ECPPM 2021 – eWork and eBusiness in Architecture, Engineering and Construction – Semenov & Scherer (eds)

Application of ontologically streamlined data for building performance analysis

D. Wolosiuk & A. Mahdavi
Department of Building Physics and Building Ecology, TU Wien, Vienna, Austria

ABSTRACT: Buildings are increasingly equipped with data monitoring infrastructures to collect multiple layers of dynamic data pertaining to the states and events related to systems' performance, indoor-environmental conditions, occupants' location, movement, and control-oriented actions representation of dynamic building related data. Efficiently utilized, this data could not only enhance the quality and effectiveness of buildings' operational regimes, but also enrich the knowledge base for building design decision support. However, to exploit the potential of this data effectively, seamless data transformation processes are needed, starting from raw monitoring data and ending in adequately structured and visualized building performance data. In the present contribution, we provide a detailed blueprint of a well-defined and ontologically supported instance of such a transformation process. To illustrate this process, we focus in this paper on a specific class of building performance queries that require information on buildings' visually relevant boundary conditions. Such queries pertain to, for example, the intensity of incident solar radiation of buildings' envelope components, the daylight availability and distribution in interior spaces, and the expected output of building-integrated solar energy harvesting systems.

1 INTRODUCTION

The developments in the last decades have shown that architecture, engineering, and construction (AEC) industry can benefit from well-defined data schemas (such as those offered by International Foundation Classes and green Building XML). Incorporating these schemas into the Building Information Modelling (BIM) software has been suggested to facilitate better communication between parties involved and to improve the overall efficiency of the building design, construction, and operation processes. However, the primary focus of most common schemas lies in the representation of primarily static building attributes, including geometry and semantic information on building components and systems. Thus, there is a potential to further enhance the existing schemas toward a more comprehensive coverage of dynamic processes – particularly those relevant to the operation phase of buildings. Specifically, buildings are increasingly equipped with sophisticated data monitoring infrastructures to collect multiple layers of dynamic data pertaining to the states and events related to systems' performance, indoor-environmental conditions, occupants' location, movement, and control-oriented actions representation of dynamic building related data. Exponentially growing volume and availability of building-related performance data enabled by wireless sensor networks and low-power microcontrollers translate into massive quantities of data on occupancy, indoor-environment, and energy systems' performance. Efficiently utilized, this data could not only enhance the quality and effectiveness of buildings' operational regimes, but also provide an empirically-grounded knowledge base for building design decision support. However, to exploit the potential of this data effectively, seamless data transformation processes are needed, starting from raw monitoring data and ending in adequately structured and visualized building performance data. In the present contribution, we provide a detailed blueprint of a well-defined and ontologically supported instance of such a transformation process. Toward this end, we first describe a recently proposed ontology for inherently dynamic building-related data, including multiple variables expressing the state of building systems, indoor environments, and occupants (position, movement, perceptual clues, behavioral manifestations). Collected raw data from buildings' monitoring infrastructure can be structured and streamlined via projection of such data into the structure of this ontology. Data thus ontologized, can then be accessed from a host of analysis-oriented applications, including data visualization, performance simulation, and optimization. To illustrate the scope and applicability of this ontology and the developed data transformation process, we focus in this paper on a specific class of building performance queries that require information on buildings' visually relevant boundary conditions. Such queries pertain to, for example, the intensity of incident solar radiation of buildings' envelope components, the daylight availability and

DOI 10.1201/9781003191476-15

distribution in interior spaces, and the expected output of building-integrated solar energy harvesting systems. Computational applications for building design and operation support typically deploy predefined or standardized formats for data input necessary for the execution of the internal algorithms. In this context, ontologically guided data processing facilitates the development of interfaces between ontologized data and performance analysis applications depending on standardized input formats. Given the generalized and scalable schema captured in the ontology, the corresponding processed data acquires a highly structured and widely deployable characteristics. This results in increased opportunities to use the same set of ontologized data for multiple purposes and by multiple stakeholders.

2 BACKGROUND

Building related ontologies have been known for quite some time. These became a foundation for primarily "static" data schemes for description of building systems and construction. There is an increasing need for mapping "dynamic" data into building information models, specifically for maintenance and control purposes. This development is related to the growing importance of energy-efficient design and the rising volume of available monitored data.

