Advanced Engineering Informatics 28 (2014) 370–380

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

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, United States
[d] Digital Enterprise Research Institute, National University of Ireland Galway, IDA Business Park, Lower Dangan, Galway, Ireland
[e] Department of Architecture and Urban Planning, Ghent University, Ghent, Belgium

ARTICLE INFO

Article history:
Received 13 August 2013
Received in revised form 5 May 2014
Accepted 21 May 2014
Available online 16 June 2014

Keywords:
Social media
Twitter
Linked data
Performance metrics
Building performance
RDF

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 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.

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 [4–6]. Modelling, 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

☆ Handled by W.O. O'Brien.

* Corresponding author. Tel.: +353 656823721.
E-mail addresses: edwardcorry@nuigalway.ie (E. Corry), james.odonnell@ucd.ie (J. O'Donnell), ed.curry@deri.org (E. Curry), Daniel.coakley@nuigalway.ie (D. Coakley), p.pauwels@uva.nl (P. Pauwels), marcus.keane@nuigalway.ie (M. Keane).

http://dx.doi.org/10.1016/j.aei.2014.05.002
1474-0346/

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 modelling (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 [16–20,11,9] 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 modelling and information handover protocols [22]. Industry and national level organisations have recognised the importance of data management and building information modelling (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.

Fig. 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. Fig. 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'.

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