Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

BIM-GIS Integration: Knowledge graphs in a world of data silos

Hou Yee Quek [1], Markus Hofmeister [1,2,3], Simon D. Rihm [1,2,3], Jingya Yan [1], Jiawei Lai [6], George Brownbridge [6], Michael Hillman [6], Sebastian Mosbach [1,2,6], Wilson Ang [1], Yi-Kai Tsai [1], Dan N. Tran [1], Soon Kang, William Tan [1], Markus Kraft [1,2,4,5]

released: June 27, 2023

[1] CARES
Cambridge Centre for Advanced Research and Education in Singapore
1 Create Way
CREATE Tower, #05-05
Singapore, 138602

[2] Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

[3] Department of Chemical & Biomolecular Engineering
National University of Singapore
4 Engineering Drive 4
Singapore, 117585

[4] School of Chemical and Biomedical Engineering
Nanyang Technological University
62 Nanyang Drive
Singapore, 637459

[5] The Alan Turing Institute
2QR, John Dodson House
96 Euston Road
London, NW1 2DB
United Kingdom

[6] CMCL Innovations
Sheraton House
Cambridge
CB3 0AX
United Kingdom

Preprint No. 311

UNIVERSITY OF CAMBRIDGE

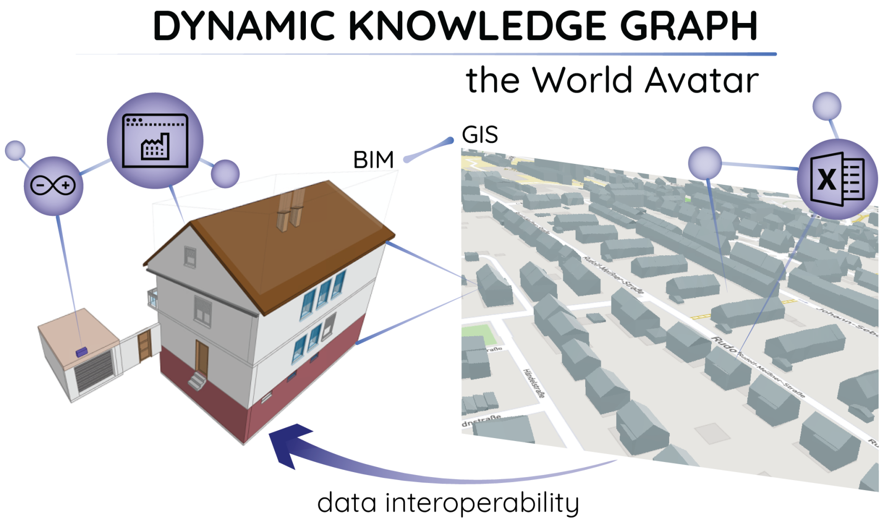
Keywords: BIM, GIS, interoperability, smart city, knowledge graph

Edited by
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom

E-Mail: mk306@cam.ac.uk
World Wide Web: https://como.ceb.cam.ac.uk/

Abstract

Cities today adopt various technologies to gather and analyse cross-domain data including Building Information Modeling (BIM) and Geographic Information System (GIS) to manage urban developments and their consequences. However, BIM-GIS integration has encountered obstacles given their disparate formats and protocols. Even when integration is achieved, solutions face difficulties incorporating other data sources such as sensors and Excel spreadsheets. Given the poor interoperability of existing urban technologies, this paper proposes a dynamic knowledge graph approach, comprising of domain ontologies, autonomous agents, and visualisation interfaces, to integrate cross-domain multi-scale data including BIM-GIS. The ontologies developed semantically annotate and represent data and their relationships with standardised definitions to align stakeholder perspectives. Agents can perform tasks such as data retrieval, processing, computation, and forecasting on the real world via the knowledge graph. The visualisation interface allows users to view and analyse real-time data for a holistic understanding of the current situation and alternate scenarios.



Highlights

- Semantic Web technologies are suited to overcome and connect current data silos
- Knowledge graphs offer promising solution to integrate both BIM and GIS data
- Integration of data beyond BIM and GIS enables real interoperability
- Dynamic knowledge graphs foster dynamism, modularity, and cross-domain collaboration
- Unified visualisation interface enables data access across domains and scales

Contents

1 Introduction

In recent years, the ever-growing digital ecosystems present in cities have gained significant attention for their transformative potential to support sustainable developments and enhance the quality of life [24, 30, 35]. Most of the technologies available generally gather data specific to their application [1, 14, 35], such as Building Information Modeling (BIM) and Geographic Information System (GIS). BIM is a tool for generating, sharing, and managing detailed building information and their geometry models throughout the building's lifecycle [8, 17]. GIS is an information system that represents, processes, and analyses data associated with a geo-location [11, 34, 37]. These technologies collect data that are then used to support various decisions in their respective domains.

However, as urban complexities and interconnectivity between various domains grow, the pressing need for cross-domain data has been hampered by the poor interoperability of technologies. Interoperability is defined as the ability of tools and systems to understand and use the functionalities of other tools. In this context, poor interoperability refers to the difficulty of integrating data from different sources due to the idiosyncrasies between these technologies' formats, protocols, and systems [30, 32]. For instance, the integration of BIM and GIS has been identified as a crucial data source to enhance the efficiency, safety, and performance of complex urban infrastructure projects [34]. But the inherently disparate data formats are a significant obstacle to their integration [11, 16, 37]. When domain specific data are not interoperable, they form data silos that withhold and isolate the access and availability of information and resources from other stakeholders [10, 15]. This hinders productivity, efficiency, innovation, and service quality [10, 15].

