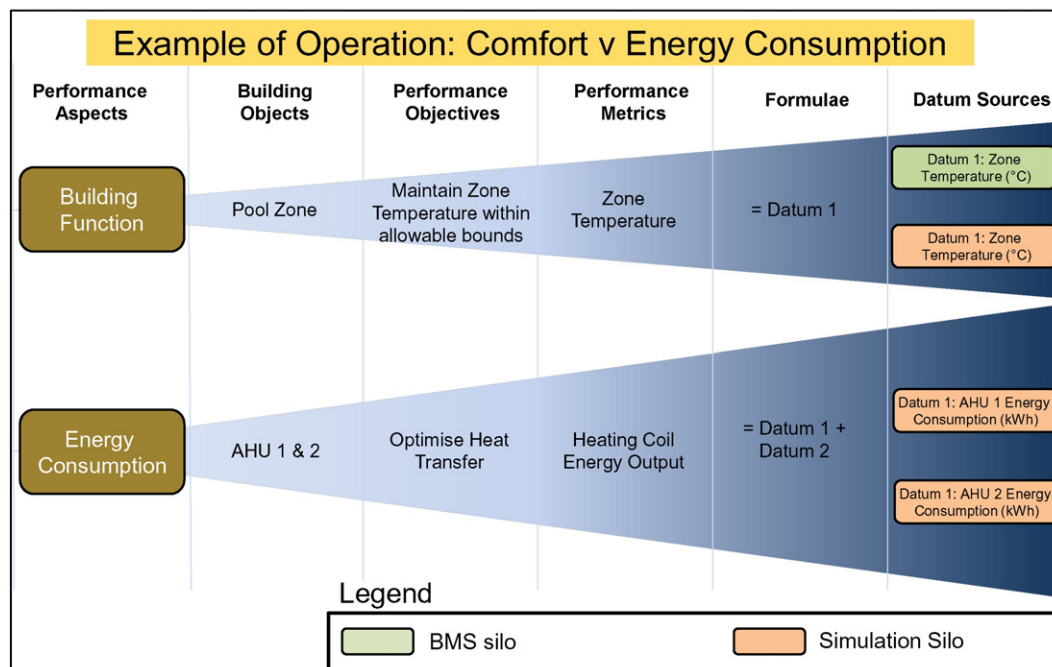


Fig. 11. Detailed information for test scenario. This model highlights building function, represented by temperature, and energy consumption of related AHUs.

Table 1

Vocabulary for presenting sensors in the building into RDF format.

RDF property	Note	Example
Resource [57]	Sensor name	Sensor1
hasLocation [58]	Sensor location	Room101
Observed property [58]	Observed property	Temperature
Feature Of interest [58]	Entity related to observed property	Room101
UOM [59]	Unit of measurement	°C

This test case generated data at ten frequencies, from 1 min per record to 10 min per record with a linear increase, to represent different volumes of data. This resulted in approximately 80 million records per month with the frequency of 1 min per record. Two methods are used to store these data for the evaluation: the first method stores data in MySQL database tables and this corresponds with the approach used in this paper. The second transforms time-series data into RDF format, storing it on the Openlink Virtuoso platform.

The queries leverage a task generator that creates rudimentary yet automated rule definitions at five different frequencies. The frequency of each analysis query varies from 1 s per query-objective to 0.2 s and the analysis interval is set to 24 h.

Both the efficiency evaluation programme and the data management programme are installed on the same computer. The computer uses a 64-bit Window 7 operating system, 64-bit MySQL platform and 64-bit Virtuoso platform, and has an E8500 CPU, 4 GB RAM (DDR2, 667 MHz), and a 250 GB hard disc (8 MB cache, 7200 RPM). This specification is considered better or equal to hardware on which most BMSs reside.

Table 3 illustrates the average time of the efficiency evaluation for different data and task frequencies. Where M denotes the MySQL approach and V indicates the Virtuoso approach. Each cell represents the average response time for 5000 tasks. Multi thread technology is used by the programme to handle tasks from the task generator, so response times are independently recorded by each thread. It is from the point when a task is generated to the point of obtaining time-series data from databases.

