SEMANTIC MAPPING ANALYSIS OF DIGITAL BULDING LOGBOOK/ PASSPORT MODELS

Sara Karami[1,4,5], José-Paulo de Almeida[2,5], Claire Ellul[3], Alberto Cardoso[1,4,5] [1]Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal [2]Department of Mathematics, University of Coimbra, Coimbra Portugal
[3]Department of Civil, Environmental & Geomatic Engineering, University of College London, London, United Kingdom
[4]Centre for Informatics and Systems of the University of Coimbra (CISUC)
[5]Institute for Systems Engineering and Computers at Coimbra (INESCC)
sarakarami@dei.uc.pt, zepaulo@mat.uc.pt, c.ellul@ucl.ac.uk, alberto@dei.uc.pt

Abstract

A Digital Building LogBook/Passport (DBL/DBP) is a repository for static ”as built” and dynamic ”Key Performance Indicators (KPIs)” throughout the life cycle of the buildings. This ongoing research uses the semantic mapping approach to analyse and compare the current recommended DBL/DBP models to identify the similarities/differences between them. The findings of this investigation show that due to the lack of comparison studies on DBL/DBP models in academic literature, the similarity percentage between proposed categories and elements is very low, and a knowledge-based method would be needed that makes the comparison not straightforward. In particular, the lack of metadata caused an increasing uncertainty percentage of elements. The focus of this paper is on assessing the DBL/DBP models (comparison of categories and elements, metadata evaluation), further work is needed to develop an integrated model.

Introduction

Buildings account for 40% of total energy consumption in the EU and are a major contributor to greenhouse gases (GHG) and energy consumers globally (Koltsios et al., 2022). Current European regulations aim to achieve the Paris Agreement as a nearly zero operating energy target in buildings (Ahmed et al., 2022). Net-Zero energy buildings are defined as a building or construction that has a zero-net consumption of energy or zero carbon emissions over a set period (Sartori et al., 2012) and are classified into four types: zero-energy sites, zero-emissions buildings, zero-energy sources, and zero-cost energy buildings (Ahmed et al., 2022). Space heating, hot water production, lighting, and the operation of various electric appliances, all use operational energy and account for the majority of total life cycle energy use (Vourdoubas, 2017).

Collecting Net-Zero emission building data in a single repository facilitates monitoring of greenhouse gas emissions across the entire life cycle of constructed assets within global, regional, and national frameworks (Society, 2020). Implementing a unified repository enables the thorough tracking and analysis of emissions data, hence facilitating the identification of patterns, trends, and areas that require improvement. Decision-makers can then use this information to guide the creation and application of strategies and policies to help the built environment reach Net-Zero emissions targets.

Building LogBooks and Passports

Töpfer (1997) introduced the first definition of a Building LogBook (BL) as a tool to improve the transparency of technical properties, standards of building services, quality of use, and operation costs of buildings by communicating comprehensive information to clients and buyers of the new property, developers, and real estate agents. The term ”Building Passport (BP)”, has now replaced the Building LogBook. A BP is a simple format for presenting both the ”birth certificate” and ”health certificate” of a building. The birth certificate specifies the key performance elements of the building to operate sustainably during the design process or after the first year of operation (Virta et al., 2012). The ”health certificate” compares the building's operation to previous years' operations and provides a short and long-term repair and retrofitting plan (Virta et al., 2012). The Global Alliance for Buildings and Construction (GlobalABC) defined the BP as a repository for all construction types and real estate activities throughout the life cycle. It will be able to create and update itself, feed on data from various data sources (e.g., Internet of Things (IoT), and Digital Twins (DT)), and enable effective virtual representations of physical assets UNEP (2020).

Semantic mapping analysis is the process of calculating the relationships between texts based on the semantic similarities methods. These methods combine automatic and knowledge-based measurements, to determine similarities and differences between phrases and sentences based on their meaning (Chandrasekaran and Mago, 2021).

This paper provides an overview of current BP/BL models to identify their main differences and similarities with semantic methods (see Experiment section), which are the fundamental basis to assess the potential for creating a single integrated model. Understanding the variation between models helps in fully comprehending the primary (minimum, lowest common denominator) BL/BP data elements, by examining element commonality and similarity. It also enables the development of a fully comprehensive model that covers all elements of all BPs. Both models help to organize data in a way that reflects the relationships among different pieces of information that can be understood by both technical and non-technical users, which is crucial for effective decision-making and system development.