Existing solutions are not yet able to overcome these silos for two reasons: Firstly, these solutions have yet to demonstrate their extensibility towards other domains. Typically, existing solutions adopt additional technologies in their workflows to overcome these data silos. Examples include FME and Dynamo in the BIM-GIS context [17, 18]. Despite the increased complexity, these solutions are usually tailored for a particular domain and cannot be generalised to other technologies. Secondly, perspectives on these solutions are usually siloed, in spite of their claims. Communities tend to focus on resolving their domain-specific interoperability challenges and fail to consider the broader context. For example, standards such as Industry Foundation Class (IFC) and CityGML have been adopted for sharing 3D urban models in their BIM and GIS domains respectively. However, such standards also represent another form of silo. This creates challenges not only for integrated BIM-GIS applications, but also for other applications. It is essential to acknowledge that BIM and GIS are merely another data source, albeit a valuable one for locations, buildings, and cities [30]. There are still other sources of crucial data for various applications. For instance, Excel spreadsheets and paper documents continue to be popular tools for managing assets [17]. The increasing deployment of sensors in public and private settings has emerged as a valuable data source of environmental measures [1, 4]. It becomes clear that as technology advances, new data sources and consequently, silos are continually being introduced [30]. This raises the question on how existing solutions can integrate old and new data sources when they are focused only on resolving domain-specific interoperability problems. Hence, the rapid technological advancements alongside siloed perspectives have given rise to a pressing research problem: the challenges posed by data silos and the growing complexity of their solutions.

The research problem is framed by two research questions:

1. What are the technologies capable of breaking down these data silos?
2. How could these technologies be designed to overcome these silos in the context of BIM-GIS integration and beyond?

This paper aims to answer these two questions by demonstrating the case for adopting the dynamic knowledge graph (KG) technology to enhance the interoperability of BIM-GIS technologies and their associated data. The dynamic KG employs standardised machine-readable ontologies to semantically annotate data and their relationships. Autonomous agents can then act on the KG to retrieve, update, and process these instantiated data for visualisations or simulations, enabling the discovery and inference of new knowledge. In addition, this paper explores the extensibility and scalability of the dynamic KG by showcasing its applicability to integrate other technologies such as energy simulation software, laboratory equipment, sensors, and utility meters, across scales.

This paper is divided into four sections. Following this introduction, section 2 highlights the prevalence of data silos and proposes the dynamic KG as a potential solution. In section 3, we outline the dynamic KG methodology and explore its applications in Section 4. Section 5 summarises the findings.

2 Current state of the BIM-GIS landscape

2.1 Data silos in BIM-GIS integration

As urban complexities and interconnectivity between various domains grow, the representation and analysis of massive urban information flows becomes critical to support sustainable urban processes and enhance the quality of life [24]. Such a task cannot be accomplished by human capabilities alone, and requires taking advantage of the ongoing technological developments [30, 34]. Nevertheless, a major impediment to this endeavour is the inherent poor interoperability between most technologies due to their distinct formats, protocols, and systems [8, 11, 30, 32]. For example, the slow progress on BIM-GIS integration is predominantly caused by the underlying idiosyncracies between the IFC and CityGML specifications [11, 41].

IFC is an open standard for sharing BIM data across software platforms [11]. BIM is valuable for generating detailed 3D building models and managing their associated asset data throughout the lifecycle: from conception and construction to the operational phase [8, 14, 17]. But the lack of viable open-source alternatives has led to the dominance of commercial BIM software like Revit and Bentley, which adopt proprietary formats that have isolated data from each other. In efforts to promote data sharing, IFC was developed to break down data silos in the BIM domain.

CityGML is an open standard for representing and sharing 3D GIS building and landscape models [11]. GIS technologies have a broader scope of applications beyond the built environment industry, and also boast a flourishing software ecosystem. Comprehensive software solutions like ArcGIS and QGIS provide user-friendly interfaces for non-programmers to store, process, and visualise geospatial data as well as perform various spatial analyses. Various map visualisation libraries such as Mapbox, Cesium, and kepler.gl are also available. They require technical proficiency to embed map views into software applications and extend their capabilities. GIS also involves the generation of 3D urban models from various methods, ranging from photogrammetry and laser scanning to the use of digital modelling software like SketchUp and ArcGIS CityEngine [39]. However, the resulting models tend to be stored in various formats and have differing level of detail [20]. To address this issue, CityGML was developed to standardise these models for data sharing and overcome the data silos in the GIS domain.

Although CityGML and IFC have improved interoperability within their respective domains, they are yet another form of silo in BIM-GIS integration research. The integration of BIM and GIS would extend their individual capabilities and provide comprehensive geometry, material, and asset data that can support the analysis and visualisation of complex urban infrastructure projects to improve their efficiency and performance across scales [34]. Despite their opportunities, IFC and CityGML standards present several challenges that hinder the adoption of this multi-scale geometry and semantic data source. First, both specifications are proposed for disparate applications, and their concepts do not align in many cases [11, 37]. For example, CityGML is intended for larger scale urban representations, inclusive of demographic, terrain, and landscape elements. This differs from the detailed building representation of IFC that includes interior facilities, cost, scheduling, and topology linking the site, buildings, storeys, and its spaces. Second, they have distinct geometry modelling paradigms that cannot be directly mapped [16]: CityGML adopts a surface-based paradigm, while IFC has a solid-based paradigm. Third, the coordinate systems of IFC and CityGML are inherently incompatible [16]: IFC adopts a local placement system that is always relative to another entity. For example, a wall is placed relative to a storey, which is placed relative to a building. CityGML adopts a world coordinate system based on absolute coordinates. Finally, IFC has a richer data schema than CityGML to accommodate the granularity and range of smaller assets. Hence, although the specifications have broken down silos within their individual domains, their differences have contributed to a technical bottleneck in BIM-GIS integration research. This bottleneck highlights how these data silos are multi-layered. Improving the interoperability of technologies within one domain may not always improve their interoperability with other domains.

In understanding the full impact of these silos, we must consider the broader context in which they operate, beyond specific applications or domains. It is critical to first acknowledge that these silos are beneficial to an extent. They emerge as a tool for organisations to isolate valuable information and augment their efficiency and innovation for a competitive advantage in a market with intensifying competition [15, 24]. For example, proprietary formats perpetuate vendor lock-ins, which are beneficial for economic profits but dampens consumer value [15, 28]. Furthermore, some industries have already established a strong reliance on specific technologies. One example is the use of BIM and GIS in the architecture, engineering, and construction industry. Given the significant resources and costs associated with their initial implementation, it can be challenging to introduce new technologies that could replace them. Consequently, these economic imperatives will cause such silos to persist even as we become more aware of their repercussions.

