Using semantic web technologies to access soft AEC data

Edward Corry[a,*], James O'Donnell[b,c], Edward Curry[d], Daniel Coakley[a], Pieter Pauwels[e], Marcus Keane[a]

[a]*Informatics Research Unit for Sustainable Engineering, National University of Ireland Galway, Galway, Ireland*
[b]*Electricity Research Centre and UCD Energy Institute, School of Mechanical and Materials Engineering, University College Dublin, Belfield, Dublin 4, Ireland*
[c]*Building Technologies and Urban Systems Department, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, MS 90R3111, Berkeley, CA 94720*
[d]*Digital Enterprise Research Institute, National University of Ireland Galway, IDA Business Park, Lower Dangan, Galway, Ireland*
[e]*Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands*

Abstract

Building related data tends to be generated, used and retained in a domain-specific manner. The lack of interoperability between data domains in the architecture, engineering and construction (AEC) industry inhibits the cross-domain use of data at an enterprise level. Semantic web technologies provide a possible solution to some of the noted interoperability issues. Traditional methods of information capture fail to take into account the wealth of soft information available throughout a building. Several sources of information are not included in performance assessment frameworks, including social media, occupant communication, mobile communication devices, occupancy patterns, human resource allocations and financial information.

The paper suggests that improved data interoperability can aid the integration of untapped silos of information into existing structured performance

*Corresponding author

Email addresses: `edwardcorry@nuigalway.ie` (Edward Corry), `james.odonnell@ucd.ie` (James O'Donnell), `ed.curry@deri.org` (Edward Curry), `Daniel.coakley@nuigalway.ie` (Daniel Coakley), `p.pauwels@uva.nl` (Pieter Pauwels), `marcus.keane@nuigalway.ie` (Marcus Keane)

Preprint submitted to Advanced Engineering Informatics *August 11, 2014*

measurement frameworks, leading to greater awareness of stakeholder concerns and building performance. An initial study of how building-related data can be published following semantic web principles and integrated with other 'soft-data' sources in a cross-domain manner is presented. The paper goes on to illustrate how data sources from outside the building operation domain can be used to supplement existing sources. Future work will include the creation of a semantic web based performance framework platform for building performance optimisation.

Keywords: social media, Twitter, linked data, performance metrics, building performance, RDF

---

1. Introduction

"You cannot manage what you do not measure." Many interested parties in the AEC domain have long placed this timeless concept as a central driver of their work [1]. In order to produce, and more importantly, operate buildings to the satisfaction of owners, occupants and legislators, a keen understanding of performance assessment and measurement is required. Decision makers need access to the information and tools required to cost-effectively assure the desired performance of buildings [2]. The lack of interoperability manifested in poor electronic data exchange, management and access has a significant cost [3] to the decision making process in general. In order to ensure optimal performance, several studies have shown that one must continually measure and monitor performance [4, 5, 6]. Modeling, measuring and benchmarking of building performance is set to become the industry norm [7] as more types of data become more available. Building performance, in the context of this paper, is defined as the deliver of functional intent of each zone in the building while accounting for the energy and cost of delivering this functional intent.

Traditionally, buildings have been managed using a small subset of the data available in a building, namely the data that is made available via building management systems (BMS). Well-recognised interoperability issues and a lack of cross-domain data exchange [8] preclude the integration of many other building data sources with existing BMS information. Successful optimisation efforts require an integrated solution including a performance assessment framework, integrated data sources and an information delivery system tailored to the skill-set of the key building stakeholder(s) [9].

This work is primarily intended to show how diverse streams of information can be captured and linked with other building data to broaden the range of data silos available for building performance optimisation. Two very different ‘soft‘ information sources, scheduling data and continuous occupant feedback, are used as initial examples of the type of soft information available in buildings. by way of a case study, the paper illustrates how these sources might be integrated into an overall assessment strategy. The paper shows primarily how semantic web technologies can be used to facilitate the required type of cross-domain data use. Finally, the paper discusses how the integration of softer data sources with such an assessment strategy could potentially resolve some of the issues outlined in this introduction.

The integration of building data using semantic web technologies was previously explored [10, 9, 11]. The resulting data structure was used to drive a building energy assessment dashboard [9]. A comprehensive performance assessment framework was illustrated in [10] for use throughout the life-cycle of the building. It showed how this approach could be integrated with existing data sources available in buildings. This paper suggests that other sources of data, outside the traditional building management systems (BMS), are available in modern buildings, often in electronic format and represent an untapped resource which can enable a greater level of cross-domain communication and engagement amongst building stakeholders. The paper explores how some of these sources could be incorporated with other building data using semantic web technologies.

These data sources are often not used in a cross-domain manner due to inertia, interoperability issues and a lack of an adequate framework into which the sources can be added. Some of the sources also tend to be hard to interpret due to the qualitative nature of the data and the lower level of trustworthiness in some cases. The paper illustrates how some of these issues can be overcome and pose the question, what can be achieved with these extra data sources?

Robust building management techniques and systems can be supplemented to include a broader interpretation of building performance, beyond typical concerns, such as energy consumption and system performance. Broader concerns regarding building operation, including cross-domain data sharing and stakeholder interaction, can also be considered when data is more easily accessible. Efforts have been made to improve interoperability in the AEC domain, particularly the various building information modeling (BIM) initiatives and processes used to describe information transfers between domains [12]. The paper generally describes the problems associated with current methods of information exchange in the AEC industry and in particular around the disjointed area of building performance assessment. Building on previous work [9, 11], the paper briefly describes how currently untapped data sources may be exposed using semantic web technologies, and interpreted using a proven technique to provide a more structured assessment of building performance, together with the more traditional sources of building performance data. The paper goes on to show how this technique may be extended to include a range of 'soft' data sources, along with more traditional hard data sources.

2. Accessing Diverse Data Sources in the Building Operation Phase

2.1. Information Exchange in Buildings: Semantic Web Technologies in the Performance Framework Tool

The Performance Framework Tool (PFT) has been conceived by the authors as a means for deriving enhanced meaning from building data sources, based on the performance metric concept [13]. The structured decision making framework is mainly aimed at providing the key building stakeholder, the building manager, with the information needed to make informed and repeatable decisions regarding the operation of a facility. It does this by providing the end user with useful information from diverse domains. Furthermore, the tool is intended to serve as an aid to building performance assessment across the building life cycle, allowing the integration of design and simulation data sources with real performance data. The PFT depends on access to various data sources from the building and the greater the range available, the more informative the tool may become.