This paper only explores approaches currently available from the literature, and hence the effort required to integrate these data sources. The resulting integrated model is aimed to be used by various use cases, who will be able to develop a sub-set (view) on the model that contains only the data that they require, and hence identify how to source the required data in the different passport contexts and countries, greatly reducing the time for data discovery and integration.

Experiment

Current DBL/DBP initiatives

Numerous research initiatives received funding through the European Horizon 2020 program to encourage increased energy renovations across the European Union. EU Commission study (EU) proposed a DBL data model to enhance energy efficiency and sustainability (Dourlens-Quaranta et al., 2020). As part of the iBRoad project, Libório et al. (2018) proposed a BL data model to support the implementation of a building's renovation road map for single-family houses based on occupant needs and preferences. As part of the ALDREN project, Sesana et al. (2020) proposed a tailored conceptual building Renovation Passport model to understand the state of the non-residential buildings and inform building owners about the technical energy performance status. Within the X-tendo project, a BL model that can be attached to the EPCs to inspire the next EPC generation was proposed (Toth et al., 2021). The EUB SuperHub project proposed a DBL model to achieve the EU's sustainable buildings (Malinovec Puček et al., 2023). The DBL report used the existing European INSPIRE initiative and proposed the DBL semantic data model (van der Ende et al., 2023). A green Building Renovation Passport (UKGFI) was proposed to enable data for energy performance estimates and retrofit assessments in (Small-Warner and Sinclair, 2022). THE BIM4EBB[1] developed an interoperable BIM-based toolkit for efficient building renovations. The BIM4Ren[2] proposed workflow for the renovation process and information requirements.
The GlobalABC proposed the main contents of BP based on the integration of Woningpas[3], the Building Renovation Logbook initiative, Ukraine[4], and the Building data Collection initiative (UNEP, 2020). Miller (2016) developed a BP to capture building energy efficiency data on residential buildings. Within the chronicle[5] project, a DBL framework is developed to achieve sustainability targets and long-term maintenance and renovation plans. National initiatives such as Madaster[6], Passeport Efficacité Énergétique[7], PAS-E[8], BASTA Logbook[9], Property Logbook[10],and Building Renovation Logbook[11], do not provide an official data model. The initiatives that provided the data model are listed in Table 1.

Previous Attempts at Integration

Although the approach of integrating BL/BP models was not addressed before, various Energy Performance Certificate (EPC) integration initiatives exist ((EPC and BP are the sources of a building's operation consumption and energy performance)). Serna-GonzÁLez et al. (2021) proposed a harmonization EPC data model by combining models from Italy and Spain. The Hale Studio tool was used to map the target data model to the source data model(Serna-GonzÁLez et al., 2021). Popa et al. (2022) created an integrated EPC model from England, France, Scotland, and Ireland to determine a dwelling's energy performance rating. Pouliot et al. (2018) used the Open II (Open Information Integration) tool to compare the geospatial standards from syntactic, structural, and semantic points of view. The Hale Studio [12] and OpenII [13] are powerful data transformation and harmonization tools, which don't include functionality for comparing and identifying similarities between texts.

Semantic similarity

Semantic similarity is a technique for assessing the degree of semantic equivalence between two texts, providing a quantified measure of similarity instead of a binary determination of whether they are similar or dissimilar (Chandrasekaran and Mago, 2021). Existing methods for short text similarity calculation can be roughly divided into two categories: word-level semantic-based and semantic modelling-based (Yang et al., 2021).
word-level semantic-based: The similarity of two short texts is calculated by aggregating the similar words in both texts, including knowledge-based and corpus-based methods.

- **Knowledge-based methods:** This method uses the shortest path method (Yang et al., 2021), Lexical (Farias et al., 2016), and syntactical similarity (Costin and Eastman, 2019), with the knowledge-based approach to computing the semantic similarity between two short texts (Chandrasekaran and Mago, 2021). This approach allows for a representation of human common sense and demonstrates improved performance compared to other methods (Yang et al., 2021). The main drawback of shortest path and lexical, respectively, is that the texts must be synonymous, and the lexical technique only considers zero and 100% similarity. The syntactically similar measure quantifies the similarity of two text strings for approximate string matching or comparison. For example, the

[1]https://www.bim4eeb-project.eu/
[2]https://bim4ren.eu/
[3]https://woningpas.vlaanderen.be/
[4]https://eeplatform.org.ua/
[5]https://www.chronicle-project.eu
[6]https://madaster.com
[7]https://www.experience-p2e.org/
[8]http://pas-e.es//en
[9]https://www.bastaonline.se
[10]https://www.propertylogbook.co.uk/
[11]https://eeplatform.org.ua/
[12]https://wetransform.to/halestudio/
[13]https://sourceforge.net/projects/openii/

strings "Sam" and "Samuel" are similar and have the same character sequence (Pradhan et al., 2015).