As urban environments become increasingly complex and interconnected, reliance on a single data source is no longer sufficient to fulfil the changing requirements over time. For example, BIM data is predominantly generated for the requirements of the construction and design phase [14]. But such data often proves insufficient for managing the facility during its later phases [14]. As a result, supplementary data must be obtained from other sources, such as paper documents or Excel workbooks, and integrated into the BIM software [17].

Moreover, past efforts in integrating multiple data sources into a singular system are not scalable beyond very narrow contexts. Given the continuous need to accommodate additional datasets, current platforms effectively forming "islands" of solutions that function within their respective technologies but prove challenging to extend to different contexts and data sources. For instance, BIM integration with asset management data can be accomplished within BIM technologies, but their integration with GIS technologies requires a different solution [11, 17]. Even if BIM-GIS solutions are available, it is merely another "island solution" that requires work to extend to other technologies. This becomes a conundrum when considering that advancements in technology will generate new sources of data that should ideally be captured as well. But based on current trajectories, new technologies are more likely to create isolated silos of data and varying interoperability solutions that may not be scalable or extensible to other technologies [30]. As a result, these data silos continue to pose a growing challenge that must be addressed.

2.2 Semantic Web and knowledge graphs

Semantic Web technologies are a potential solution for bridging these silos. These technologies semantically annotate data and their relationships through ontologies based on a standardised machine-readable Resource Description Framework (RDF) format and the principles of Linked Data [6, 7]. An ontology is a formal specification of knowledge, including the concepts and relationships of any domain, that can be described and understood by humans and/or machines [3]. They usually comprise of a terminological component (TBox) and an assertional component (ABox) [36]. A TBox describes the concepts and relationships, while an ABox contains data that realises concrete instances of these abstract constructs based on real-world entities. For example, a TBox might state that every individual is a person who has another person as a parent. Conversely, an ABox will specify that Joe is an instance of a person and has a parent – Mary, who is also an instance of a person. Each instance in the ABox can be accessed with a unique internationalised resource identifiers (IRIs) on the World Wide Web, with semantics to describe the concept, its attributes, and their context. In this manner, the instances in an ontology enable the discovery, integration, and transfer of information between different domains and systems [7]. Some applications extend these capabilities using KGs to derive new knowledge and support larger scale applications [2, 21]. A KG is a collection of interlinked ontologies in a directed graph, where the nodes refer to entities of interest and the edges represent their relationships with each other [21]. Effectively, the KG and ontologies in the Semantic Web act as a layer of abstraction to standardise, integrate, and store data for any applications across domains and scales.

In the architecture, engineering, and construction industry, there has been a growing uptake of Semantic Web technologies to not just cut down data silos, but also meet the surging demand for granular and dynamic urban data across domains and scales [29, 39]. Current technologies face significantly greater difficulties in expressing meaning and adding dynamism into data. One example is in the BIM domain, where a BIM model for a window can contain information on its dimensions, materials, and manufacturer details. But in the absence of semantics, this information cannot be interpreted by a computer to infer new knowledge; it requires a human interpreter. Moreover, there is still no consistent and persistent approach to update assets' information as they are developed, added, or removed over time throughout the lifecycle [39]. On the other hand, Semantic Web technologies have already demonstrated its potential to enrich data models semantically and achieve interoperability, logical inference, consistency, and scalability across complex systems and domains [29, 39].

When implementing these technologies for BIM-GIS integration, it is crucial to note that the current state of Semantic Web approaches has not yet fully realised its potential to break down these silos in a larger context. Despite their claims of enhanced interoperability, a majority of the literature presented in recent reviews [11, 17, 41] have typically implemented their proposed solutions within a single application and context. Few have yet to demonstrate the extension of their work to new applications. This lack of continuity may lead readers to doubt the potential of Semantic Web technologies as a whole. However, we argue that this shortfall is an outcome of rigid ontology designs that are not reusable or scalable to new domains.

For example, the `IfcOwl` and `CityGML` ontologies are developed as equivalent representations of the geometry and semantic concepts in IFC and CityGML [13, 29]. Regardless, these two ontologies are not suitable for cross-domain applications for several reasons. Firstly, they are not concise as they include their original schema's intermediary relationships, which are necessary to link entities in an XML-based schema, but are redundant concepts in ontologies. One example is the `IfcRelAggregates` property in IFC that could link a chair to the floor it is found on. When instantiated into the `IfcOwl` ontology, users will have to generate three relationships:

```
:IfcRelAggregates_01 rdf:type ifc:IfcRelAggregates .
:IfcRelAggregates_01 ifc:relatingObject_IfcRelDecomposes :Storey_02 .
:IfcRelAggregates_01 ifc:relatingObject_IfcRelDecomposes :Chair_03 .
```

Instead, the chair instance could be directly linked to its storey using one relationship. The additional relationships add complexity to the query syntax and leads to longer data retrieval times. Secondly, as per their schemas, the representation of specific domain knowledge are often incomplete even in the ontologies. For example, CityGML has incorporated landscaping and infrastructure concepts like vegetation and tunnels. Regardless, these concepts do not include any semantic attributes besides their geometry. Such concepts are not rich or granular enough to suffice the requirements of some applications. Lastly, these ontologies require users to possess expertise in the BIM and GIS domains, and a thorough understanding of the IFC and CityGML schemas and their workflows. But such technical expertise demands considerable time and resources that have opportunity costs and could be dedicated to other tasks. Effectively, existing ontologies are not sufficiently concise or complete for applications outside of their initial domains.