Central to the PFT (and building management) is the integration of information from various domains. No building stakeholder retains (or can retain) a complete picture of all building-related information and although the building manager can access perhaps the greatest range of information about a building and its performance, typically, building information is created, maintained and lost by many stakeholders throughout the building life-cycle [12]. This loss of information and lack of interoperability across domains has been well documented [14, 15, 3]. Several initiatives have been made to develop technologies [16, 17, 18, 19, 20, 11, 9] and define procedures [13, 21] to capture and retain information amongst various stakeholders and across domains. However, due to the lack of information interoperability, it is (near to) impossible to get a cross-domain view of a building in terms of interaction of data streams in a clear and structured manner. It is not the purpose of this tool to provide such a complete view. Instead, the PFT tool aims at providing access to various information sources, so that the building manager gets the option to choose the criteria according to which he assesses building performance.

Considering the building as a whole, there are several streams of data that currently exist to serve particular domains and remain untapped in the building performance sphere. A detailed analysis of the integration challenges is provided by Shen et al. [8]. Technologies are emerging which can bridge the interoperability gap across several domains in the AEC industry. New information exchange definitions are being generated to describe all manner of domains, including such diverse areas as curtain wall modeling and information handover protocols [22]. Industry and national level organisations have recognised the importance of data management and building information modeling (BIM) in particular and are driving advances in this area by making BIM a requirement of projects [23, 24]. Taken as a whole, advances in the interoperability question pose some very interesting questions as to what use may be made of these technologies to generate an enhanced view of building performance.

Figure 3 illustrates the concept behind exposing previously remote data sources in a Resource Description Framework (RDF) format [25]. The paper identifies ways in which semantic web technologies can serve as a unifying set of technologies aiding interoperability across previously remote data sources. Utilising semantic web technologies, previously unused sets of building data are exposed and integrated with relating datasets. Figure 3 is a representation of the platform this research effort is currently working towards with a view to semantically integrating building data into a performance assessment platform.

2.2. Semantic Web Technologies

The semantic web was conceived in [26] as a network that describes the meaning of its concepts through a directed, labelled graph. Each node in this graph represents a particular concept or object in the world and each arc in this graph represents the logical relation between two of these concepts or objects. When viewed together, the graph represents a set of logic-based declarative sentences. Relationships can then be created between these sentences or 'triples'.

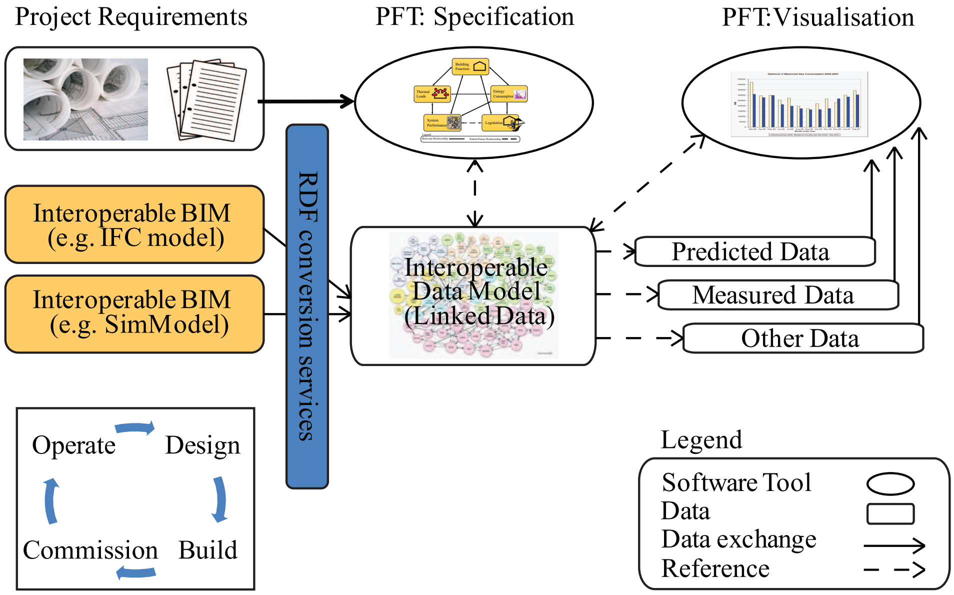


Figure 1: Semantic Web based building performance assessment platform.

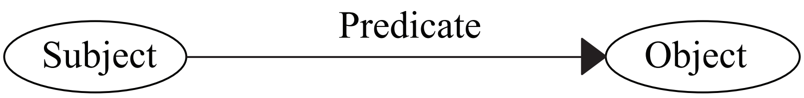
All kinds of data can thus be linked together, resulting in a web of information that both humans and machines can read.

The Resource Description Framework (RDF) [20] is the data model used for information representation. An RDF graph is constructed by applying a logical AND operator to a range of logical statements containing concepts or objects in the world and their relations. These statements are often referred to as RDF triples, consisting of a subject, a predicate and an object, implying directionality in the RDF graph Fig. 4. Every concept described in an RDF graph, whether this be an object, subject or predicate, is uniquely defined through a uniform resource identifier (URI). The resulting RDF graph can be converted into a textual representation that follows a specific syntax [27].

Several triples can be joined together and, in this manner, a collection of information can be exposed. For instance, other information can be published relating to the room, or the other occupants. The strength of the technique lies in the ability to uniquely reference the subject, predicate and object using a URI, allowing data sharing to take place at the data level, rather than the application level.

RDF is especially powerfully when attempting to integrate cross-domain data as a series of triples can be quickly accumulated concerning the same

RDF Triple Structure:



Sample triple (RDF graph):

Sample triple (RDF/XML syntax):

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:vocab="http://www.vocab.deri.ie/">
<rdf:Description rdf:about="http://www.iruse.ie/staff/name/edwardcorry">
  <vocab:roomsoccupant>
    <rdf:Description
      rdf:about="http://www.lab.linkeddata.deri.ie/2010/deri#roomsr202e">
    </rdf:Description>
  </vocab:roomsoccupant>
</rdf:Description>
</rdf:RDF>
```

Sample triple (N3 syntax):

```
@prefix iruse:  <http://www.iruse.ie/staff/name/> .
@prefix vocab:  <http://www.vocab.deri.ie/> .
@prefix deri:  <http://www.lab.linkeddata.deri.ie/2010/deri#> .