2.1 *Building ontologies and "dynamic" data*

Recently, there have been a few attempts to address the need for an ontology and data schema for dynamic building related data for performance assessment processes and applications. Instances of such efforts include an ontology for building monitoring (Mahdavi & Taheri 2017), and a building performance indicator ontology (Mahdavi & Taheri 2018, Mahdavi & Wolosiuk 2109a, b). The ontology for building monitoring intended to capture the aspects of data streams coming from various monitoring systems. Based on some of the former efforts in that area (e.g., Mahdavi et al. 2011, 2016) main categories (occupants, indoor environmental conditions, external environmental conditions, control systems and devices, equipment, and energy flows) and pertaining sub-categories were defined based on numerous instances of monitored variables. In addition, a data schema was developed that captures necessary attributes concerning diverse monitored variables.

Building performance indicators (frequently included in standards, technical literature, and simulation applications) play a major role in evidence-based evaluation of building design and operation. There have been very little attempts to capture the characteristics of this domain toward formulating an explicit BPI ontology. The BPI ontology proposed in Mahdavi (2018, 2019a) was based on a review of a large number building performance indicators pertaining to different domains. Five relevant categorical groups were identified, namely energy and resources, indoor air quality, thermal performance, acoustical performance, and visual performance. Moreover, essential data attributes were considered that are needed to comprehensively capture the complexities of various performance indicators. The most recent development combined these two ontologies into a Building Performance Data ontology.

2.2 *Building performance data ontological schema*

Over the course of development of BPI ontology, it became clear that that ultimately most performance indicators are compounds of measured data, or they can be traced back to a single measured variable. This observation indicated that attributes used to describe monitored data should be mostly covered by attributes domain of BPI data. For example, measured indoor temperature or CO2 concentration level can function BPIs, for instance in thermal performance and indoor air quality categories. Compound BPIs can be derived from multiple measured variables. For instance, daylight factor (an indicator in the visual performance domain) is derived by dividing measured or computed indoor illuminance by the simultaneously obtained outdoor illuminance. It was therefore logical to slightly modify the elaborate indicators schema to include monitored data, arriving thus at a comprehensive Building Performance Data (BPD), that is an ontological schema that would encompass both previously mentioned ontologies (Mahdavi & Wolosiuk 2019b).

The general schema for BPD is largely similar to the BPI schema proposed in Mahdavi (2018). Table 1 presents the main features of the BPD schema. Each considered variable falls under a specific main performance category and a sub-category, and has a unique name. Each variable is of a certain type (discrete or continuous), has a magnitude, possibly a direction (in case of vector-type variables), and a unit. Properties included in the spatial domain allow to associate a variable with a position in a Cartesian space or a certain topologically relevant location. Those included in the temporal domain provide necessary (time stamp) and supplementary (duration, time step, or aggregation method) details on variable's temporal features. Properties in the Frequency domain help to specify details relevant for specific performance variables (for example in the acoustics domain).

The Agent ID property is relevant if a variable is attributed to a building user or other relevant agent. The properties included in notes help to provide further details, annotations, and specifications concerning the captured data.

3 FROM RAW DATA TO APPLICATIONS

Structured transformation of raw data (in the context of this work mostly understood as a numerical measure of a performance variable) into an organized ontological schema involves a number of challenges and

Table 1. General BPD schema (modified based on Mahdavi 2018).

<table>
<tr><td colspan="4">Category</td></tr>
<tr><td colspan="4">Sub-category</td></tr>
<tr><td rowspan="24">Variable</td><td colspan="3">Name</td></tr>
<tr><td rowspan="18">Value</td><td colspan="2">Type</td></tr>
<tr><td colspan="2">Magnitude (size)</td></tr>
<tr><td colspan="2">Direction (vector)</td></tr>
<tr><td colspan="2">Unit</td></tr>
<tr><td rowspan="6">Spatialdomain</td><td>Point</td></tr>
<tr><td>Plane</td></tr>
<tr><td>Volume</td></tr>
<tr><td>Topological ref.</td></tr>
<tr><td>Aggregation method</td></tr>
<tr><td>Grid size</td></tr>
<tr><td rowspan="4">Temporaldomain</td><td>Time stamp</td></tr>
<tr><td>Duration</td></tr>
<tr><td>Time step</td></tr>
<tr><td>Aggregation method</td></tr>
<tr><td rowspan="4">Frequencydomain</td><td>Range</td></tr>
<tr><td>Band (filter)</td></tr>
<tr><td>Weighting</td></tr>
<tr><td>Aggregation method</td></tr>
<tr><td>Agent</td><td colspan="2">ID</td></tr>
<tr><td rowspan="3">Notes</td><td rowspan="2">Data sources</td><td>Category</td></tr>
<tr><td>ID/name</td></tr>
<tr><td>Derivationmethod</td><td>Details (formula, link, etc.)</td></tr>
</table>

questions at each step. Figure 1 presents an overview of the proposed process, where primary “low-level” performance related data undergoes various processing stages to finally be utilized in various “high-level” applications.