Nevertheless, certain ontologies have suggested the potential of an alternate method involving the use of modular domain ontologies. One example is the `BOT` ontology which provides a minimal representation of the topological hierarchy of IFC models, connecting sites, buildings, storeys, spaces, and elements [31]. Users can then extend the ontology with more elaborate domain or application ontologies that includes the attributes necessary for their specific use case. This approach views each domain ontology as a module that is interconnected through key concepts. Moreover, ontologies can be developed for each phase of the building life cycle to incorporate the dynamic nature of their data. For example, a construction ontology could represent material information and costs, schedules, and other information during the construction and design phase. At the operational phase, an asset management ontology could represent each asset, their date of purchase, replacement dates, and maintenance schedules. Hence, in meeting current urban needs, the proposed Semantic Web approach will have to incorporate principles of modularity, interoperability, dynamism, and scalability.

3 Methods

3.1 The World Avatar

As conveyed by its name, the "World Avatar" aims to develop an all-encompassing digital twin that can connect data and computational agents in real-time to create a living digital "avatar" of the real world, inclusive of abstract concepts and processes [2]. A digital twin is a digital representation of assets, processes, or systems in the built or natural environment that creates the opportunity for positive feedback into the physical world [9]. The World Avatar (TWA) adopts a dynamic KG, which is a KG that is continuously updated and restructured by autonomous agents. Acting on the real-time status of the physical world, these agents continuously interact with the KG for updates, analysis, decision-making, and control of real-world entities. Within a digital twin, the agents in a dynamic KG can update the digital twin's specifications and influence the real world depending on predefined objectives.

Originating from a chemical engineering perspective, TWA was applied to the decarbonisation of the chemical industry in Singapore. It is currently able to describe a number of concepts around chemistry, chemical processes, laboratories, power systems, city and environmental planning [25]. This paper presents our approach to describe, integrate, and apply asset, building, and infrastructure data from BIM, GIS, and other sources in use cases such as laboratory automation and smart cities. As these data are linked to real-world entities with physical geometries, this paper will also introduce how TWA incorporate and visualise 3D urban models from the BIM and GIS domains.

3.2 Dynamic knowledge graph approach

As illustrated in **Fig. 1**, the dynamic KG consists of three key components – domain ontologies, autonomous agents, and a visualisation interface.

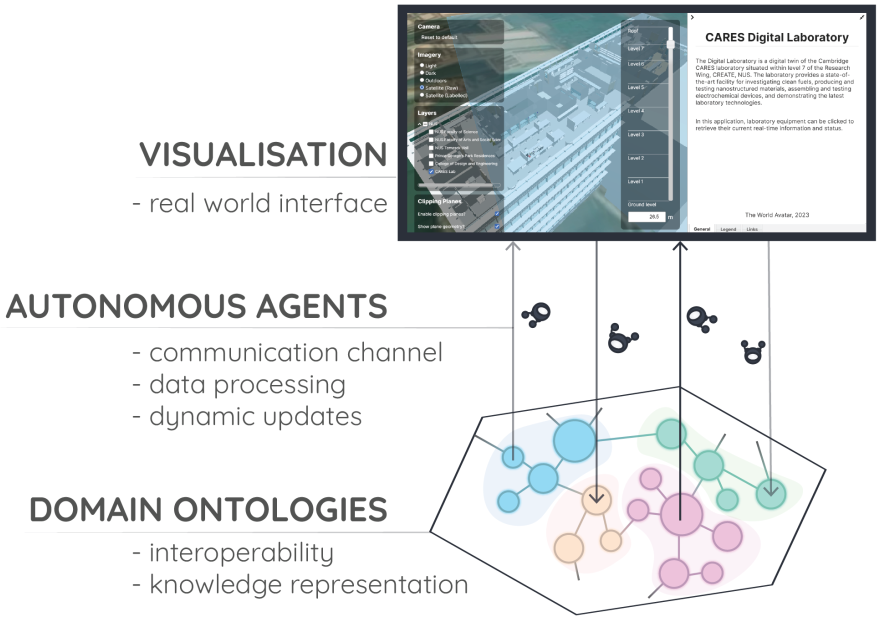


Figure 1: *The three key components and features of a dynamic knowledge graph: Domain ontologies, autonomous agents, and a visualisation interface*

3.2.1 Domain ontologies

Domain ontologies represent the concepts and relationships involved in the domain of interest. When implemented with a modular design, they can be extended to connect different domains and their knowledge. In TWA, ontologies are designed as modules specific to their domains, reusing existing concepts where possible and extending them otherwise. **Fig. 2** illustrates the modularity and interoperability of ontologies in TWA for the BIM and GIS domains, namely `OntoBuiltEnv`, `OntoBIM`, and `OntoCityGML`. The `BOT` ontology has been reused in `OntoBIM` to represent the building topology [31]. We distinguish between geometry and semantic representation as their entities may not have digital geometry representations of any kind, but will always hold functional information in the real world. For example, a building entity will always have a `bot:Building` instance with semantic relationships in the KG, but it might not have a digital representation of its geometry.

At the geometry level, `OntoBIM` and `OntoCityGML` represent the geometry concepts and relationships contained in 3D IFC and CityGML models respectively. The entity's concept is linked to their geometry concepts through the optional `hasIfcRepresentation` and `hasOntoCityGMLRepresentation` relationships. Yet, this is dependent on the availability of the IFC and/or CityGML geometric representation. Such a distinction between geometry and semantic concepts is significant for two reasons: First, it avoids any assumptions that the IFC and CityGML models are perfect geometric representation of their real-world counterparts. Second, functional information remains available in the dynamic KG, which can support other tasks that do not involve geometry.