iruse:edwardcorry     vocab:roomsoccupant     deri:roomsr202e
```

Figure 2: A triple consists of a subject, predicate and object. Each of these has a unique URI. A sample RDF graph is given in three forms: graph syntax, RDF/XML syntax and N3 syntax.

object. Several vocabularies or ontologies have emerged to describe specific domains of data including FOAF, Dublin Core and SIOC. These vocabularies provide further meaning to domain objects and relationships. An object may be referenced in a number of domains, using different ontologies. This research applies semantic web techniques in the AEC sector to enable greater cross-domain data sharing.

2.3. Hard and Soft Building Data

Hard data sources are understood as sources which are readily accessible to existing BMS and consist of quantifiable data that is easy to aggregate and infer information from. On the other hand, *soft* data sources are sources that are not generally accessible to the building management infrastructure and are often qualitative rather than quantitative in nature, making it difficult to draw particular inferences from.

Modern buildings encompass a diverse range of information domains, between which an acknowledged interoperability deficit exists [3], as illustrated in Fig. 1. The list of building-related data in Fig. 1 is far from exhaustive, but it illustrates how the various domains independently retain an array of building-related data that is most often not integrated with the building management structure or made available on a cross-domain basis. These data sources can serve a purpose in the optimisation of building performance when incorporated into a comprehensive performance management platform [28], by supplementing the existing hard data sources in the performance assessment framework.

While a performance framework aimed at optimising building performance can certainly benefit from enhanced building data access, a building should engage with all building stakeholders and not just the building manager. There is significant scope to use qualitative, soft data sources to inform building users as to the impact of their preferences on building performance and to persuade them to modify behaviour accordingly. Ultimately, the purpose of most buildings is to provide a comfortable and safe environment for occupants to live and work. By enabling building occupants to engage with the building and understand the impacts of their actions on building performance, it is possible to engender a sense of involvement with the building community.

Some of the hard data sources currently used in the building performance optimisation space are illustrated in Fig. 2, together with some of the possible softer data sources. These additional sources could complement existing hard

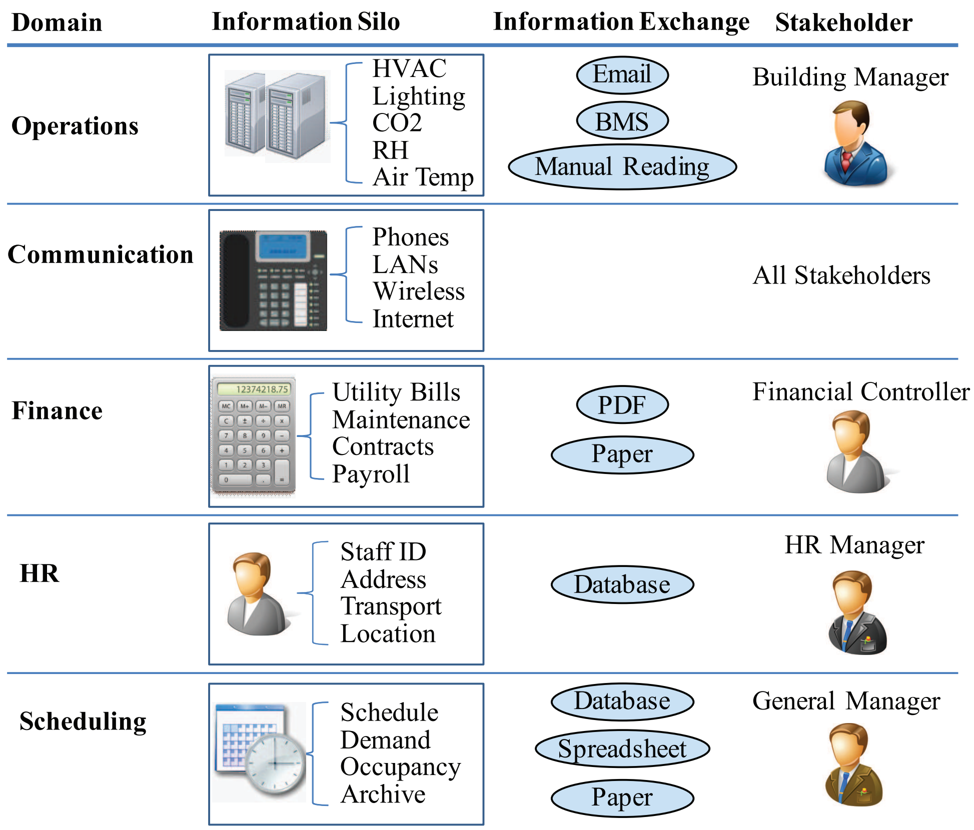


Figure 3: Some of the disconnected data silos across AEC domains resulting in incomplete representations of building performance.

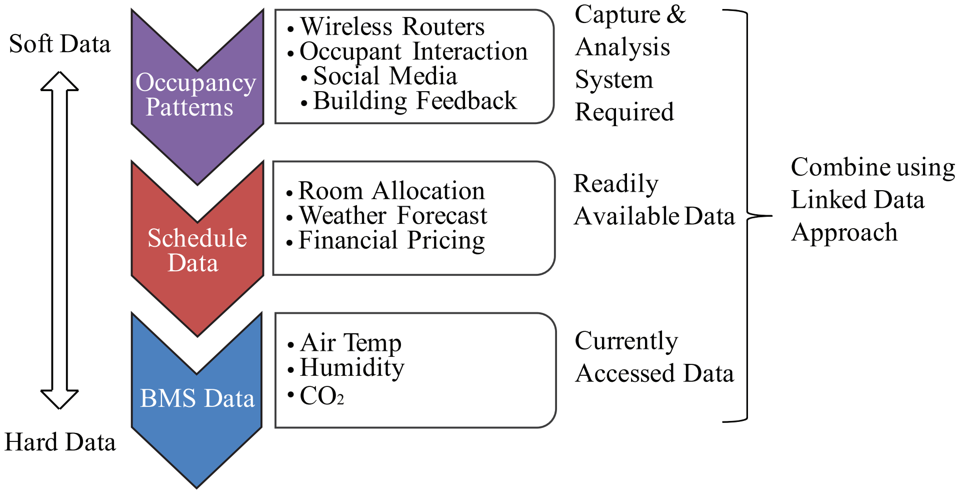


Figure 4: Continuum of hard and soft data sources relating to occupancy and scheduling currently not integrated in any meaningful way during building operation.

data sources by providing a further layer of data, for example in the area of fault detection. There is a wide spectrum of data sources available throughout a building, even when just considering the narrow area of scheduling and occupancy patterns. Some of these sources are readily accessible and exist in a format that lends itself to analysis, whilst others require a greater degree of assessment and interpretation before they can be used to drive performance optimisation efforts.

The paper explores how two of these data sources could be integrated with existing data sources using semantic web technologies. The authors have developed a number of software tools aimed at displaying building data in an informative and structured manner. These solutions are tailored to suit the needs of the end user or building stakeholder and in the case of [9], are aimed at motivating the building occupant to pursue specific energy saving measures.

3. Demonstrators

In the remainder of this paper, two demonstrators are documented to show how cross-domain data could be integrated with existing data sources using semantic web technologies. These demonstrators illustrate the concept and work is on-going on the technical implementation of these data exchanges. The demonstrators 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 diverse data sources can be accessed and transformed using semantic web technologies. The demonstrators illustrate how data from random sources can be easily transformed into RDF and integrated with other data.

3.1. Demonstrator motivation

In the first demonstrator, the paper explores how scheduling data might be integrated with building operation data to illustrate how such data can be used in a cross-domain manner. This experiment is not intended to predict actual savings from the integration of cross-domain occupancy data, but is designed to show how data from non-connected domains can be integrated to allow a greater degree of understanding of building requirements.

In a typical university or other large scale campus, the scheduling software built into the individual building's BMS is manually populated. In many cases, the system is configured to operate during office hours, when the facility is occupied, taking account of holidays, etc. During the design phase of the project life-cycle, expected occupancy patterns are taken into account when deciding on the optimum schedule. Often, little attention is actually paid to occupancy patterns during the operational phase of a facility, leading to uncomfortable and over-conditioned situations in the building. Controlling HVAC systems using occupancy data is a recognised means of optimising performance [29]. At the same time, room occupancy numbers are often scheduled by a different function in the university, the admissions office. The schedule and occupancy pattern changes from year to year, but this is not reflected in the BMS settings. Essentially, the activities of one domain can have knock on effects on other domains in the building.

The second demonstrator focuses on soft data related to building use which is difficult to quantify and integrate with existing operational structures. The idea behind this demonstrator is to generate a sense of ownership and ambient awareness amongst a group of building occupants and to encourage them to post tweets describing some of their interactions with the building. It is felt that this type of feedback would provide building managers with instant feedback on building issues as they arise and could also serve as a type of barometer for occupant satisfaction. Again, this is not a typical source of data for building managers. In this demonstration the paper illustrates how this type of data can be captured and transformed using