Starting with data sources, one has to address the multiplicity of possible source data structure and formatting. There are multiple standards of raw data storage (e.g., tabular files, database files, spread sheet files etc.) and formatting (e.g., date formatting, number formatting, header structure, measurement unit, etc.). These are related, in part, to manufacturers’ standards (e.g., for sensors), performance application output standards, or individual users’ needs or preferences. To prepare the data for further processing, frequently a universal or automated extraction of measured data is not feasible. Rather, customized approaches are often necessary. Other preprocessing steps such as temporal aggregation/segmentation and quality check can be performed depending on the data specifics. The need for aggregation or segmentation depends on data source’s sampling type. Fixed temporal intervals in ontologized dataset could increase usability in future applications (e.g., in statistical analyses). Quality check is important especially when data comes from physical sensors. These are prone to power outages or communication interruptions.

Finally, the ontologization process – categorization and attribution of variables – can also become a challenging task. This step is highly dependent on the nature of the variable in question. More specifically, in many cases the properties pertaining to a variable cannot be scraped from the header of a data source file. More often they need to be identified or provided manually, making this step potentially prone to errors that could be hard to detect in the final semantically enriched dataset.

When choosing a deposition format, a list of requirements should be fulfilled and technical specifications considered. First and foremost, such a format must capture and store variables and variables’ metadata (of a various type) in a hierarchical manner. Secondly, given the nature of building performance data and possible large size volumes, provision of effective support for queries is of a high importance.

4 PBD ONTOLOGY IMPLEMENTATION

The proposed ontology and the data schema were tested as a part of a structured transformation workflow (see Figure 1), where performance data collected by multiple sensors monitoring indoor and outdoor environment were put through the transformation steps to form ontologically structured dataset.

The primary data source type in this implementation was multiple database files containing sensor readings supplemented with a timestamp. The source files’ standardized naming convention facilitated the identification of the performance variable and its relevant (e.g. temporal, spatial) attributes. To facilitate the attributes supplementation and deposition process, all variables and their attributes were aggregated in a single csv file. The content of this file followed the design specifications of conversion algorithm. Some of the variables required individual approaches in preprocessing or conversion (e.g. sky scanner, luminance camera) due to non-standard data source formats.

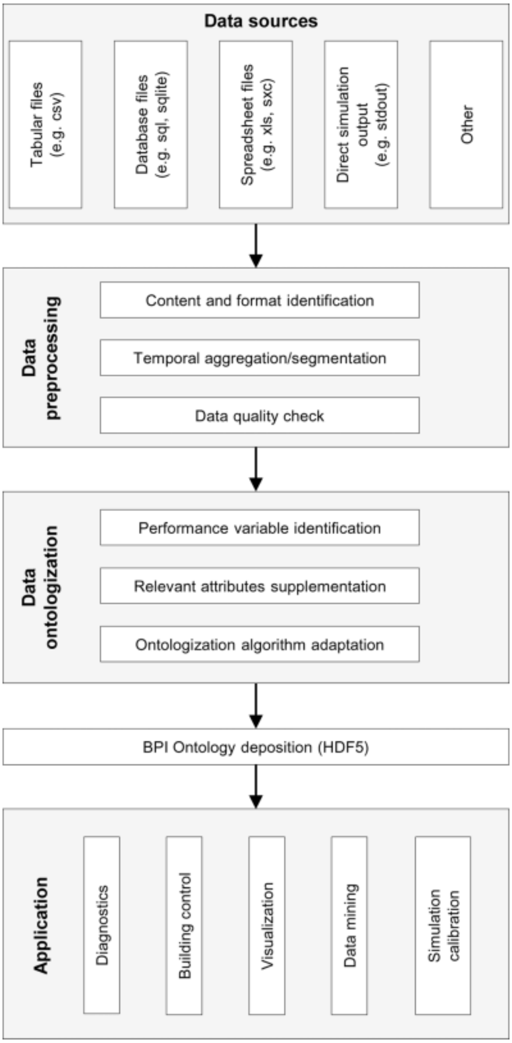

Figure 1. Schematic process overview for transformation (preprocessing, ontologization, storage) of performance-data for use in various down-stream applications (modified based on Mahdavi & Wolosiuk 2019b).

