Integration of operational data in building information modelling: From ontology to application

Ardeshir Mahdavi*, and Dawid Wolosiuk

Department of Building Physics and Building Ecology, Karlsplatz 13, Vienna, Austria

Abstract. Building design, construction, and operation has been suggested to benefit from comprehensive and well-formed repository of associated information. Building Information Modeling (BIM) has pursued the supposition that repositories could facilitate seamless exchange of information amongst the multiple stakeholders in the building delivery process. Related research and development efforts have primarily focused on representation of buildings' geometry and specifications of their constituent structural and constructional components. More recently, representations of building environmental control equipment and systems have begun to become incorporated in BIM applications. Ongoing work in this area has likewise targeted development of common schemes to incorporate, in BIM, buildings' sensory networks and elements that serve, for instance, the operation of HVAC (Heating, Ventilation, and Air-Conditioning) systems. To achieve these targets, comprehensive and robust ontologies of building monitoring data and building performance indicators are essential. In the present contribution, we first present a recently introduced original proposal for such ontologies, covering data regarding dynamic data relevant to state, operation, and performance of buildings. This ontology was developed based on an extensive review of building monitoring data and performance indicator catalogues in thermal, air quality, visual, and acoustical domains. The ontology's structural core basically involves a systematic specification of the generic attributes of building performance variables. We then illustrate various benefits and applications of this ontology. It is shown to support data quality check, data visualization, building operation optimization, and preventive building systems maintenance. It can also add to the clarity of building performance requirements specifications, advance the understanding of building performance principles in educational and training settings, and provide and early integration of buildings' operational attributes in BIM applications.

1 Introduction

Recent trends in building design and construction processes display increased efforts in the AEC community (architecture, engineering and construction) to deploy building information modeling (BIM), especially in large-scale projects. This is also reflected in the increasingly prescribed use of BIM in publicly procured construction projects. This is suggested to be beneficial to the stakeholders involved by supporting seamless communication and collaboration toward cost reduction and design errors minimization.

BIM is all about data. Data organization and structure that can be commonly recognized and shared by digital tools is the key to successful information exchange amongst pertinent professionals. In BIM, data organization is enabled (conveyed) by utilization of standardized data models and ontologies. An example of a well-known data model is Industry Foundation Classes (IFC) [1]. It is used for representation of construction (walls, roofs, windows, etc.) and facility management (maintenance details, installation dates, etc.) data. Whereas IFC may represent an instance of a wellestablished, continuously evolving format for primarily static data (e.g., building components), there is a paucity of data models and ontologies for representation of dynamic data (e.g., monitored states). The latter, would include, in the context of built environment, measurement data acquired from various sensor networks as well as simulated data generated via computational building performance assessment tools. A performance assessment of both building designs and existing buildings rely on monitored or computed data concerning buildings' behavior toward derivation of the values of key building performance indicators (BPIs). Both primary performance data (i.e., sensor data or simulated data) and high-level building performance indicator values come in various forms, degrees of resolution, and application domains. Efficient and effective processing of such information could greatly benefit from a well-structured ontology that would cover the multiple levels of complexity involved. Such ontology is essential for scientific community to facilitate data analysis towards building design, operation, and retrofit optimization throughout the buildings' life cycle. It would also provide a solid basis for development of visualization engines that could further support optimization or BMS (Building

^{*} Corresponding author: <u>amahdavi@tuwien.ac.at</u>

Management Systems) applications and provide deeper insight into the data.

Multiple efforts have targeted the development of such ontologies. A previous effort in the building performance assessment domain involved the definition of an ontology based on an extensive review of existing building performance indicators in thermal, visual, acoustical and air quality domains [2, 3]. Another effort proposed an ontology for building monitoring data addressing the diversity and complexity of monitored data streams [4, 5, 6].

The following section of the paper provides an overview of these ontologies. Moreover, it includes a universal schema that captures the common characteristics of building monitoring and performance data.