- **Corpus-based methods:** This method extracts the context information of words from large corpora and then induces the distributional properties of words (Yang et al., 2021). It does not capture the semantic relationships between words. It treats each term in isolation, without considering synonyms, related terms, or context.

Semantic modelling-based: Semantic modelling-based methods model a text as a whole to obtain the semantic vector and calculate the similarity of two short texts by performing algebraic operations on the obtained vectors (Yang et al., 2021).

- **Explicit Semantic Representation:** ESR models focus solely on word-level representations, overlooking the valuable information conveyed by syntax (Yang et al., 2021).

Methodology

To evaluate the potential of combining existing BP towards creating a comprehensive passport model, their similarities and differences in elements (attributes) classification, and metadata availability need to be investigated. The overall approach is organized as follows:

Step 1: Identify LogBooks/Passports. This included a comprehensive literature review, using the following search terms in the Google Scholar engine in all fields of the publications, such as titles, abstracts, keywords, and the entire text of the papers: "LogBook", "Building LogBook", "Building Passport", "Material Passport", "District LogBook", "Building Renovation Passport", "Green Building", "Building Certificate". Publications from 2000 to 2023 in the form of journal and, conference papers were searched, with language limited to English to allow for semantic matching.

Step 2: High level comparison. The papers that provided information about the BP structure and classification were selected and reviewed to identify the detailed data model presence that could be used for comparison, and the presence of metadata describing the model. Following this stage, BPs with sufficient detail for semantic comparison were selected. To enable the comparison process, a single BP was selected, to be the model to which all other models will be compared.

Step 3: Detailed data model comparison for the selected BP

- **Format, type, and level of classification investigation:** This step involved understanding how different elements in the models are classified (grouped), the types of classifications used, and the hierarchical classification level.

- **Manually Transfer Models into Excel:** The models were initially in PDF format, and for the initial comparison, each model was manually entered into a distinct Excel sheet. This involved taking each element in the model and manually typing it out into the corresponding sheet. A sheet was created for each classification identified above, with a column created for the elements from each BP.

- **Exclude the dynamic data:** To evaluate the infrastructural elements of BP and due to the different natures of static and dynamic data, the dynamic elements were excluded.

- **Import into a Database:** SQL (Structured Query Language) is commonly used for querying relational entities (Date, 2011). Each sheet model in CSV format was imported into a database; i.e. one entity (table) was created for each classification for each BP. This enables semantic comparisons to be performed using Transact-SQL queries.

Step 4: Semantic comparison. The similarities and differences between the classifications and elements of each model by the other models were examined. The knowledge-based and syntactical methods were used to conduct a more detailed semantic comparison of the category and element at the static information level. The LIKE query was executed to model the syntactical method, which allows finding all textual elements similar to the source element. This method is acknowledged as a semi-automatic method in this paper. The following is the comparison process:

- **Identifying the elements with similar terminology:** A semi-automatic method was used to find textually similar element names.

- **Identifying the elements with similar meaning but different terminology:** A knowledge-based approach was used to find the elements that did not share the same terms but were similar in meaning. Where available, metadata was used where this was not available a comprehensive exploration of the documentation was undertaken to confirm any interconnection between the data element names.

- **Identifying the elements with similar terminology potentially different meaning:** As above, after using the semi-automatic method, a knowledge-based approach was used to review and clarify the elements, using metadata and BP documentation where possible.

- **Identifying the elements with uncertain meaning:** A final category of elements are those where the meaning is uncertain and where the documentation and metadata do not provide sufficient clarity to match the elements.

Step 6: Key information comparison. The main BP content proposed by the GlobalABC UNEP (2020) was compared to categories and elements of identified models to assess their similarity.

Step 7: Similarity methods comparison. The average similarity percentage was examined to determine whether semi-automatic or knowledge-based approaches performed better.