Figure 5: New Engineering Building (NEB), NUI Galway, Ireland.

semantic web technologies.

The outcome from both demonstrators is a set of building-related data exposed in RDF graphs, which can then be easily accessed and queried using semantic web technologies. In the concluding section, the paper describes how semantic web technologies forms the basis of a performance management tool used to integrate these data sets in a cross-domain manner.

3.2. Demonstration Building at the National University of Ireland, Galway campus

The building used to carry out the demonstrations is the 14,000 m^2 new engineering building (NEB) on the National University of Ireland, Galway campus (Fig. 4). This is an ideal demonstrator as it is a heavily instrumented building and utilises a complex mixed-mode heating and cooling system together with an innovative climate facade used to provide extensive natural ventilation.

The NEB is particularly interesting given that 90 percent of the building's occupants are students who attend lectures and engage in practical coursework in the building. They generally do not see themselves as stakeholders in the building and are often not aware of the controls available to them in the building or how the building management function operates. The building is managed remotely based entirely on hard data emanating from the BMS. The onsite building manager on the other hand deals almost entirely with soft data feedback in the form of queries from the building occupants.

4. Demonstrator 1: Integrating scheduling data with building operating strategy

4.1. Available data

The university admissions office uses timetabling software (MS Excel) to administer the use of university lecturing facilities. This centralised room booking/scheduling service operates separately from the BMS, a Cylon Unitron System [30]. The room booking schedule changes from year to year and as a result, spaces are conditioned when no occupants are present, whilst others are not conditioned, despite students being present. This type of scheduling mismatch is replicated in many buildings.

Some studies show that occupancy can be used as an indicator to schedule demand-led air conditioning systems, together with the traditional air temperature, external air temperature and relative humidity readings [31, 32], whilst others suggest methods of interpreting occupant satisfaction with indoor ambient temperatures [33]. Buildings are generally conditioned to satisfy maximum occupancy, but this level often does not describe occupancy patterns. Existing systems used in other domains that provide ordinal data can provide a basis for performance analysis [34] and, when considered with other hard data sources, can provide further qualifying data about performance. Sources of this sort of data include facility scheduling software, infrared sensors, swipe card systems, wireless routers and personal radio frequency identification (RFID) trackers. Other studies have investigated methods of measuring real time occupancy using a variety of technologies, including infrared detectors and door and window opening sensors [35], RFID sensors [36] and Wi-Fi connection hotspots [37].

Many of these technologies are highly complex and rely on complex algorithms to determine the occupancy level of a space. Furthermore, these methods do not overcome the interoperability issues associated with cross-domain data analysis. The paper illustrates how semantic web technologies can be used to expose occupancy scheduling data from a completely separate, autonomous building domain and deliver it to other interested parties in the facility. Although questions exist over the usefulness of static occupancy schedules to drive HVAC scheduling, this type of softer data can serve as an indicator of building use and, when viewed in conjunction with other traditional hard data sources, can serve an important function.

Table 1 shows the BMS schedule for the lecture theater G018, indicating the hours when the space is being conditioned. This pattern reflects an effort on behalf of the university to maintain a conditioned space, whilst keeping costs low. This is the type of information currently available to the building manager about this space, as returned by the BMS.

By comparing this schedule to the room booking schedule (Table 2), those moments in the week can be found when a fully occupied room is conditioned and when an empty room is not conditioned.

4.2. Combining the data sets

Using semantic web technologies, it is possible to explicitly link semantic representations of building objects, such as rooms, while they are retained in various different data silos. In Fig. 6, the room concept is used by four

Time	Mon	Tue	Wed	Thu	Fri
08:00-09:00	Off	Off	Off	Off	Off
09:00-10:00	On	On	On	On	On
10:00-11:00	On	On	On	On	On
11:00-12:00	Off	Off	Off	Off	Off
12:00-13:00	Off	Off	Off	Off	Off
13:00-14:00	Off	Off	Off	Off	Off
14:00-15:00	Off	Off	Off	Off	Off
15:00-16:00	On	On	On	On	On
16:00-17:00	On	On	On	On	On
17:00-18:00	Off	Off	Off	Off	Off

Table 1: BMS schedule of operation for lecture theatre G018. Cells coloured grey represent times when the space is conditioned.

different data models, each model representing a different context. Firstly, the *BMS* uses the concept of the room to represent the location of sensors and HVAC services. *Human resource management (HRM) software* uses the room concept to define where a staff member is based. The *BIM modeling environment* uses the room concept to define a geometric space with respect to the remainder of the building, while the *campus scheduling software* uses the room concept to define where an event, in this case a lecture, takes place with a given number of participants.

By exposing these four diverse data streams in RDF and linking them together as in Fig. 5, ways of analysing this data with a view to greater operational efficiency in the space, based on optimising the BMS schedule can be explored. Taking the BMS scheduling data, a rudimentary calendar using Google calendar and exported to the iCal file format (Listing 1). The iCal file format was used as a means to capture calendar data as it is a schema which can be easily transformed to RDF, using an existing conversion service. One of the key pillars of the semantic web initiative is the reuse of existing ontologies to describe data.

Listing 1: BMS Schedule in iCal format.