2 An ontology for dynamic building data

2.1. Ontology for monitored data

To provide a robust classification framework for representation of data from building monitoring systems, a number of basic data categories must be identified. Based on the aforementioned efforts in the area of building monitoring, six data categories where identified. These include: occupants, indoor environmental conditions, external environmental conditions, control systems and devices, equipment, and energy flows. Table 1 provides an overview of five of these categories together with examples of subcategories and monitored variables.

 Table 1. General schema for building performance data (modified based on [6])

Category of monitored data	Subcategory of monitored data	Examples of monitored variables	
	Position	Movement	
Inhabitanta	Control action	Window contact	
innaonants	Attributes	Clo-value	
	Attitudes	Thermal sensation	
	Hygro-thermal	Air temperature	
Indoor	Acoustical	Sound level	
conditions	Air quality	CO2 concentration	
	Visual	Task illuminance	
	Heating/cooling	Thermostat setting	
Control	Ventilation	Air flow rate	
devices	Lighting	Dimming setting	
	Shading	Blind position	
Equipment	Office	Operational state	
	Household	Operational state	
	Heating/cooling	Fuel consumption	
Energy	Ventilation	Power usage	
Energy	Lighting	Power usage	
	Equipment	Power usage	

2.2 Ontology for building performance indicators

A comprehensive review of all building performance indicators would be a difficult task, as they are constantly extended and modified. A recent effort [7] reviewed a large number of performance indicators in following domains: energy efficiency, hygro-thermal performance, thermal comfort, indoor air quality, indoor visual environment, and indoor acoustical environment. The indicators span from ones that capture strictly technical systems performance to others that describe building's "habitability" (i.e., indoor environmental quality) [3, 8].

Figure 1 presents an overview of indicator domains (main categories) and examples of their subsets together with illustrative indicator instances.

2.3 Universal schema

To capture the common characteristics of data streams (feeding in the ontology) generated by both physical (meters, sensors) and virtual (simulation tools, numeric models) data sources, we proposed a comprehensive

 Table 2. General schema for building performance indicators (modified based on [3])

Category						
Sub-category						
	Name					
		Туре				
		Magnitude (size)				
		Direction (vector)				
		Unit				
			Point			
			Plane			
			Volume			
		Spatial domain	Topological ref.			
			Aggregation			
	Value		method			
			Grid size			
X 7 · 11			Time stamp			
Variable		_	Duration			
		Temporal	Time step			
		domain	Aggregation			
			method			
			Range			
		Frequency	Band (filter)			
-		domain	Weighting			
			Aggregation			
			method			
	Agent	ID				
	Notes -	Data sources	Category			
		Dum Sources	ID/name			
		Derivation	Details (formula,			
		method	link, etc.)			



Fig. 1. An overview map of five building performance domains with subcategories and illustrative performance indicator examples [3]

ontological schema (see Table 2). This schema can be shown to fulfill requirements of both monitored data and BPI values. Each variable falls under a specific category and subcategory. Given a specific time and space, each variable can assume a specific value. Each value can have a number of assigned properties and attributes. The variable's type suggests, primarily, if it is quantitative, or qualitative. Quantitative data should be supplemented with the magnitude, in case of vectors also direction and a relevant unit for valid processing and interpretation.

Depending on the category of the variable, a number of additional properties can be specified in three domains. Spatial domain properties allow to associate a variable to a specific point in Cartesian coordination system or to topologically specified location (e.g., room tag). Temporal domain properties can be expressed in the schema via a time stamp (e.g. for a sensor reading). A time step denotes recurrent temporal intervals to which measured or simulation data could be assigned (e.g., hourly heating loads). Duration denotes the overall time frame to which a given variable value corresponds (e.g., annual cooling load). As such time step and grid size can specify the discretization resolution of pertinent temporal and spatial continua. The Frequency domain attributes are relevant to measured or simulated values that display wave characteristics (e.g., light/radiation, sound).

2.4 Application of the proposed schema

To illustrate the working and potential of the ontology, Table 3 includes three exemplary variables from different domains. Thereby, categories, subcategories, and variable attributes are captured.

3 Implementation and testing

3.1 Ontologically consistent data storage

One of the most important aspects regarding the actual application potential of the proposed ontology concerns the selection of a proper data container. Thereby, large sets of semantically enriched data would have to be structured so as to conform to the proposed schema.