Both methods achieve stable performance when the data frequency is less than 20 record/h. The response time of MySQL is around 60 milliseconds (ms) and the response time of Virtuoso is around 80 ms. Some negligible fluctuations are due to the caching mechanism of the operation system. In cases with additional reading load, i.e. data frequency ≥ 20 record/h, the response time starts to rise. Significantly larger variations arise for recording frequencies of 60 and 30 record/h.

In addition, tables generated with different data frequencies hold the same row numbers and indexing information. The reason for the noticeable differences in query time is that the reading load from requests already exceeds bottlenecks of the server hard disc. Results in the first row and the second row of Table 3 can be improved with a faster hard

Table 2

Sensors deployed in each room of the hypothetical building.

Device type	Instances
Temperature sensor	3
Humidity sensor	2
Air-speed sensor	1
Infrared sensor	3
Smoke sensor	2
Illumination sensor	2
CO ₂ sensor	2

disc. With development of hardware technology, general RAM and hard disc devices already achieve more efficient performance. In addition, most BMSs collect time-series data with a lower frequency (usually 4 record/h) and the objective generating frequency for a building is lower than 1 task/s. Therefore, both approaches are sufficient for lower resolutions of time-series data but the hybrid approach can provide sufficient efficiency for performance data of contemporary and future buildings.

6. Conclusion

The lack of adequate integration of building performance data with other contextual building data has long been an issue for the AEC industry. The inability to consider building information holistically hinders the optimisation of building performance and the narrowing of the gap between actual performance and design intent.

Many of the existing approaches to data integration in the industry call for significant conversion efforts of native data formats to a central data format, or the creation of adapters to link differing formats together. Other approaches describe information semantically. This work recognises that not all AEC information needs to be integrated and that it is usually more efficient to leave data such as performance measurements in their native format.

This paper describes and implements a novel method of linking traditionally disconnected data and information and constructing integrated data sources to enable in-depth and insightful building performance assessment. The approach adopts a three-step methodology (the DB-RDF, the connection module and the mapping module) which governs the presence the representing of database scheme and platform information, the access to data stores and the mapping of time-series data to contextual building data.

The approach provides a hybrid architecture on top of the existing building management infrastructure, which enables the provision of a cross-domain information assessment function. Based on this new development, building performance can be comprehensively assessed using existing performance assessment approaches, without significant reinterpretation and conversion from building managers. In addition, the design of the hybrid architecture provides sufficient efficiency of

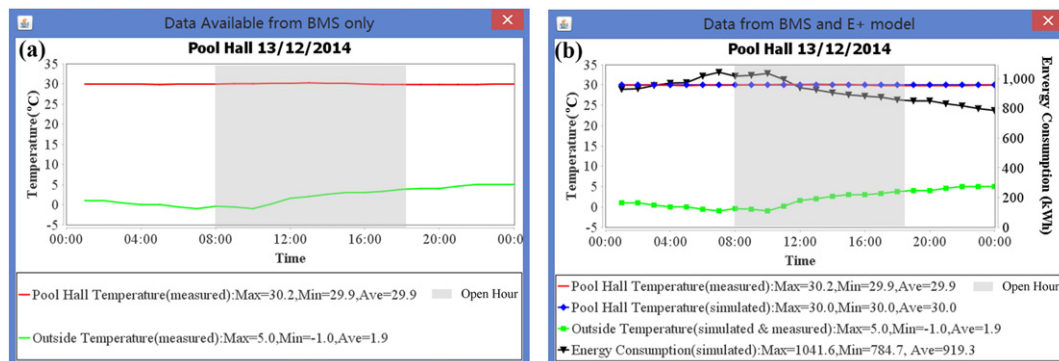


Fig. 12. (a) Zone temperature and relative humidity (b) total energy consumption of heating coils in AHU1 and 2.

Table 3
Average response time of efficiency evaluation (ms).

Data frequency (record/h)	Task frequency (task/s)									
	5		2.5		1.67		1.25		1	
	M	V	M	V	M	V	M	V	M	V
60	4620	10,837	1885	4620	842	1820	580	1106	547	1019
30	1120	3085	428	1216	230	470	120	271	118	250
20	102	789	82	257	68	142	64	128	63	115
15	64	88	61	85	60	83	62	81	63	84
12	63	84	61	83	60	82	61	81	61	83
10	61	82	60	81	59	81	62	82	62	84
8.6	62	82	59	80	60	80	61	82	63	82
7.5	61	80	60	78	59	79	62	81	61	84
6.7	60	78	58	77	58	78	60	81	62	85
6	59	79	58	75	59	78	61	82	63	84

data integration, which is essential for real-time building performance assessment, especially for analysis cases involving a large number of buildings. Furthermore, this approach could provide financial controllers with new insights into asset and organisational energy consumption and environmental performance, particularly at different levels of granularity.