Step 8: Statistical analysis. Aggregating all of the outcomes above, the similarity percentage between categories and elements was calculated to assess how much information could be collected from each model. Elements with a high level of certainty were chosen for analysis, while those with unclear meanings were excluded. It is important to note that elements lacking consistent terminology; i.e. where meaning was uncertain, have been removed from the analysis.

Results

Analyzing BL/BP models

A total of 7 BL/BP data models were identified and summarised in Table 1. They have a different number of categories that provide different information. The first column of the Table 4, Table 5, Table 6, Table 7, and Table 8 provides the category description.

Table 1: Summary of Identified BP

Source	Summary	Metadata Available	Number of Categories	Total Number of elements
EU	Proposing a DBL model to achieve energy efficiency and sustainability	Partial	7	76
X-tendo	Proposing a DBL model to inspire the next generation of EPCs information	No	7	173
EUB	Proposing a data structure for DBL to achieve energy efficiency and sustainability	Partial	7	167
ALDREN	Proposing a Building Renovation Passport (BRP) model to understand the technical energy performance status of non-residential buildings	No	5	66
iBRoad	Proposing a framework for managing energy performance, executed maintenance activities, and building plans	Partial	4	73
UKGFI	Proposing a green BRP model to improve energy efficiency	No	8	75
DBL report	Proposing a semantic DBL data model	Partial	7	86

Performing semantic comparison

A preliminary comparison was performed to select a more complete model regarding the building's fundamental elements. EU and X-tendo models provided more information on the buildings, and EU was selected as a starting point. In the semi-automatic part, a script with the LIKE operator was executed per category, and the identified elements with the same input terms were displayed in the results. Similar results were then imported into the repository, which was used for the next comparisons. This process was repeated for each model until it reached the final one, allowing all models to be compared. It is important to note that some steps were iterative since previously analyzed information had to be reanalyzed several times along the line, as the need to understand the problem better increased in subsequent rounds. The following are some illustrative examples of the comparison process.

- **elements with similar terminology:** The "Floor area" element in the EU model's *Building descriptions and characteristics* category, was identified as a match by the automated checking process to the "Total floor area (unheated and heated area)" in the EUB model's *General Building Information* category, "Floor area" in the X-tendo model's *Building descriptions and characteristics* category, "Net floor area" in the iBRoad model's *General and administrative information * category and "Floor area and building volume" in the UKGFI model's *Building type* category.

- **elements with dissimilar terminology but similar meaning:** The "Building information model/Design and plans of the building" element in the X-tendo, EU model's *Building descriptions and characteristics* category, could be assumed to be similar to the "Layout of the whole building for multi-unit properties (i.e. block of flats, terrace housing) " in the UKGFI's *Building type* category, and "3D model/Architectural plans" in the EUB model's *Building Documentation BIM* category. One of the main limitations of an automated approach is identifying similarities, where there are no identical terms between models.

- **elements with similar terminology but uncertain meaning:** The "Address" has been considered in the EU and X-tendo models's*Administrative* category, which is undefined and is related to the building owner's address, occupier address or the building's address. The meaning of the "Storage" element in the X-tendo model's *Building descriptions and characteristics* category, and the "Energy supply and storage" in the UKGFI's *Building Services* category are unclear.

- **elements with uncertainty meaning:** The meaning of "Renewable energy systems" in the EU model's *Building descriptions and characteristics* category, "Any existing and planned local energy schemes" in the UKGFI model's *Building Service* category, "Optimizing self-consumption of locally generated energy" in the ALDREN model's *Technical system* category, and the "isDescribedByNativeGIS" in the DBL report's *Location* category, are unclear. In the EU and X-tendo models, the "U value of different components" is proposed as an element in the "Building Performance" category, although its meaning and definition are unclear.

- **elements with uncertainty metadata:** The description of "Ventilation systems" or "Cooling systems" in iBRoad model's *Building construction information* category is ”Description of the building ventilation systems” and ”Description of the building cooling systems”, which do not encompass the essential requirements that should be incorporated into it.

duplicated elements: The ”Safety Manual” proposed in the EU model includes building operations and maintenance, and security procedures, whereas the ” Maintenance records and information” element, also has been proposed by EU, X-tendo, UKGFI, and iBRoad models. The ”Building pictures” in the EUB report's *Building Documentation BIM* category was repeated in the EUB report's *General Building Information* category.