We selected HDF (Hierarchical Data Format) file format, specifically its most recent version HDF5 [9] as the container for ontologically structured data. This format is suggested to be suitable for storing and managing large and complex data sets. The following points are quoted from developers' summary of the format "...Advantages of HDF5:

- Versatile data model that can represent very complex, heterogeneous data objects and a wide variety of metadata through an unlimited variety of datatypes
- Ready for high speed raw data acquisition
- Portable and extensible with no limits on file size, allowing applications to evolve in their use of HDF5
- Self-describing, requiring no outside information for applications to interpret the structure and contents of a file...
- Long-term data archiving solution" [10]

Table 3. Illustrative representation of three exemplary p	erformance	variables	(monitored data,	performance	indicators)	following the
structure of the proposed ontology (modified based on [3	3])					

Cate	gory			Energy and resources	Thermal performance	Occupants
Subc	ategory			Energy performance indicator	Environmental indices	Control action
	Name			Heating Load	Air flow velocity	Window contact
		Туре		Quantitative	Quantitative	Quantitative
		Magnitude		50	0.25	0
		Direction		-	[0 0 -1]	-
		Unit		kWh.m ⁻²	m.s ⁻¹	-
			Point	-	1.50, 2.00, 1.10	1.00 , 0.00, 1.20
			Plane	-	-	-
			Volume	Building A	-	-
		Spatial domain	Topological reference	Building A	Room R_1	Room R_1
iable			Aggregation method	-	-	-
Var	Value		Grid size	-	0.20 m	-
		Temporal domain	Time stamp	-	08.05.2018 10:30:00	08.05.2018 09:20:00
			Duration	Annual	-	-
			Time step	1 hour	-	5 min
			Aggregation method	Arithmetic summation	-	-
		Frequency domain	Range	-	-	-
			Band	-	-	-
			Weighting	-	-	-
			Aggregation method	-	-	-
Ager	nt	ID		-	-	Occupant_1
		Data	Category	Simulation	Simulation	Sensor
Note	s	source	ID/name	Sim_20180729_1	Sim_20180823_3	Con_15
110165		Derivation Method			https://www.iso.org/	

3.2 Examination of the robustness of the ontology

We have been testing the robustness of the proposed ontology and implementation in HDF5 file format. Testing involves high resolution measurement data gathered from multiple sensors installed at the office space in a university building (TU Wien) in Vienna, Austria. Table 4 gives an overview of the observed variables selected for testing. The selected dataset includes 116 unique variables observed over a period of 3 years resulting in about 16 millions of single data points.

As an initial step, data had to undergo migration from a multi-table database format to the ontology's treestructure schema. This operation required detailed review of the observed variables concerning missing attributes. A workflow was developed to assign supplementary information to monitored variables and to store the enriched data according to the developed schema in a single HDF5 file. Any application of the proposed ontology, whether it is data quality check, optimization, visualization, or analysis relies on accurate data extraction. The structure of the proposed ontology enables efficient and intuitive data queries to browse, locate, or lookup data of interest.

A series of algorithms were created in the Python [11] environment to test querying efficiency of ontologically structured data. The main focus of the test was to extract target variables that fulfill a specified combination of spatial, temporal, and categorical criteria. After successful extraction, the data of interest was further processed in terms of descriptive statistics and data visualization (e.g., box plots, histograms, line plots).

To illustrate this process, consider the example of a test query to find, extract, and process all available variables from the "Indoor conditions" category. Toward this end, the implemented ontological structure (see Table 2) facilitates the search and extraction process in a highly efficient manner. For instance, spatial (e.g., X[3-5m]; Y[0-2m]; Z[0-3m]) and temporal (e.g., March to June 2016) filters can rapidly narrow down the search space and return – almost instantaneously – the results.