With time-series data integrated by the hybrid solution, abnormal behaviours in buildings can be detected using Fault Detection and Diagnosis (FDD) algorithms for building operation optimisation in the future.

Acknowledgements

This work is supported by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Programme, project title SuPerB (Project number 631617).

Funding narrative: Dr. Will Turner, ERC GI, is funded under the Programme for Research in Third Level Institutions PRTL (Cycle 5) and co-funded under the European Regional Development Fund (ERDF).

References

- [1] M. Hamdy, A. Hasan, K. Siren, A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010, *Energy Build.* 56 (2013) 189–203.
- [2] P. De Wilde, The gap between predicted and measured energy performance of buildings: a framework for investigation, *Autom. Constr.* 41 (2014) 40–49.
- [3] K.J. Butcher, Guide F: Energy Efficiency NEW 2012, The Chartered Institution of Building Services Engineers, 2012 (978-1-906846-22-0).
- [4] EVO, International Performance Measurement & Verification Protocol. Core Concepts (English-EVO 10000-1:2014), Efficiency Valuation Organization, June 2014 (URL: www.evo-world.org).
- [5] G.C.B.B.W. Group, The performance gap: causes & solutions, 2012 (URL: http://www.greenconstructionboard.org/images/stories/pdfs/performance-gap/2013-03-04%20Closing%20the%20Gap_Final%20Report_ISSUE.pdf).
- [6] N. GCR, Cost Analysis of Inadequate Interoperability in the US Capital Facilities Industry, National Institute of Standards and Technology (NIST), 2004.
- [7] J. Schein, S.T. Bushby, N.S. Castro, J.M. House, A rule-based fault detection method for air handling units, *Energy Build.* 38 (2006) 1485–1492.
- [8] J. Granderson, Energy Information Handbook: Applications for Energy-Efficient Building Operations, 2013.
- [9] J. O'Donnell, M. Keane, E. Morrissey, V. Bazjanac, Scenario modelling: a holistic environmental and energy management method for building operation optimisation, *Energy Build.* 62 (2013) 146–157.
- [10] E. Corry, A Semantic Web Approach to Enable the Holistic Environmental and Energy Management of Buildings (Ph.D. thesis) Informatics Research Unit for Sustainable Engineering, National University of Ireland Galway, 2014 (URL).
- [11] E. Corry, J. O'Donnell, E. Corry, D. Coakley, P. Pauwels, M. Keane, Using semantic web technologies to access soft AEC data, *Adv. Eng. Inform.* 28 (2014) 370–380.
- [12] T. Liebich, Y. Adachi, J. Forester, J. Hyvarinen, S. Richter, T. Chipman, et al., Industry Foundation Classes IFC4 Official Release, 2015.
- [13] D. Tzovaras, T. Brennan, D1.3 state-of-art and industry analysis, Adapt4EE occupant aware, intelligent and adaptive enterprises, 2012 (URL: <http://www.adapt4ee.eu/adapt4ee/files/document/deliverables/Adapt4EE-Deliverable-D1.3.pdf>).
- [14] T. Liebich, K. Stuhlmacher, P. Katranuschkov, R. Guruz, D2.1 BIM Enhancement Specification, Platform for Holistic Energy Efficiency Simulation and Lifecycle Management of Public Use Facilities (HESMOS) – Integrated Virtual Lab, Project ICT 26088, 2011 (URL).
- [15] M. Look, D1.2 Report Detailing Neighbourhood Information Model Semantics, COOPERATE Control and Optimisation for Energy Positive Neighbourhoods, 2013 (URL).
- [16] T. Greifengberg, M. Look, B. Rumpe, Integrating Heterogeneous Building and Periphery Data Models at the District Level: the NIM Approach, 2014 (arXiv preprint arXiv:1412.2961).
- [17] P. Pauwels, Reconsidering Information System Support for Architectural Design Thinking (Ph.D. thesis) Ghent University, 2012 (URL: <https://biblio.ugent.be/publication/2919567>).
- [18] P. Pauwels, Supporting decision-making in the building life-cycle using linked building data, *Buildings* 4 (2014) 549–579.
- [19] D.L. McGuinness, F. Van Harmelen, OWL Web Ontology Language Overview, W3C Recommendation, 10 (2004), 2004.
- [20] G. Klyne, J.J. Carroll, Resource description framework (RDF): concepts and abstract syntax, 2006 (URL: <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>).
- [21] J. O'Donnell, E. Corry, S. Hasan, M. Keane, Building performance optimization using cross-domain scenario modeling, linked data, and complex event processing, *Build. Environ.* 62 (2013) 102–111.
- [22] E. Corry, J. O'Donnell, E. Corry, S. Hasan, M. Keane, S. O'Riain, Linking building data in the cloud: integrating cross-domain building data using linked data, *Adv. Eng. Inform.* 27 (2013) 206–219.