Our results demonstrated that significant progress in defining a comprehensive BL/BP model has been made, though there is still a lack of interoperability between elements, data consistency, metadata availability, and information exchange.

Comparing some key information

The percentage of similar elements proposed by more than 3 models is 6 %, and the similarity percentage between elements is low (See section Statistical analysis), making it challenging to identify the key elements. Table 2 compares the main content proposed by GlobalABC and identified BP models (categories and elements). As can be seen, in some cases, the contents are similar to the group of model elements (e.g., Building description, Technical features, Use and operation). In these cases, similar elements were found using the knowledge-based method in addition to the semi-automatic method. The semi-automatic method was used to find similar elements to the *Contracts*, *Material inventory*, *Building description*, *Operation*, *Energy Performance Certificate*, and *Maintenance*. The knowledge-based was performed to find similar elements to the *Identification of buildings*, *Identification of the plot*, *Material inventory* and *Technical features*. The models did not include the *Environmental performance and carbon footprint*, *Surfaces*, *Dismantling and recycling strategy*, *Results of user satisfaction survey* and *Certificate* elements. The EUB model only incorporated the *Proof of maintenance* (as the maintenance report element), *Design documents* (as the Architectural plans and As-built plans), and *Operational cost*.

Comparing similarity methods

Table 3 compares the semi-automatic and knowledge-based methods. The similarity was calculated by the average number of similar certain and uncertain elements (elements were unclear or did not have similar words but were similar in their meaning) by the total number of its elements. The semi-automatic method performed better for the X-tendo and EUB models, while the others were processed using the knowledge-based approach.

Table 2: Comparison of some key information

Content	EU	X-tendo	EUB	ALDREN	iBRoad	UKGFI	DBL report
Identification of plot and plot-related characteristic	**Element:** Building owner	**Element:** Buiding owner	**Element:** Building owner	**Element:** Owner name		**Element:** Owner's details	**Element:** Owner
Identification of building	**Element:** Unique building identifier	**Element:** Unique building identifier	**Element:** Unique building identifier			**Element:** Unique building identifier	
Contracts	**Element:** Utility,Service contract,Insurance	**Element:** Utility,Service contract,Insurance	**Element:** Utility,Service contract,Insurance				
Energy Performance Certificate / sustainability label	**Element:** EPC rating	**Element:** EPC rating	**Category:** EPC certificate		**Element:** EPC number,Issue date		
Material inventory	**Category:** Building material inventory	**Category:** Building material inventory	**Element:** Wall,Floor,Roof			**Element:** Materials Used	**Category:** Structure & Material
Building description	**Category:** Building descriptions and characteristics	**Category:** Building descriptions and characteristics					
Technical features and characteristics	**Category:** Building operation and use	**Category:** Building operation and use	**Category:** Building Element Information	**Category** :Building operation and use	**Category:** Building operation and use		
Use and operation data / consumption	**Category:** Building performance	**Category:** Building performance	**Category:** Building performance		**Category:** Building energy performance	**Category** :Energy consumption & use behaviour	**Category** :Prformance
Maintenance manuals	**Element:** Maintenance log	**Element:** Maintenance log	**Element:** Maintenance log		**Element:** Maintenance information	**Element:** Maintenance records	

Table 3: Comparison of similarity methods

	X-tendo [%]		EUB [%]		ALDREN [%]		iBRoad [%]		UKGFI [%]		DBL report [%]	
Source	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based
EU	37	4	20	6	0	4	1	4	0	10	6	3
X-tendo	-	-	16	4	1	2	0	4	1	5	2	1
EUB	-	-	-	-	2	4	1	7	0	6	2	3
ALDREN	-	-	-	-	-	-	1	5	1	6	3	11
iBRoad	-	-	-	-	-	-	-	-	0	6	2	4
UKGFI	-	-	-	-	-	-	-	-	-	-	0	2

Performing statistical analysis

The similarity percentage of classifications for each model to the other models was calculated by the average number of similar categories divided by the total number of categories in the model. The EU model with 57%, X-tendo with 57%, and EUB with 51 % were the most identical to the other models. ALDREN with 40%, iBRoad with 35%, DBL report with 25%, and UKGFI with 10 % were in the subsequent ranks. As can be seen in Table 4, Table 5, Table 6, Table 7, and Table 8, the ”Building Documentation BIM” category is not included in the other models, or the ”Envelope” category of the ALDREN model is relevant to the ”Building construction” category of iBRoad. The location element has been considered in the other models but not in a separate category; ALDREN's model is the only one considered a category for location.