At the semantic level, `OntoBuiltEnv` encompasses the larger scale GIS concepts and

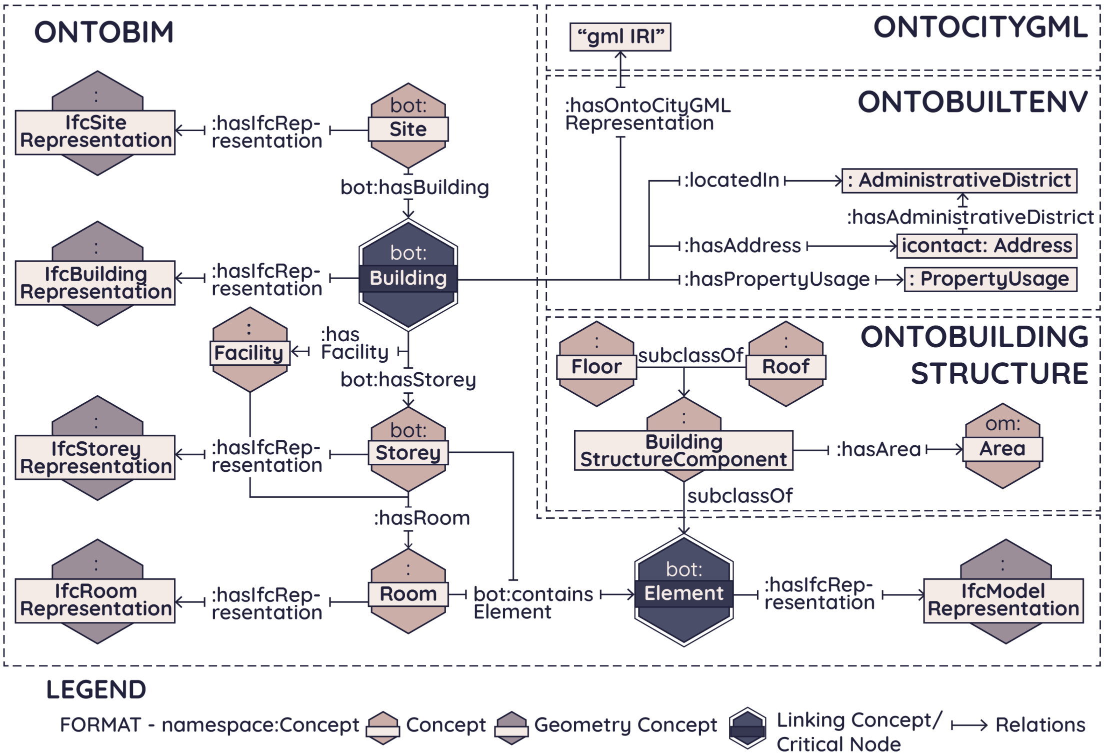


Figure 2: *An extract of the modular ontology design (TBox) connecting both BIM and GIS perspectives about buildings, by linking cross-domain knowledge through a common 'Building' or 'Element' concept. All referenced namespaces are declared in Appendix A, with ":" referring to their current domain namespace.*

relationships such as address, land use, and property type, while `OntoBIM` incorporates the building topology between sites, buildings, storeys, facilities, and rooms. In accommodating for multi-storey facilities, `OntoBIM` connects them directly to their building and room instances, instead of following the storey relationships. The critical node is the `bot:Building` concept that links `OntoBuiltEnv` and `OntoBIM` to access both ontologies' concepts and relationships. Critical nodes act as gateways to connect at least two ontologies and their knowledge together.

The next critical node is the `bot:Element` concept in `OntoBIM`, which is linked to the building topology via the `bot:containsElement` relationship to either a storey or room. Acting as a bridge to other domain ontologies, this concept can be utilised to represent any smaller scale element such as devices, sensors, walls, doors, or furniture. For example, `OntoBuildingStructure` describes the building structure components such as walls, roofs, and stairs, and their specific dimensions and properties. When their parent `BuildingStructureComponent` class is linked to the `bot:Element` concept, we are able to access their structural attributes, topology, and geospatial relationships in the ABox. Thus, the `bot:Element` concept enables the ease of extending TWA's KG to include more domain ontologies such as sensors, furniture, and even chemical reactions.

3.2.2 Autonomous agents

In TWA, autonomous agents are part of the dynamic KG, and are represented with an agent ontology [25]. Each agent is a Semantic Web service that acts upon the KG or the real world to fulfill tasks such as data retrieval, update, forecast, and even simulation. For example, some agents continuously retrieve real world data such as from sensors and mobile phones, and update their representation in the dynamic KG accordingly to enable real-time dynamism. In the dynamic KG, the agents communicate and interact with each other or the KG through HTTP requests. Ontologies support this multi-agent system by establishing shared definitions of their knowledge.

In the context of BIM-GIS integration, data instantiation agents are crucial to manage and process specific data formats such as BIM, GIS, sensors, and Excel spreadsheets into the dynamic KG. An example is the `Ifc2OntoBIMAgent` that converts IFC models into an `OntoBIM` ABox that is stored within the KG. Furthermore, the visualisation of these data on the interface requires several agents. Visualisation agents like the `Ifc2TilesetAgent` generate 3D tilesets and glTF files from the IFC and CityGML models alongside the ABox. Each element will have an associated IRI that information retrieval agents such as the `FeatureInfoAgent` can use to retrieve their real-time data and time series from the dynamic KG. Users can then interact with the visualisation interface running on the web browser to navigate the building or floor, interact with the assets, view real-time data, and more.

3.2.3 Visualisation interface

In the body of literature, the visualisation of BIM and GIS geometry data generally involves the integration of BIM data into GIS environments or GIS data into BIM environments, with the former being the prevailing pattern [1, 37, 41]. Some of these prevailing workflows process and visualise BIM and GIS data in a separate Common Data Environment (CDE), usually in a GIS-based web platform [33, 41]. An advantage of these CDEs is the reduction of upfront resources and costs when designed with intuitive user interfaces. Users do not have to spend resources to learn the different workflows and interfaces of each BIM and GIS software as well as their specific domain expertise. Coupled with other reasons, which we will expand on, we have decided to visualises both BIM and GIS data in a CDE.