```
UID:qhfaru4esobl8ts8mm7qi0jgl8@google.com
CREATED:20130531T221206Z
DESCRIPTION:
LAST-MODIFIED:20130718T131250Z
```

Time	Mon	Tue	Wed	Thu	Fri
08:00-09:00					
09:00-10:00	237		237	200	237
10:00-11:00		237	237	237	200
11:00-12:00	237	180	180	145	237
12:00-13:00	237	200	237	200	149
13:00-14:00			145		
14:00-15:00	221	237	145		140
15:00-16:00	221		120	160	140
16:00-17:00	149		250	160	
17:00-18:00	200			160	

Table 2: BMS schedule overlaid with occupancy pattern. The blue background indicated when the room is conditioned and the numbers relate to the amount of students scheduled to be in the room at that time.

```
LOCATION:G017
SEQUENCE:0
STATUS:CONFIRMED
SUMMARY:OFF
TRANSP:OPAQUE
END:VEVENT
BEGIN:VEVENT
DTSTART:20130606T070000Z
DTEND:20130606T080000Z
DTSTAMP:20130718T132918Z
```

The web-based iCaltoRDF converter [38] is used to convert this output to RDF, using the RDF calendar ontology [39] (Listing 2). This system uses the RDF Calendar [39] to integrate calendar data with other semantic web data.

Listing 2: BMS schedule in iCal format converted to RDF data.

```
<component>
    <Vevent>
        <dtstart rdf:parseType='Resource'>
            <dateTime>2013-06-07T07:00:00Z</dateTime>
        </dtstart>
        <dtend rdf:parseType='Resource'>
            <dateTime>2013-06-07T08:00:00Z</dateTime>
```

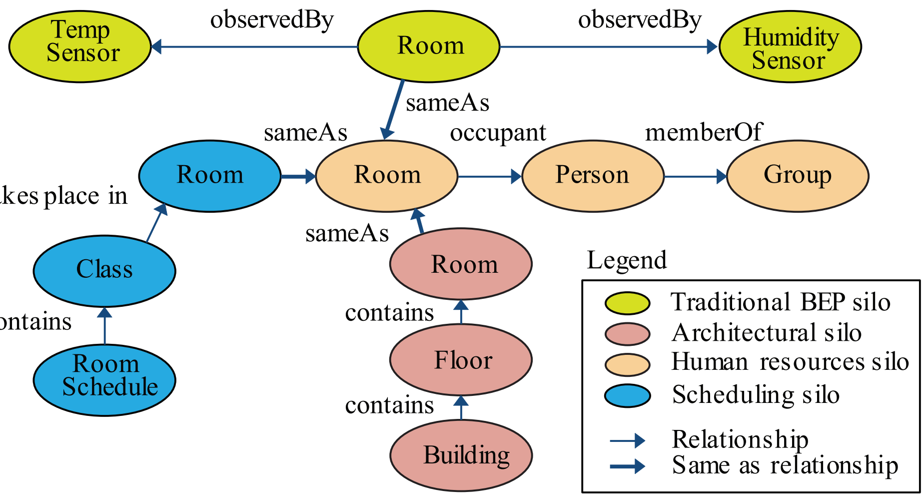


Figure 6: Diagram illustrating the relationship between the BMS, the room booking system (MS Excel), BIM and human resource management (HRM) systems, linked using the Room entity.