Table 4 demonstrates the similarity percentage of EU elements to the X-tendo, EUB, ALDREN, iBRoad, UKGFI, and DBL report models using the semi-automatic and knowledge-based methods. It was calculated by the number of similar elements with clear meaning in the EU model by the total number of its elements. Table 5, Table 6, Table 7, and Table 8 shows the similarity comparison for the other models. As can be seen, the overall matching percentage in terms of category and elements, with X-tendo and then EUB, is significantly higher than with the other models. The low percentage of similarities arises from diverse elements incorporated within the same categories across different models. For example, "Building age" is part of the iBRoad "General and administrative information" category, while the "Year Built" is included in the EU "General Building Information" category.

Table 4: EU similarity comparison with X-tendo, EUB, ALDREN, iBRoad, UKGFI, DBL report

	X-tendo [%]		EUB [%]		ALDREN [%]		iBRoad [%]		UKGFI [%]		DBL report [%]	
EU	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based
Administrative	33	0	17	17	0	0	0	0	0	17	0	0
Building descriptions and characteristics	14	0	5	0	0	0	0	10	0	5	4	0
General Building Information	29	0	14	14	0	14	0	14	0	14	0	14
Building operation and use	0	0	0	0	0	0	0	0	0	0	0	0
Building performance	33	0	0	0	0	33	16	0	0	17	0	0
Finance	46	0	7	0	0	0	0	0	0	0	0	0
Building material inventory	0	0	0	0	0	0	0	0	0	0	0	0

Table 5: X-tendo similarity comparison with EUB, ALDREN, iBRoad, UKGFI, DBL report

	EUB [%]		ALDREN [%]		iBRoad [%]		UKGFI [%]		DBL report [%]	
X-tendo	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based
Administrative	38	0	0	0	0	0	0	8	0	0
Building descriptions and characteristics	3	0	3	0	1	0	1	0	1	0
General Building Information	11	0	0	0	0	0	0	0	0	0
Building operation and use	0	0	0	0	0	0	0	0	0	0
Building performance	20	0	0	0	0	0	20	0	0	0
Finance	23	0	0	0	0	0	0	0	0	0
Building material inventory	0	0	0	0	0	0	0	0	0	0

Discussion

The BL/BP is a summary of all the key information about the building, including the original design, commission-

Table 6: EUB similarity comparison with ALDREN, iBRoad, UKGFI, and DBL report

	ALDREN [%]		iBRoad [%]		UKGFI [%]		DBL report [%]	
EUB	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based
Administrative	19	0	0	0	0	0	10	4
General Building Information	8	4	8	8	8	8	4	0
Building Element Information	0	7	0	13	0	0	0	0
Building operation and use	0	0	0	0	0	0	0	0
Building performance	0	0	10	20	0	0	0	0
Finance	0	0	0	0	0	0	0	0
Building Documentation BIM	0	0	0	0	0	0	0	0

Table 7: iBRoad similarity comparison with UKGFI, DBL report & UKGFI to DBL report

	UKGFI [%]		DBL report [%]	
iBRoad	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based
General and Administrative	7	18	7	14
Building construction	0	0	0	0
Building energy performance	0	0	0	0
Building operation and use	0	0	0	0

	DBL report [%]	
UKGFI	Semi-automatic	Knowledge-based
Ownership & governance	0	0
Building Service	0	0
Energy consumption & use behaviour	0	0
Building Type	0	0
Building Fabric	0	0
Climate resiliency	0	0
Circular economy considerations & enhanced climate information	0	0
Building information	0	0

Table 8: ALDREN similarity comparison with iBRoad, UKGFI, DBL report

	iBRoad [%]		UKGFI [%]		DBL report	
ALDREN	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based	Semi-automatic	Knowledge-based
Building features	0	20	0	20	10	30
Location data	0	0	0	0	12	37
Building geometry	16	0	16	0	8	0
Envelope	0	0	0	0	0	12
Technical system	0	0	0	0	0	0

ing, and handover details, and information on its management and energy performance, the material used, sustainability performance, indoor environmental quality, and potential energy throughout the building life cycle (Small-Warner and Sinclair, 2022).