In TWA, an open-source JavaScript library called CesiumJS has been incorporated into the visualisation interface, as shown on **Fig. 3**. CesiumJS is popular amongst the research community to generate 3D map views and showcase 3D geospatial data and models for cities, buildings, and assets [1, 18, 27, 33, 40]. The library can load 3D urban models in the 3D Tiles, glTF, or KML formats. Cesium 3D Tiles is an open specification based on glTF, that is designed and optimised to stream and render massive heterogeneous data sets [12]. It should be noted that glTF acts as a general container for geometry formats that can accommodate both the solid boundary representation modelling paradigm in IFC as well as the surface-based paradigm in CityGML. This circumvents the technical bottleneck of BIM-GIS integration at the geometry level [37, 41]. Furthermore, TWA's open source approach prevents vendor lock-in, especially to the commercial BIM software, and further promote the sharing and fusion of 3D data. Users are able to reproduce the workflows with similar libraries or other vendors based on their requirements.

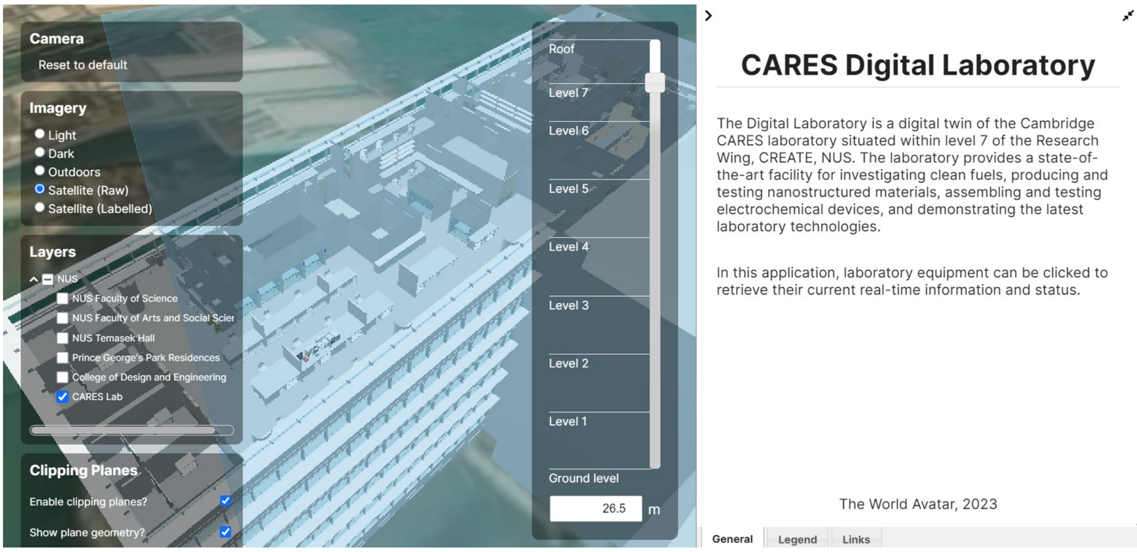


Figure 3: *An example of the World Avatar's visualisation interface running on any web browser.*

4 Applications

This section showcases TWA's capabilities to seamlessly connect BIM, GIS, and other domains across scales. At the city scale, a smart city application in Germany features the opportunities of accessing granular BIM and other information in a GIS environment. At a facility level, a digital laboratory application in Singapore highlights the opportunities of connecting disparate niche multi-scale domains such as building management systems, chemistry, experiments, and sensors, and the capacity to scale this up.

4.1 Smart city

In one instance, TWA has enabled smart city capabilities within the town of Pirmasens, Germany. Given a population of approximately 40,000 people, Pirmasens strikes a balance between being a representative size for an urban centre in Germany and still being small enough to facilitate an efficient implementation for a proof of concept. Despite previous digitisation efforts of municipal data and governance processes, the majority of data remains fragmented across multiple city administration departments and in various formats (e.g., Excel, csv, xml, and shapefiles). As a result, many of these datasets remain untapped and underutilised.

In this context, the dynamic KG is applied to demonstrate the value of integrating these diverse datasets from different domains, including the built environment and building energy. More precisely, this work features the integration of building and energy data to support the transition towards renewable energy, a priority theme on Germany's energy agenda [26]. While previous subsidies for photovoltaic installation costs have been a crucial enabler for initial photovoltaic adaption [26, 38], the curtailment of these subsidies in 2012 and 2014 has led to a significant decline in photovoltaic installations [38]. Given the rising costs of conventional energy sources and the pressing environmental concerns, other levers are required to continue this transition. One recommendation is to provide home owners with easier access to information about the photovoltaic potential of their properties and improving information transparency through a user-friendly publicly-accessible platform [26].

Presently, a web platform (`https://solarkataster.rlp.de/start`) is publicly accessible to display solar cadastre data such as potential solar energy yields of individual buildings; however, the platform lacks sufficient information to actually empower its users. For instance, it lacks complementary information about energy consumption or actionable installation instructions to estimate prospective cost savings [26]. In this regard, a seamless integration of BIM and GIS information could address this limitation: BIM can provide detailed energy consumption data, and infer the number of installable photovoltaic panels based on the dimensions provided by various vendors and the actual geometry of the roof and walls. The geolocation information in GIS facilitates the computation of theoretical solar energy yields based on local climate and weather conditions as well as the influence of surrounding buildings due to shading.

Figure 4: *Smart city representation of Pirmasens using semantic data: The GIS representations allow for planning and management on a city level, while the BIM representations (where available) enable inclusion of individual buildings' functional data within the same application.*

As illustrated in **Fig. 4**, TWA integrates the city scale GIS and detailed BIM data to create a rich smart city representation of Pirmasens. By selecting a building, users can then seamlessly identify and navigate the connections between BIM-GIS data and other semantic information for example, the energy consumption of a particular building. Notably, these data connections are extended through additional domain ontologies such as

OntoDevice and OntoUbemmp that are linked via the central bot:Element concept. While OntoDevice describes the solar devices and energy consumption measures from monitoring devices, OntoUbemmp contains the building utility and energy concepts related to consumption and production. The latter ontology, alongside the dimensions of a building's wall facades and roof, are provided as inputs to the City Energy Analyst Agent to compute the energy consumption and theoretical solar energy outputs via the City Energy Analyst simulation tool [19]. Thus, the agent, ontologies, and BIM-GIS representation connects cross-domain data to provide a holistic user interface for energy management.