- [23] P. Pauwels, D.V. Deursen, R. Verstraeten, J.D. Roo, R.D. Meyer, R.V.d. Walle, et al., A semantic rule checking environment for building performance checking, *Autom. Constr.* 20 (2011) 506–518.
- [24] P. Pauwels, D.V. Deursen, J.D. Roo, T.V. Ackere, R.D. Meyer, R.V.d. Walle, et al., Three-dimensional information exchange over the semantic web for the domain of architecture, engineering, and construction, *Artif. Intell. Eng. Des. Anal. Manuf.* 25 (2011) 317–332.
- [25] P. Barnaghi, W. Wang, C. Henson, K. Taylor, Semantics for the internet of things: early progress and back to the future, *Int. J. Semant. Web Inf. Syst.* 8 (2012) 1–21.
- [26] J. Euzenat, P. Shvaiko, Ontology Matching, Springer, Heidelberg, 2007 (978-3-642-38721-0).
- [27] T. Maile, V. Bazjanac, M. Fischer, A method to compare simulated and measured data to assess building energy performance, *Build. Environ.* 56 (2012) 241–251.
- [28] A.C. Menezes, A. Cripps, D. Bouchlaghem, R. Buswell, Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap, *Appl. Energy* 97 (2012) 355–364.
- [29] T.W. Kim, M. Fischer, Automated generation of user activity-space pairs in space-use analysis, *J. Constr. Eng. Manag.* 140 (2014) 04014007.
- [30] E.F. Codd, Relational database: a practical foundation for productivity, *Commun. ACM* 25 (1982) 109–117.
- [31] M.A. Bornea, J. Dolby, A. Kementsietsidis, K. Srinivas, P. Dantressangle, O. Udrea, et al., Building an efficient RDF store over a relational database, The 2013 ACM SIGMOD International Conference on Management of Data 2013, pp. 121–132.
- [32] G. Augenbroe, The role of simulation in performance based building, building performance simulation for design and operation, in: J.L. Hensen, L. Roberto (Eds.), *Building Performance Simulation for Design and Operation*, Spon Press, London 2011, pp. 15–36 (978-0415474146).
- [33] G. Augenbroe, C.-S. Park, Quantification methods of technical building performance, *Build. Res. Inf.* 33 (2005) 159–172.
- [34] D. Pati, C.-S. Park, G. Augenbroe, Roles of building performance assessment in stakeholder dialogue in AEC, *Autom. Constr.* 15 (2006) 415–427.
- [35] R.J. Hitchcock, Standardized Building Performance Metrics, Lawrence Berkeley National Laboratory, 2003.
- [36] C. Neumann, D. Jacob, Results of the project building EQ: tools and methods for linking EPBD and continuous commissioning Fraunhofer ISE, 2010 (URL: http://www.buildingeq-online.net/fileadmin/user_upload/Results/BEQ_publishable_results_report_final_rev_100624.pdf).
- [37] E. Corry, P. Pauwels, S. Hu, M. Keane, J. O'Donnell, A performance assessment ontology for the environmental and energy management of buildings, *Autom. Constr.* 57 (2015) 249–259.
- [38] E. Corry, S. Hasan, S. O'Riain, Enterprise energy management using a linked dataspace for energy intelligence, Sustainable Internet and ICT for Sustainability (SustainIT), 2012, IEEE 2012, pp. 1–6.
- [39] J. Steel, R. Drogemuller, B. Toth, Model interoperability in building information modelling, *Softw. Syst. Model.* 11 (2012) 99–109.
- [40] S. Törmä, N.V. Hoang, P. Pauwels, Tool: IFC-to-RDF conversion tool, 2013 (URL: <http://linkedbuildingdata.net/tools/tool-ifc-to-rdf-conversion-tool/>).
- [41] J. O'Donnell, R. See, C. Rose, T. Maile, V. Bazjanac, P. Haves, SimModel: a Domain Data Model for Whole Building Energy Simulation, Ibpsa Building Simulation 2011 Sydney, Australia, 2011.
- [42] P. Pauwels, E. Corry, J. O'Donnell, Representing SimModel in the Web Ontology Language, American Society of Civil Engineers, 2014 2271–2278.
- [43] M. Compton, P. Barnaghi, L. Bermudez, R. García-Castro, O. Corcho, The SSN ontology of the W3C semantic sensor network incubator group, *Web Semant. Sci. Serv. Agents World Wide Web* 17 (2012) 25–32.
- [44] A.H. Kazmi, M.J. O'grady, D.T. Delaney, A.G. Ruzzelli, G.M.P. O'hare, A review of wireless-sensor-network-enabled building energy management systems, *ACM Trans. Sens. Networks* 10 (2014) 66:61–66:43.
- [45] F. Priyatna, O. Corcho, J. Sequeda, Formalisation and experiences of R2RML-based SPARQL to SQL query translation using morph, The 23rd International Conference on World Wide Web, ACM, New York, NY, USA 2014, pp. 479–490.
- [46] R. Colomo-Palacios, J.L. Sánchez-Cervantes, G. Alor-Hernández, A. Rodríguez-González, Linked data: perspectives for IT professionals, *Int. J. Hum. Cap. Inf. Technol. Prof.* 3 (2012) 1–15.
- [47] C. Liu, J. Urbani, G. Qi, Efficient RDF stream reasoning with graphics processing units (GPUs), Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, Seoul, Republic of Korea 2014, pp. 343–344.