The main purpose of our approach was to evaluate the differences and similarities between the data models that underpin each BP and assess whether semi-automated or knowledge-based approaches would be useful to create an integrated data model of the various BP. This work showed that the currently proposed BL/BP models, firstly have different categories (groupings or classifications of elements) which, when examined in detail, overlap in some cases and are not aligned with the proposed categories by GlobalABC (UNEP, 2020). This adds a layer of complexity to the potential for automated integration, which is further complicated by the fact that elements (attributes) within each classification sometimes overlapped or were included in different groupings in the different BP. Underpinning the automation challenges are gaps in the lack of an agreed and standardized conceptual BL model accompanied by detailed metadata, which would provide clear element definitions, including the data type, and sufficient data descriptions. This could be further complicated when it is considered that not all these BP originals in English-speaking countries - and local terminology - and translations- could cause additional issues. Methodologically, the use of text-based comparison (LIKE) for matching could also introduce false positives (See section Step 4 for examples) and manual checking was required to validate the results. The approach of selecting a single model (EU model) towards full integration of models available from the literature, was however very beneficial as it was avoided.

The comparison results showed that although the most common categories are Administrative information, Building Identification, Building performance, and Building operation (Table 2), the similarity percentages between elements are low (Table 4, Table 5, Table 6). There is a lack of clear consensus regarding the Building geometry category, suggesting the need for further discussion. The similarity percentage between models on Building equipment information (e.g., Heating, Cooling, DHW, etc) is low, which is substantial in proposing the integrated model. The models made no distinction between Building and Building unit elements, which are required for creating an integrated mode. The EUB model incorporated the EPBD indicators (Energy Performance Building Direction) [14], however further discussion is required to distinguish between EPBD and integrated BP elements. The categories in the integrated BP model must be harmonized according to the proposed GlobalABC report's categories. Although the static information was examined, the EUB model only included the KPI indicators. This indicates that identifying real-time data and considering the Building Automation System (BAS) in the integrated model requires significant effort.

Overall, therefore, it can be concluded that to develop an integrated model an extensively manual process will be required, starting from creating digital versions of PDF models and moving from there. Indeed, the automated SQL comparisons did not prove reliable enough to even be a first pass for the approach, given that in 32 cases similar terminology was used with different meanings, so false positives were observed (Section Methodology- Step 4). An extensive effort to identify key information in an integrated model is required due to the 42 elements shared by more than 3 models.

[14] https://eur-lex.europa.eu/resource.html?uri=cellar:c51fe6d1-5da2-11ec-9c6c-01aa75ed71a1.0001.02/DOC_1

Conclusions and further work

The BL/BP is a repository for the comprehensive collection of the most important performance characteristics and technical data feed in from various data sources (e.g., Internet of Things(IoT)) and should enable effective virtual representations of physical assets (Tharma et al., 2018). The overall purpose of this specific task of our ongoing research was to examine both the parallels and distinctions of the current recommended BL/BP models in terms of their element's contribution to the physical, energy performance characteristics, and technical data of a building using semantic similarity methods. The results showed that, while great progress has been made, the similarities between proposed models are quite low, and a knowledge-based strategy would be required. The outcome of the process above will be an integrated UML and physical data model that will allow users to trace back each element to its source or sources (BP or EPC) and make a connection between different entities. Users of this integrated model will thus be able to select the required subset of elements from the full model (as a view) and easily understand whether such data is available from data currently collected by BP or EPC. Future work will also entail the integration of the BP identified above with EPC models. These provide the necessary energy consumption information and may help achieve Net-Zero carbon objectives.

Acknowledgments

This research is being partly supported by the Portuguese Foundation for Science and Technology (FCT) I.P./MCTES through national funds (PIDDAC), within the scope of INESC-Coimbra R&D Unit (Ref. UIDB/00308/2020 Ref. UI0308/geoBIM.1/2022) and of CISUC R&D Unit (Ref. UIDB/00326/2020).

References

Ahmed, A., Ge, T., Peng, J., Yan, W.-C., Tee, B. T., and You, S. (2022). Assessment of the renewable energy generation towards net-zero energy buildings: a review. Energy and Buildings, 256:111755.

Chandrasekaran, D. and Mago, V. (2021). Evolution of semantic similarity—a survey. ACM Computing Surveys (CSUR), 54(2):1–37.

Costin, A. and Eastman, C. (2019). Need for interoperability to enable seamless information exchanges in smart and sustainable urban systems. Journal of Computing in Civil Engineering, 33(3):04019008.