Python programming language was used in the process of performance data transformation. Several functions for conversion, quality check, attributes supplementation, and serializations were written in a HDF5 file (The HDF Group 2019). HDF5 is a primarily scientific data format capable of storing various data type structures. It is capable of storing these structures in hierarchical manner and it makes it possible to assign and accommodate complex metadata to the elements of the structure. Given these features, the HDF5 file format appeared to provide a suitable foundation for testing the BPD Ontology and the proposed data schema. Note that we do not suggest that this is the only or the ultimate implementation solution. Another implementation approach, for instance, could be based on Semantic Web technologies.

The semantically enriched dataset was then tested in a series of applications (see Mahdavi & Wolosiuk 2019a, 2019b). The following section presents an illustrative instance of an advanced application.

4.1 *Illustrative example*

In this illustrative example a task specific interface between ontologized data, stored in an HDF5 file, and Ladybug Tools was created. The interface relies on defined terms and data structure of BPD Ontology.

Ladybug Tools (Roudsari 2013) integrates the potential of well-known performance simulation engines such as the EnergyPlus (Crawley et al. 2001) or Radiance (Ward 1994), with the Rhino 3D (MCNeel 2019) modelling software. It is a collection of small applications that couple these simulation engines with a 3D modelling and visualization potential of the Rhino software. This linking is enabled through Grasshopper – a visual programming environment built-into the Rhino software.

4.2 *From BPD ontology to Grasshopper*

Grasshopper quickly grew beyond initial 3D algorithmic modelling and parametric design platform. This was enabled by giving the community a possibility of creating custom components and component packages that could be shared online. The Ladybug tools is an instance of such component package development. Grasshopper supports a number of programming languages for creating new components. These components can be of a universal nature (simple mathematical operations on input) as well as complex or task specific nature, such as calling external software for output generation. The latter is the case of this implementation, where several custom components written in Python language were created to enable and test interfacing between BPD ontology serialized in a HDF5 file and elements of Ladybug Tools. In general, the created custom components take HDF5 file input, and some input options to extract specific performance variable in a desired quantity and form that is consistent with particular requirements of Ladybug's component input. Other basic scenarios that were tested involved browsing HDF5 content for specific performance variable selection (based on attribute filtering) for data visualization in Rhino 3D.

4.3 *Location-specific daylight studies*

Many Ladybug elements take EnergyPlus Weather Files (EPW) as an input. EPW is a tabular-style file containing detailed (hourly) typical meteorological year (TMY) weather data for a specific location. One of the elements of the EPW file is the information on global, diffused horizontal, and direct normal solar radiation. This data is used as an input for some of the daylight study simulations provided in Ladybug. Specifically, it is used for generation of sky model for the Radiance simulation software.

Figure 2. A custom component for modification of EnergyPlus weather file, for use with climate-based sky generator in solar radiation studies.

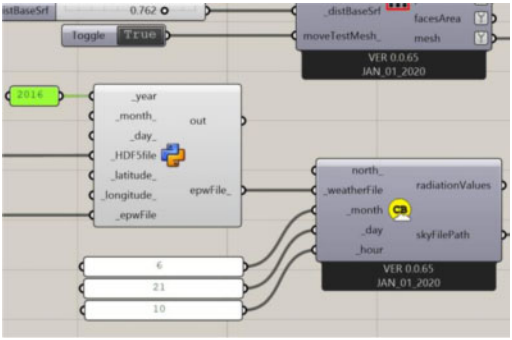

Figure 3. Custom BPD interfacing component integrated into simulation setup.

In this scenario, an application specific component was created to take measured radiation data from the HDF5 file and to replace TMY values in the original EPW file. Values recorded in a specified year (or a specific period within a year) are *i*) aggregated in terms of hourly values, *ii)* direct normal component is calculated from global and diffused horizontal radiation (if not provided), *iii)* latitude and longitude information is replaced (if provided), *iv)* and finally a new EPW file is generated, stored locally, and provided as an input for the *Generate Climate Based Sky* component. Figure 2 shows this component with the required and optional inputs on the left and outputs on the right side. Being able to modify parts of an existing EPW allows for seamless integration of localized environmental data into the standard Ladybug design or analysis workflow.

The created BPD interfacing component was tested by integration in two illustrative simulations, involving snapshot-type Illuminance and annual Daylight Autonomy analyses. The test case for the created component involved modified sample studies provided by the creators of Ladybug software.