- [48] C.P.d. Laborda, S. Conrad, Relational.OWL: a data and schema representation format based on OWL, The 2nd Asia-Pacific Conference on Conceptual Modelling 2005, pp. 89–96.
- [49] M. Hert, G. Reif, H.C. Gall, A comparison of RDB-to-RDF mapping languages, The 7th International Conference on Semantic Systems, ACM, New York, NY, USA 2011, pp. 25–32.
- [50] P.T.T. Thuy, N.D. Thuan, Y. Han, K. Park, Y.-K. Lee, RDB2RDF: completed transformation from relational database into RDF ontology, The 8th International Conference on Ubiquitous Information Management and Communication, ACM, New York, NY, USA, 2014.
- [51] T. Bosch, R. Cyganiak, A. Gregory, J. Wackerow, DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data, LDOW, 2013.
- [52] P. Cudré-Mauroux, I. Enchev, S. Fundatureanu, P. Groth, A. Haque, A. Harth, et al., NoSQL Databases for RDF: An Empirical Evaluation, The Semantic Web – ISWC 2013, Vol. 8219, 2013 310–325.
- [53] J.T. O'Donnell, Specification of Optimum Holistic Building Environmental and Energy Performance Information to Support Informed Decision Making (Ph.D. thesis) Department of Civil and Environmental Engineering, University College Cork, 2009 (URL: <http://cora.ucc.ie/handle/10468/603>).
- [54] H. Neuhaus, M. Compton, The Semantic Sensor Network Ontology, AGILE Workshop on Challenges in Geospatial Data Harmonisation, Hannover, Germany 2009, pp. 1–33.
- [55] P. Pauwels, W. Terkaj, EXPRESS to OWL for construction industry: towards a recommendable and usable ifcOWL ontology, Autom. Constr. 63 (2016) 100–133.
- [56] H. Bergsten, Improved performance with a connection pool, <http://www.webdevelopersjournal.com/1999>.
- [57] The RDF Concepts Vocabulary (RDF), 1999 (URL:) <http://www.w3.org/1999/02/22-rdf-syntax-ns>.
- [58] Semantic Sensor Network Ontology, 2005 (URL:) <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn>.
- [59] SPITFIRE Ontology, 2010 (URL:) http://spitfire-project.eu/cc/spitfireCC_n3.owl.