```
        </dtend>
        <dtstamp rdf:parseType='Resource'>
            <dateTime>2013-07-18T13:29:18Z</dateTime>
        </dtstamp>
        <uid>qhfaru4esobl8ts8mm7qi0jgl8@google.com</uid>
        <created>2013-05-31T22:12:06Z</created>
        <description></description>
        <lastModified rdf:parseType='Resource'>
        <dateTime>2013-07-18T13:12:50Z</dateTime>
        </lastModified>
        <location>G018</location>
        <sequence>0</sequence>
        <status>CONFIRMED</status>
        <summary>OFF</summary>
        <transp>OPAQUE</transp>
    </Vevent>
</component>
```

A similar process is used to convert the room occupancy schedule to RDF. The key idea here is that further information is gathered about the component relating to each time slot. In this case, the time slot relating to Mondays, from 7 to 8 AM, can include a summary reference of `off`, but also a summary reference of `237`. In this manner, two separate schedules can be integrated. For our purposes, the resulting data set can be used by semantic web technologies to illustrate the occasions when the space is being conditioned, although no occupants are present. Armed with this information, the manager can review the BMS schedule and perhaps decide to modify it. Using a performance metric [13] to describe this objective, the building manager can be provided with quantifiable data on the efficiency of the BMS schedule.

Expanding the range of data sources available and transforming these sources into specific metrics gives the building manager greater awareness of what is happening throughout the building. In this case, the lecture theater is conditioned for 20 hours a week. By incorporating occupancy schedules into this analysis, it can be seen that the room is being conditioned for 5 hours when no lectures are scheduled. Furthermore, the room is not conditioned at all when the bulk of the students are present, during the middle of the day.

Of course this is a simplified example and these correlations should not be looked at in isolation but rather should be used as part of the entire solution, incorporating simulation outcomes, temperature and CO_2 profiles and soft data, including emails, twitter feedback, etc., to optimise performance on a continuous basis.

Time	Mon	Tue	Wed	Thu	Fri
08:00-09:00					
09:00-10:00	237		237	200	237
10:00-11:00		237	237	237	200
11:00-12:00	237	180	180	145	237
12:00-13:00	237	200	237	200	149
13:00-14:00			145		
14:00-15:00	221	237	145		140
15:00-16:00	221		120	160	140
16:00-17:00	149		250	160	
17:00-18:00	200			160	

Table 3: A modified BMS schedule, still operating for 20 hours. Cells coloured grey represent times when the space is conditioned.

Table 3 shows an example of a modified schedule that may be implemented, based on a variety of other factors.

4.3. Discussion of Results

Using the suggested approach, the BMS schedule can be considered in conjunction with other relevant data sources. Research is on-going at present to expose various types of data in RDF. A similar type of analysis can be performed using other data sources, including financial pricing for utilities [9] and comparison of operating conditions with weather data. When such data is available and incorporated with existing BMS data, various possibilities for the optimisation of building performance emerge. These possibilities fall into a number of categories:

1. Optimisation of building performance
   (a) Minimal use of energy whilst meeting stakeholder requirements
   (b) Meeting stakeholder requirements at reduced cost
2. Understanding stakeholder requirements
   (a) Base decisions on actual operation rather than design stage requirements
   (b) Use stakeholder information to optimise stakeholder satisfaction

4.4. Further Work in this Area

Capturing occupancy patterns in buildings is quite a difficult undertaking. In the case of a university building, some indication of occupancy might be gathered from the room booking service. Another data source that might additionally be used, is provided by the wireless network. Students can remotely access course information through this network using a wireless enabled device. An analysis of wireless router patterns throughout the week would also be informative when trying to gauge the true occupancy of the space.

Neither approach provides a complete solution to the issue. The room booking service does not take into account absenteeism amongst students or cancelled lectures, whilst the mobile phone analysis requires each student to have a wireless enabled phone in class.

Using semantic web technologies, it is possible to gather this type of information for the rooms in the building. This type of data is delving more into the realm of soft data and with that it becomes more difficult to infer useful information from it. For instance, in this case, students are not required to log into the wireless network and it is feasible that a room could be full, without anybody accessing the wireless network. Looking at a chart illustrating usage patterns of the wireless network will not be particularly useful for the building manager in terms of an occupancy analysis, but it may serve as a pointer when used in conjunction with other data sources, such as the room booking and BMS schedules.

5. Demonstrator 2: Determining Occupant Comfort Levels

5.1. Assessing occupant comfort

The second demonstrator identifies a range of data sources which may be generated around the area of occupant comfort. These data sources tend to be more qualitative in nature and in some cases may be difficult to derive meaning from. The purpose of this demonstrator is to outline how these sources might be captured and interpreted using semantic web tools. The study was based on the area of occupant comfort, particularly thermal comfort. This work consisted of a Twitter survey, a measurement-based predicted mean vote (PMV) [40] study, a survey-based PMV study, and a simulation-based PMV study.

With the advent of social media, a new range of data sources have now emerged, providing softer, but no less useful information in the form of chatter and instant feedback. These information sources represent an opportunity to engender a sense of connection amongst all stakeholders in a building community. It is now possible to open dialogue with building stakeholders and these dialogues can be focused to encourage feedback, on a range of topics, not least being building operation.

Furthermore, dialogue can be instigated outside the traditional formal channels of information transfer of building operation where information is restricted to a hierarchical gatekeeper approach, where all information is diverted to a centrally placed manager who interprets or filters this information. The paper proposes a range of scenarios which outline the relevance of social media to stakeholder dialogue and demonstrate how these scenarios might be realised by linking the social media information silo with existing building information silos.

5.2. Available data

An aspect of building performance that is studied in the second demonstrator is that of stakeholder satisfaction [41, 42]. More precisely, an experiment was carried out using the Twitter micro-blogging site. Using the NEB as a test bed (Fig. 4), a group of 65 final year engineering undergraduates were encouraged to follow a particular Twitter account (`CE454`) and to post commentary on building performance as they encountered it, throughout the day. This work differs from other studies [43, 44] in this area by the manner in which the data is extracted from the social media domain and exposed in RDF. The key point of this work is to make information more accessible using semantic web technologies.

Based on an initial survey of the group, 35 percent declared that they used social media more than 8 hours per week, with Facebook (89 percent), YouTube (78 percent) and Twitter (78 percent) being the dominant sites accessed. Although almost half the respondents to the survey declared that they never or rarely accessed social media sites during class time, the remainder of respondents accessed such sites throughout the college day. It is important to note here that the group of students surveyed take an Energy Systems course and should thus not only be more keen to use information technology, including social media, but they should also be more aware of the energy systems surrounding them in a building.

The students were asked to comment specifically on a number of zones within the building and these were each given a specified `#` name. The zones included a large lecture hall (`#NEBG017`), two computer suites (`#NEBCompG` and `#NEBComp1`) and the restaurant area (`#NEBCanteen`). The students were asked to reply using the following format: `@CE454 #Location, PMV, comment`. In this way, related tweets could be identified easily on Twitter. The students tended to spend a lot of time in these spaces and they were encouraged to comment on the thermal comfort conditions in the spaces, based on the PMV thermal comfort scale [40], ranging from +3 to -3 as shown in Table 4. It is important to add here that the computer suites (`#NEBCompG` and `#NEBComp1`) tend to be considerably warmer than the other rooms.

PMV value	Thermal Comfort
+3	Too Hot
+2	Warm
+1	Slightly Warm
0	Neutral
-1	Slightly Cool
-2	Cool
-3	Cold

Table 4: PMV thermal comfort scale.

Figure 7: Twitter response relating to the main lecture theater.

By encouraging building occupants to tweet about the comfort levels in the building and comment on general building issues, a Twitter feed can be created for the building (example in Fig. 7). These tweets can also be structured in a particular format which lends itself to analysis.

Students were also asked to comment generally on the building and in this case, the `#NEBGen` tag was used. It was unclear what type of feedback would emerge from this channel and whether it could be a useful flow of information about unknown issues encountered by building occupants.

5.3. Findings of Social Media Experiment

Although most students in the group signed up to Twitter and followed the research account, there was little activity on the account regarding spaces where the thermal conditions were neutral, or classed as 0 on the PMV scale. The twitter handle `CE454` was used to post 26 tweets in total. The twitter response to the main lecture theater, `#NEBG017`, was quite limited, with perhaps 3 tweets in total, and consistently placed the occupant satisfaction level at 0. This corresponded strongly with actual thermal comfort measurements in the space, suggesting a PMV reading between -0.8 and 0, throughout the day.

In contrast to this, the computer suite 1, `#NEBComp1`, generated much more comment on Twitter, around 10 tweets (Fig. 7). Many of the respondents felt the temperature in the space was too hot. This correlated strongly with the thermal comfort analysis of the space, which tended toward a PMV of +3 (too hot).

Figure 8: Twitter results for `#NEBComp1`, indicating an issue with the thermal comfort levels in the space.

When students were asked specifically about the thermal conditions in the computer room, some evidence of ambient awareness was evident, where a user could see a relevant response and respond to that also (Fig. 9).

52 responses were received in total, over a period of three weeks. Users seemed to respond only when something was making them uncomfortable.

Figure 9: Twitter feedback on uncomfortable computer room.

For example, 'loud mechanical' and excessive 'wind' noises were reported, together with high temperatures in the computer suites. People were less motivated to respond when conditions were satisfactory.

Some of the responses were quite interesting from a building management perspective. For instance, the building has a main fresh water supply that is used to service a number of water dispensers located throughout the building. This system was inoperative recently and this featured in a couple of tweets. Similarly, unusual noises were reported in a tweet, including excessive wind noise and loud mechanical sounds. When these issues were discussed with the building manager, he described an on-going issue with the fountain system in the building and an air handling using (AHU) problem with the computer suite. The Twitter experiment is on-going and is being used to ascertain occupant satisfaction levels on a range of issues throughout the building.

5.4. Combining the data sets

Having identified the Twitter data source, this information could be exposed semantically. The Online Presence Ontology [45, 46] can be used to describe a twitter message as an RDF statement (Listing 3). This statement can then be interpreted using semantic webs tools. The aim of Modeling Online Presence is to enable the integration and exchange of online presence related data and utilises a Semantic Web ontology (OPO) to represent data about Online Presence in RDF. This ontology describes data generated using various online messaging and blogging services and how it might be published in RDF. Again, the goal of the semantic web initiative is to utilise existing ontologies to expose data using RDF.

Listing 3: RDF representation of a Twitter message sent by the `CE454` account, based on examples created by the Online Presence Ontology working group.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:opo="http://online-presence.net/opo/ns#"
   xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:rdf="
   http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:
   sioc="http://rdfs.org/sioc/ns#">
  <sioc:UserAccount rdf:about="http://online-presence.net
     /opo/examples#CE454">
    <foaf:accountServiceHomepage rdf:resource="http://www
       .twitter.com/" />
    <foaf:accountName>CE454</foaf:accountName>
  </sioc:UserAccount>
```

```
  <opo:OnlinePresence rdf:about="http://online-presence.
     net/opo/examples#CE454Presence">
    <opo:customMessage>
      <sioc:Post rdf:about="http://online-presence.net/
         opo/examples#CE454Status">
        <sioc:content>@Ce454 #NEBGen What are conditions
           like in the NEB today? Computer rooms seem to
           be an issue? Do people miss the water
           fountains?</sioc:content>
      </sioc:Post>
    </opo:customMessage>
    <opo:startTime>2013-01-25T09:50:11</opo:startTime>
    <opo:declaredOn rdf:resource="http://online-presence.
       net/opo/examples#CE454TwitterAccount" />
  </opo:OnlinePresence>
</rdf:RDF>
```