The illuminance simulation results are based on selected diffuse horizontal and direct normal irradiance values obtained from modified EPW file generated by our custom component. The new EPW file contains radiation values for the entire year as recorded by pyranometers in 2016 in Vienna city center. The

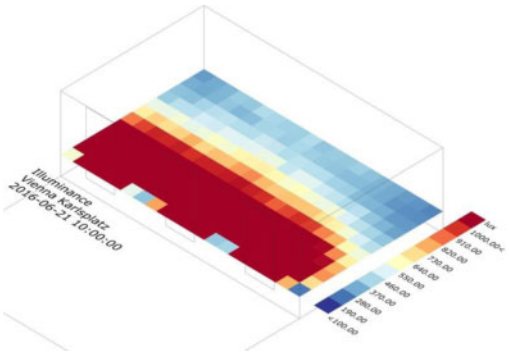

Figure 4. Visualization of the indoor illuminance simulation results based on local historical data for Vienna.

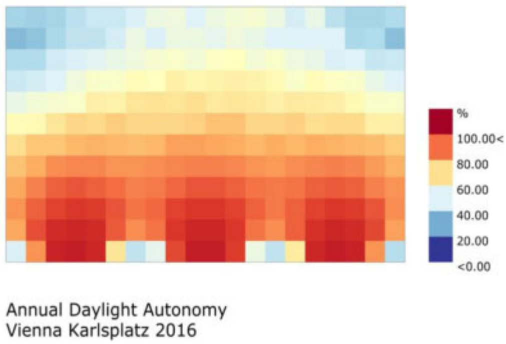

Figure 5. Daylight autonomy studies based on local data extracted from BPD stored in a HDF5 repository.

selected point in time for the simulation is 2016-06-2110:00:00. Figure 3 presents the integrated BPD interfacing component into Ladybug components-based simulation setup (only a small part of the setup canvas is visible here). The resulting daylight illuminance distribution values are visualized in Figure 4 in terms of a color scale.

In a similar manner, an annual Daylight Autonomy for the same space was performed. This time the local historical solar radiation data from the entire year 2016 was used to visualize and analyze Daylight Autonomy (DA) based on a default office type occupancy schedule. The simulated results represent the percentage amount of occupancy time when the illuminance is above the given threshold of 300lux. Again, results are visualized as a color-mapped grid representing value threshold for a given analysis grid tile (see Figure 5).

5 CONCLUDING REMARKS

This paper discussed the recent version of the BPD ontology and its data schema. We discussed the process and challenges concerning the standard ontological data transformation process. We further presented instances of advanced application of semantically structured datasets. Toward this end, "ontologized" dataset was deployed, which is based long-term daylight monitoring data from our Department's microclimatic observatory. We illustrated how this dataset was used to support building performance analyses and visualization pertaining to indoor illuminance distribution and Daylight Autonomy.

As such, the experiences thus far confirm the promising potential of the proposed approach. Thereby, multiple streams of monitored or computed building performance data, once structured according to a standardized ontological schema (BPD) can be made available for seamless utilization in multiple application scenarios, involving not only building design, but also building operation and management support.

REFERENCES

Crawley, DB., Lawrie, L K., Winkelman, F.C. et al. 2001. EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4), 319–331.

Mahdavi, A. 2011. People in Building Performance Simulation. In Hensen, J.L.M., Lamberts R. *Building Performance Simulation for Design and Operation*. Taylor & Francis Group.

Mahdavi, A., Glawischnig, S., Schuss, M., Tahmasebi, F. & Heiderer, A. 2016. Structured Building Monitoring: Ontologies and Platform. *Proc. of 11th ECPPM Conference*.

Mahdavi, A., Taheri, M. 2017. An Ontology for Building Monitoring. *Journal of Building Performance Simulation*. 10(5–6), 499–508.

Mahdavi, A., Taheri, M. 2018. A building performance indicator ontology. *Proc. of 12th ECPPM Conference*, 385–390.

Mahdavi, A., Wolosiuk D. 2019a. Integration of operational data in building information modelling: From ontology to application. CLIMA 2019 Congress. *E3S Web Conf.Volume 111*.

Mahdavi, A., Wolosiuk D. 2019b. A Building Performance Indicator Ontology: Structure and Applications. *16th International IBPSA Conference*.

McNeel, R. & Associates 2019. Rhinoceros Version 6. Available from https://www.rhino3d.com/

Roudsari, M.S. & Pak, M. 2013. Ladybug: a parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design. *Proc. of the 13th International IBPSA Conference*.

The HDF Group. Hierarchical Data Format Version 5 (HDF5) 2019. Available from https://www.hdfgroup.org/

Ward, G.J. 1994. The RADIANCE Lighting Simulation and Rendering System. *Proc. of the 21st SIGGRAPH Conference*.