Table 4.	Variables	selected	for	initial	ontology testing	
----------	-----------	----------	-----	---------	------------------	--

Category	Subcategory	Monitored variable		
	Hygro-thermal	Temperature		
	conditions	Relative humidity		
Indoor conditions	Visual	Overhead illuminance		
	Indoor air quality	CO2 concentration		
		Power usage		
Equipment	Electrical	Energy usage		
Control system and devices	Heating/Cooling system	Radiator surface temperature		
0	Position	Presence		
Occupants	Control actions	Window contact		

As the query result are already well-formed and clearly indexed, they can be conveniently subjected to further processing in different applications (e.g., visualization, data mining, trend analysis).

Figure 2 presents, as an example of a basic application scenario, the visualizations (line graph, histogram, box plot) of one of the extracted variables in the aforementioned query, namely the indoor air relative humidity.







Fig. 2. An example of descriptive statistics analysis and visualization of a monitored variable (obtained from a specific indoor air relative humidity sensor) in a specific space over a specific period of time (top: trend; middle: frequency distribution; bottom: box plot)

4 Concluding remarks and future work

In this contribution, we argued for a more in-depth view of BIM application in AEC addressing not only to represent building fabric and facility information, but also to cover dynamic data relevant to buildings' optimal operation. Toward this end, appropriate data models and ontologies for representation of dynamic data (both primary performance data and high-level building performance indicator values) are necessary. Such an ontology can support data analysis towards building design, operation, and retrofit optimization throughout the buildings' life cycle.

Base on a review of previous efforts, the present contribution described an ontology that captures the common characteristics of building monitoring and performance data. To demonstrate the usability and robustness of the ontology, a concrete implementation was targeted, using the HDF5 format for data storage and query. Moreover, the extraction of the stored data toward statistical data analysis and data visualization was demonstrated using the specific instance of an existing building with a comprehensive monitoring infrastructure. This implementation points to a number of critical issues and future challenges. For instance, in case of data obtained from legacy resources, certain data treatment processes and steps may be necessary. This is required to ensure the computability of the structured data with the specifics of the proposed ontology. Nonetheless, the implementation also demonstrated the potential for a fast and seamless data extraction process.

We are currently exploring the potential of further developments in this area to effectively support a number of operations, including automated (or semi-automated) data cleansing, data discretization, derivation of virtual data points based on numeric simulation, as well as deeper application in model generation, data mining, control optimization, and preventive building systems maintenance.

Acknowledgement: The treatment presented in this paper benefited from authors' participation in the IEA EBC Annex 79 and the related discussions.

References

- 1. ISO 16739:2013 Industry Foundation Classes (IFC) for Data Sharing in the Construction and Facility Management Industries (2013)
- 2. A. Mahdavi, J. Bachinger, G. Suter, Building Simulation 2005, *Toward a Unified Information Space for the Specification of Building Performance Simulation Results*, p.671-676 (2005)
- A. Mahdavi, M. Taheri, Proceedings of ECPPM 2018, *A building performance ontology*, p.385-390 (2018)
- A. Mahdavi, S. Glawischnig, M. Schuss, F. Tahmasebi, and A. Heiderer, Proceedings of ECPPM 2016, *Structured Building Monitoring: Ontologies and Platform*, (2016)
- A. Mahdavi, S. Glawischnig, M. Schuss, F. Tahmasebi, M. Taheri, Exploring Occupant Behavior in Buildings, *Structured Building Data Management: Ontologies, Queries, and Platforms*, p.261-286 (2017)
- 6. A. Mahdavi, M Taheri, Journal of Building Performance Simulation, *An ontology for building monitoring*, Vol. **10**, p.499-508 (2017)
- 7. N. Constantinou, Master thesis. Department of Building Physics and Building Ecology, TU Wien, *A comprehensive multi-domain building performance indicator catalogue*, (2017)
- 8. A. Mahdavi, Human Ecology Review, *Steps to general theory of habitability*, Vol.5, No.1, (1998)
- 9. The HDF Group, https://www.hdfgroup.org
- 10. The HDF Group, *Hierarchical Data Format Version 5 (HDF5)*, https://www.hdfgroup.org/wpcontent/uploads/2017/12/HDF512-17.pdf
- 11. Python Software Foundation, Python programming language, https://www.python.org