5.5. Discussion of Results

There are a number of findings from this experiment. First of all, it is not clear that *Twitter or micro-blogging in general can be used to accurately survey the population of a large building*. Taking the engineering building as an example, it is inhabited by a large group of technically capable people, with access to a free building-wide wireless network. The group of students surveyed are positively disposed to the question of building operation as they take an Energy Systems course. Despite this, the participation levels of the group were low. Perhaps the main observation was that people were more motivated to respond when directly affected by a specific issue.

Second, the *experiment showed some potential in the area of fault detection*, or issues in the building that may not be obvious to the building manager. The feed returned some unexpected responses, including feedback on noise levels throughout the building and the quality of the fresh water. As an information source, however, the Twitter feed can only be analysed to a limited level. No matter how many predefined hash-tag names (#) are used, the information will always be qualitative in nature rather than quantitative.

A third conclusion that can be made, is that *micro-blogging occupants could easily become a type of mobile sensor*, identifying issues with building performance and posting those issues in a visible way to the wider building community, focusing the attention of the building manager on the issue. The authors feel that this is the area in which Twitter might be most useful, the identification and publication of issues as they arise.

Lastly, it can be concluded that *semantic web ontologies exist which allow the interpretation of micro-blog posts semantically.* These ontologies can then be used by the appropriate semantic web technologies to form an improved and integrated perspective on available building data.

6. Conclusion

In conclusion, identifying and accessing other data sources is a very relevant step in trying to optimise building performance. It has been illustrated, using just two examples of building-related data, how cross-domain scheduling data can be captured and used and also, how micro-blogging sites such as Twitter could be used to identify occupant issues with building performance.

When integrated into a wider building management framework, these extra data sources are particularly useful. Developing this level of integration has proved to be a significant challenge, particularly when integrating cross-domain data. The paper has illustrated the benefits of using semantic web technologies to resolve some of these interoperability issues. This work is ongoing and focuses on converting remote data silos to RDF and developing a performance framework platform capable of capturing and interpreting these streams of data. This work requires a performance framework ontology to describe this process and will be presented in a further paper.

Ultimately, not all building-related data sources will be of equal use and developing interoperability between some of the more qualitative sources is of limited value. By the same token, data sources which can give a clear indication of real-time building occupancy patterns are very worthwhile and there are a host of such sources throughout modern buildings. The authors suggest that quantitative data that exists in separate AEC domains lends itself more easily to analysis and there are clear benefits to exposing these data sources to the building management framework. There are over 200 million buildings in the EU and as enabling technology develops, it is clear that vast quantities and types of softer data will emerge from modern buildings, in the areas of communication systems, automated control systems, financial, human resources, etc.

A robust methodology needs to be in place to capitalise on this data and drive operational efficiency. The authors feel that a comprehensive performance measurement platform is required that takes data from traditional hard building sources, together with softer data sources.

7. Acknowledgments

This work has been funded by the Irish Research Council. This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

References

[1] D. Crawley, J. Hand, M. Kummert, B. Griffith, Contrasting the capabilities of building energy performance simulation programs, Build Environ 43 (2008) 661–73.

[2] R. J. Hitchcock, M. A. Piette, S. E. Selkowitz, Performance metrics and life-cycle information management for building performance assurance., in: 1998 ACEEE Summer Study on Energy Efficiency in Buildings, 1998.

[3] M. Gallagher, A. O'Connor, J. Dettbar, L. Gilday, Cost analysis of inadequate interoperability in the U.S. capital facilities industry, Technical Report NIST Report GCR 04-867, National Institute of Standards and Technology (NIST), 2004.

[4] M. Liu, D. Claridge, W. Turner, Continuous commissioning of building energy systems, J Sol Energy Eng 125 (2003) 275–81.

[5] M. A. Piette, S. T. Khalsa, P. Haves, Analysis of an information monitoring and diagnostic system to improve building operations, Energy and Buildings 33 (2001) 783–91.

[6] N. C., D. Jacob, Guidelines for the evaluation of building performance, Fraunhofer Institute for Solar Energy Systems, Freiburg, DE, 2008.

[7] Green Buildings Discussion Group, Implications of A Data-Driven Built Environment, 2012. Available online: http://www.nyas.org/DataBuiltConf. Last accessed on 25 July 2013.

[8] W. Shen, Q. Hao, H. Mak, J. Neelamkavil, H. Xie, J. Dickinson, R. Thomas, A. Pardasani, H. Xue, Systems integration and collaboration in architecture, engineering, construction, and facilities management: A review, Advanced Engineering Informatics 24 (2010) 196–207.

[9] E. Curry, 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, Advanced Engineering Informatics 27 (2013) 206–19.

[10] J. O'Donnell, E. Corry, S. Hasan, M. Keane, E. Curry, Building performance optimization using cross-domain scenario modeling, linked data, and complex event processing, Build Environ 62 (2013) 102–11.

[11] P. Pauwels, D. Van Deursen, J. De Roo, T. Van Ackere, R. De Meyer, R. Van de Walle, J. Van Campenhout, Threedimensional information exchange over the semantic web for the domain of architecture, engineering and construction, Artificial Intelligence for Engineering, Design and Manufacturing 25 (2011) 317–32.