Date, C. J. (2011). SQL and Relational Theory: How to Write Accurate SQL Code. O'Reilly Media, Inc.

Dourlens-Quaranta, S., Carbonari, G., De Groote, M., Borragán, G., De Regel, S., Toth, Z., Volt, J., and Glicker, J. (2020). Building logbook state of play. report 1 of the study on the development of a european union framework for buildings' digital logbook. Report 1.

Farias, T. M. d., Roxin, A., and Nicolle, C. (2016). A brief overview of semantic interoperability for enterprise information systems. In International Conference on Informatics in Economy, pages 38–52. Springer.

Koltsios, S., Fokaides, P., Georgali, P.-Z., Tsolakis, A. C., Chatzipanagiotidou, P., Klumbytė, E., Jurelionis, A., Šeduikytė, L., Kontopoulos, C., Malavazos, C., et al. (2022). An enhanced framework for next-generation operational buildings energy performance certificates. International Journal of Energy Research, 46(14):20079–20095.

Libório, P., Fragoso, R., Monteiro, C. S., Fernandes, J., Silva, E., and Casteele, T. V. (2018). The logbook data quest: Setting up indicators and other requirements for a renovation passport.

Malinovec Puček, M., Khoja, A., Bazzan, E., and Gyuris, P. (2023). A data structure for digital building logbooks: Achieving energy efficiency, sustainability, and smartness in buildings across the eu. Buildings, 13(4):1082.

Miller, W. (2016). Capturing sustainable housing characteristics through electronic building files: the australian experience. In Sustainable Built Environment Conference 2016 Proceedings, pages 190–199. Karlsruher Institut fur Technologie (KIT).

Popa, A., Ramallo González, A. P., Jaglan, G., and Fensel, A. (2022). A semantically data-driven classification framework for energy consumption in buildings. Energies, 15(9):3155.

Pouliot, J., Larrivee, S., Ellul, C., and Boudhaim, A. (2018). Exploring schema matching to compare geospatial standards: application to underground utility networks. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives, 42(4/W10):157–164.

Pradhan, N., Gyanchandani, M., and Wadhvani, R. (2015). A review on text similarity technique used in ir and its application. International Journal of Computer Applications, 120(9):29–34.

Sartori, I., Napolitano, A., and Voss, K. (2012). Net zero energy buildings: A consistent definition framework. Energy and buildings, 48:220–232.

Serna-GonzÁLez, V., HernÁNdez, M. G., Miguel-Herrero, F. J., Valmaseda, C., Martirano, G., Pignatelli, F., and Vinci, F. (2021). Harmonisation of datasets of energy performance certificates of buildings across europe.

Sesana, M. M., Rivallain, M., and Salvalai, G. (2020). Overview of the available knowledge for the data model definition of a building renovation passport for non-residential buildings: The aldren project experience. Sustainability, 12(2):642.

Small-Warner, K. and Sinclair, C. (2022). Green building passports: a review for scotland.

Society, R. (2020). Digital technology and the planet: Harnessing computing to achieve net zero uk.

Tharma, R., Winter, R., Eigner, M., et al. (2018). An approach for the implementation of the digital twin in the automotive wiring harness field. In DS 92: Proceedings of the DESIGN 2018 15th International Design Conference, pages 3023–3032.

Töpfer, K. (1997). Gebäudepaß–eine serviceleistung für bauherren und nutzer. Bundesbaublatt, 9(97):608.

Toth, Z., Maia, I., Rosa, N., and Volt, J. (2021). Technical specifications of energy performance certificates data handling: Understanding the value of data.

UNEP (2020). The building passport: A tool for capturing whole life data in construction and real estate – practical guidelines.

van der Ende, M., Flickenschild, M., Borst, T., Raes, N., Cai, A., Gankema, Y., and Schinkel, R. (2023). Dbl semantic data model, providing standard form and meaning to digital building logbook data.

Virta, M., Hovorka, F., and Lippo, A. (2012). Building passport as a tool to evaluate sustainability of building. World Green Building Council Publication: London, UK.

Vourdoubas, J. (2017). Creation of zero co2 emissions residential buildings due to operating and embodied energy use on the island of crete, greece. Open Journal of Energy Efficiency, 6(4):141–154.

Yang, J., Li, Y., Gao, C., and Zhang, Y. (2021). Measuring the short text similarity based on semantic and syntactic information. Future Generation Computer Systems, 114:169–180.