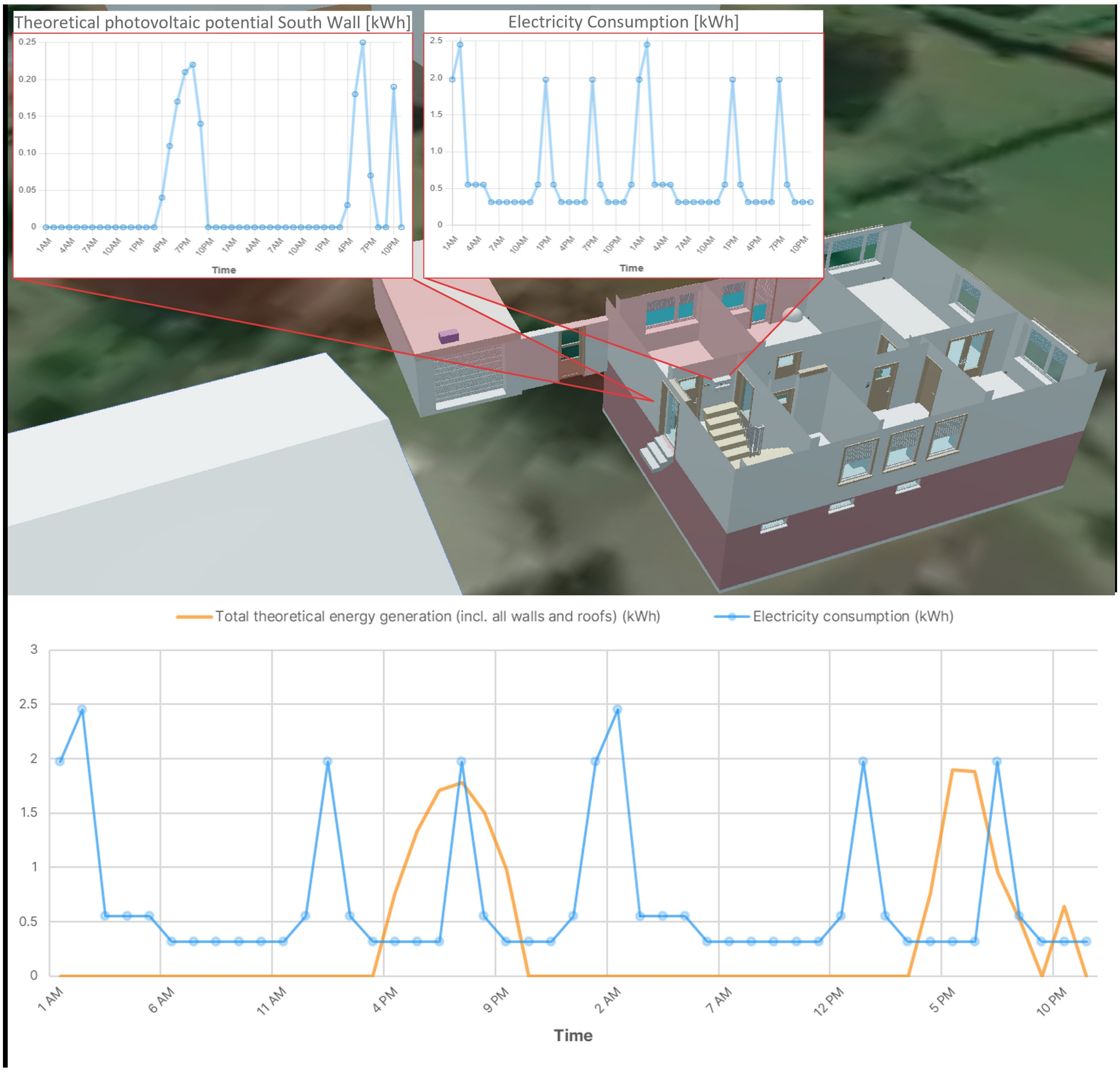


Figure 5: *Dynamic cross-domain data visualisation in the World Avatar: Comparing a residential building's actual electricity consumption (BIM data) against its total photovoltaic potential (assessed using location and proximity information, i.e., GIS data) enables cost-benefit analyses of solar panel installation.*

As shown in **Fig. 5**, users can view their energy consumption and theoretical energy yields through an unified visualisation interface. This has several advantages:

First, the integration of BIM-GIS parameters plays a key role in calibrating and enhancing the accuracy and precision of city-scale simulation models. This is crucial for the City Energy Analyst, which is a general purpose simulation tool that depends on various assumptions which differ between regions and building types. By incorporating BIM, precise building dimensions and actual energy consumption data are available. The geocoding process of GIS complements BIM data to incorporate weather conditions and evaluate the impact of surrounding buildings on solar energy reception for the simulation model. Notably, the time series representation also captures the dynamic nature of this domain, considering the influence of seasonal variations, weather patterns, and new construction developments that may impact solar energy generation rates. The consolidation of these data also acts as a plausibility check, validating computational models against real-world values. Consequently, these inputs can support simulation tools to calculate precise and reliable solar energy estimates that are extendable to other contexts.
Second, accessible energy assessments consolidated in a single user-friendly interface enhances the information transparency of energy savings estimates to the public. Referencing the visualisation presented in **Fig. 5**, users can easily relate their current energy consumption to the theoretical solar energy generation and costs to estimate their energy savings.
Last, upscaling these detailed building level analyses to district and city scale, can help to derive more accurate aggregates as evidence to support energy policy decisions.