[12] C. Eastman, P. Teicholz, R. Sacks, K. Liston, BIM handbook: a guide to building information modeling for owners, managers, designers, engineers and contractors, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008.

[13] K. Gillespie, P. Haves, R. Hitchcock, J. Deringer, K. Kinney, A Specifications Guide for Performance Monitoring Systems, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 2007.

[14] P. Teicholz, Aecbytes: Labor productivity declines in the construction industry: Causes and remedies, 2004. Available online: http://www.aecbytes.com/viewpoint/2004/issue_4.html. Last accessed on 25 July 2013.

[15] BuildingSMART, Industry foundation classes (ifc) data model - buildingsmart, 2013. Available online: http://www.buildingsmart.org/standards/ifc. Last accessed on 25 July 2013.

[16] BuildingSMART, National bim standard - united states, 2013. Available online: http://www.nationalbimstandard.org/. Last accessed on 25 July 2013.

[17] E. W. East, Construction operations building information exchange (cobie) — whole building design guide, 2013. Available online: http://www.wbdg.org/resources/cobie.php. Last accessed on 25 July 2013.

[18] BLIS Project, Building lifecycle interoperable software - project brief, 2002. Available online: http://www.blis-project.org/index2.html. Last accessed on 25 July 2013.

[19] R. H. Katz, D. A. Culler, S. Sanders, S. Alspaugh, Y. Chen, S. Dawson-Haggerty, P. Dutta, M. He, X. Jiang, L. Keys, A. Krioukov, K. Lutz, J. Ortiz, P. Mohan, E. Reutzel, J. Taneja, J. Hsu, S. Shankar, An information-centric energy infrastructure: The berkeley view, Sustainable Computing: Informatics and Systems 1 (2011) 7–22.

[20] G. Augenbroe, An overview of the combine project, in: Proceedings of the 1st European Conference on Product and Process Modelling in the building industry, 1994, pp. 547–55.

[21] C.-S. Park, G. Augenbroe, Benchmarking of a building performance assessment toolkit, in: Architectural Engineering 2003: Building Integration Solutions, 2003, pp. 1–4.

[22] IFC Solutions Factory, The model view definition site, 2010. Available online: http://www.blis-project.org/IAI-MVD/. Last accessed on 25 July 2013.

[23] UK Cabinet Office, Government construction strategy, 2011. Available online: https://www.gov.uk/government/publications/government-construction-strategy. Last accessed on 25 July 2013.

[24] GSA, General services administration 3d-4d building information modeling, 2013. Available online: www.gsa.gov/bim. Last accessed on 17 Feb 2014.

[25] F. Manola, E. Miller, RDF Primer, W3C Recommendation, 2004. Available online: http://www.w3.org/TR/rdf-primer/. Last accessed on 25 July 2013.

[26] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific American 284 (2001) 34–43.

[27] P. Pauwels, D. Van Deursen, R. Verstraeten, J. De Roo, R. De Meyer, R. Van de Walle, J. Van Campenhout, A semantic rule checking environment for building performance checking, Automation in Construction 20 (2011) 506–18.

[28] J. 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, Cork, IE, 2009.

[29] V. L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A. E. Cerpa, M. D. Sohn, S. Narayanan, Energy efficient building environment control strategies using real-time occupancy measurements, in: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, 2009, pp. 19–24.

[30] Cylon, Building - energy management - systems - bms, 2013. Available online: http://www.cylon.com/. Last accessed on 25 July 2013.

[31] J. Kwak, P. Varakantham, R. Maheswaran, Y. Chang, M. Tambe, B. Becerik-Gerber, W. Wood, Tesla: An energy-saving agent that leverages schedule flexibility., in: Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems. International Foundation for Autonomous Agents and Multiagent Systems, 2013., 2013.

[32] J. Kwak, P. Varakantham, R. Maheswaran, M. Tambe, F. Jazizadeh, G. Kavulya, L. Klein, B. Becerik-Gerber, T. Hayes, W. Wood, Saves: A sustainable multiagent application to conserve building energy considering occupants., in: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[33] F. Jazizadeh, A. Ghahramani, B. Becerik-Gerber, T. Kichkaylo, M. Orosz, User-led decentralized thermal comfort driven hvac operations for improved efficiency in office buildings, Energy and Buildings 70 (2014) 398 – 410.

[34] ITOBO, Itobo: Information and communication technology for sustainable and optimised building operation, 2008. Available online: http://zuse.ucc.ie/itobo/. Last accessed on 25 July 2013.

[35] P. Daz, T. Olivares, R. Galindo, A. Ortiz, F. Royo, T. Clemente, The ecosense project: An intelligent energy management system with a wireless sensor and actor network, in: R. Howlett, L. Jain, S. Lee (Eds.), Sustainability in Energy and Buildings, Springer, Berlin Heidelberg, DE, 2011, pp. 19–24.

[36] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an rfid based system for demand-driven hvac operations, Automation in Construction 24 (2012) 89–99.

[37] C. Martani, D. Lee, P. Robinson, R. Britter, R. C., Enernet: Studying the dynamic relationship between building occupancy and energy consumption, Energy and Buildings 47 (2012) 584–91.

[38] D. Connolly, ical2rdf – icalendar to rdf convertor, 2003. Available online: http://www.kanzaki.com/courier/ical2rdf. Last accessed on 25 July 2013.

[39] D. Connolly, L. Miller, Rdf calendar - an application of the resource description framework to icalendar data, W3C Interest Group Note, 2005. Available online: http://www.w3.org/TR/rdfcal/. Last accessed on 25 July 2013.

[40] P. Fanger, Thermal Comfort, Danish Technical Press, Copenhagen, DK, 1970.

[41] D. Pati, C.-S. Park, G. Augenbroe, Roles of building performance assessment in stakeholder dialogue in aec, Automation in Construction 15 (2006) 415–27.

[42] D. Pati, C.-S. Park, G. Augenbroe, Roles of quantified expressions of building performance assessment in facility procurement and management, Building and Environment 44 (2009) 773–84.

[43] D. Lehrer, J. Vasudev, Evaluating a social media application for sustainability in the workplace., in: CHI'11 Extended Abstracts on Human Factors in Computing Systems, 2011, 2011.

[44] J. Mankoff, D. Matthews, S. Fussell, M. Johnson, Leveraging social networks to motivate individuals to reduce their ecological footprints, in: Proc. of the Hawaii International Conference of System Sciences (HICSS-40), 2007, 2007.

[45] M. Stankovic, J. Jovanovic, Online presence in social networks, in: Proceedings of W3C Workshop on the Future of Social Networks, 2009.

[46] M. Stankovic, J. Jovanovic, A. Passant, N. Milikic, F. Radulovic, The online presence project, 2013. Available online: http://online-presence.net/ontology.php. Last accessed on 25 July 2013.