4.2 Laboratory automation

In supporting laboratory automation, TWA has been utilised to create a digital twin of a chemistry laboratory in the Cambridge Centre for Advanced Research and Education in Singapore (CARES) at the National University of Singapore's CREATE campus. There has been growing interest in the automation of research activities to accelerate scientific discoveries in the face of looming sustainability threats [23]. However, existing platforms have been a hindrance to these goals. Current scientific technologies generally offer isolated perspectives on chemistry and experiment design, but may neglect to consider the influence of humans supervising and executing experiments as well as available infrastructure and related costs [5].

For example, the role of a laboratory manager has become somewhat complex due to the poor interoperability of these platforms. In coordination with a dedicated facility manager on building-level operations, this personnel has to ensure the availability, compliance, maintenance, and safety of laboratory equipment and chemicals, while managing the massive energy consumption of laboratories [22]. In the present state, they often have to access multiple user interfaces and comprehend different software workflows to grasp and apply the cross-domain knowledge to their tasks. Namely, such a role requires knowledge on asset management, BIM, building management systems, chemistry related to ongoing experiments, paper-based floor plans, and sensors. Moreover, in a typical research campus housing multiple laboratories, coordinating data from these different laboratories can be complex due to the poor interoperability and scale, which can interfere with the laboratory manager's duties [5]. BIM-GIS integration is anticipated to offer new opportunities for comprehensive benchmark analyses and propose recommendations for modifying laboratory designs, equipment, and layouts to support laboratory managers. Specifically, BIM provides granular information on energy consumption and other specifications for each asset, which can be linked and aggregated to a facility, building, and even campus level using the precise location provided by GIS. When coupled with the equipment and environmental data supplemented by existing building management systems and the Internet of Things, laboratory managers can then formulate best practices and benchmarks for laboratory equipment, layouts, and designs. This enables them to develop effective energy management strategies and promote sustainable practices not only at the individual laboratory level but across the entire research campus.

One instance of the ongoing TWA work to augment and automate some aspects of the laboratory manager's role can be seen in **Fig. 6**. A single accessible user interface displaying cross-domain data enables laboratory managers to monitor various facility operations and device conditions. The representation of these cross-domain data in the dynamic KG have been integrated through various ontologies such as `OntoDevice`, `OntoBMS`, `OntoLab`, and `OntoSpecies`. `OntoDevice` describes the device concepts and their generic properties and measurements, which can be extended to represent specific device domains in niche ontologies such as the representation of building management systems via `OntoBMS`. `OntoLab` describes the concepts within the laboratory, inclusive of their equipment and experiments, while `OntoSpecies` describes the chemicals and their properties within the chemistry domain. These ontologies are linked directly to the `bot:Element` in `OntoBIM`, thus, enabling cross-domain linkage.

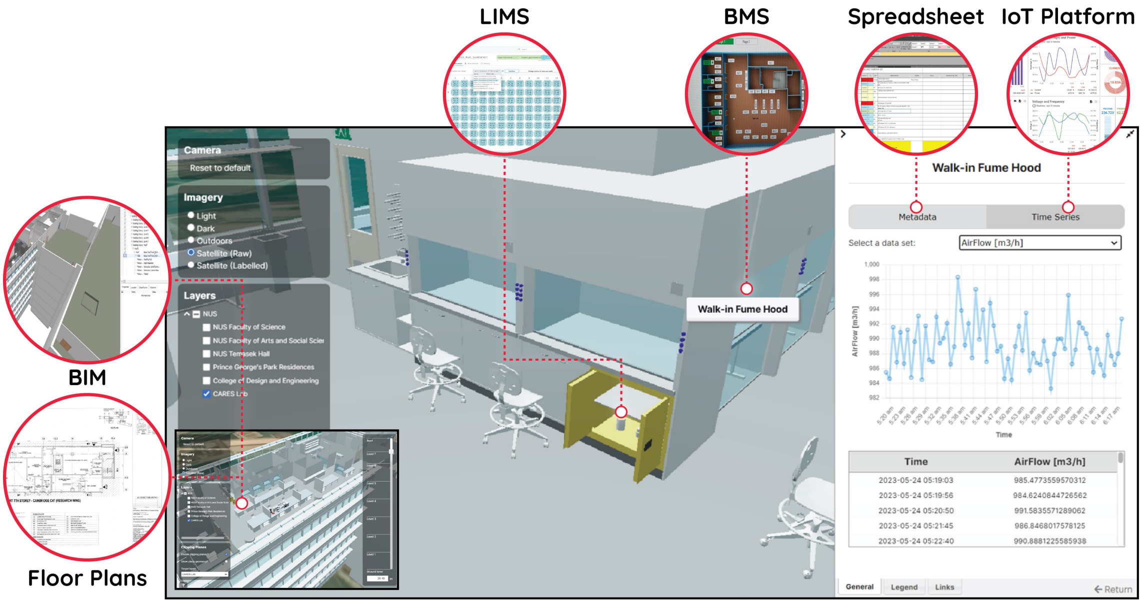


Figure 6: *Dynamic cross-domain data visualisations in the World Avatar: Monitoring facility operations and device conditions on multiple scales through a general visualisation interface replacing a variety of previous "island solutions" (in red circles).*

Despite the diversity in domains and technologies, the dynamic KG seamlessly integrates a variety of heterogeneous data formats and systems for a rich digital representation of the laboratory. This digital laboratory does not merely advance user interactions with cross-domain data, but can influence the real world through an automated control system. Specifically, TWA is able to directly send signals to initiate device state changes based on predefined thresholds. In cases where remote control is not feasible, the system can generate email notifications for manual intervention by the laboratory manager. This capability is made possible due to the availability of granular and dynamic data from the Internet of Things and building management systems. Leveraging on these data enables changes of individual assets to be monitored over time, regardless of their domains. In addition, the future status of these assets can be predicted through the time series inputs and `Forecasting Agent` to achieve predictive maintenance. Hence, TWA has achieved a higher level of automation for laboratories that can augment the laboratory manager's duties. Given the integration with GIS, this work can also be scaled up to the campus level to support sustainable laboratory practices and accelerate scientific discoveries.

5 Conclusion

The current state of digital ecosystems in cities is characterised by a variety of isolated data silos and poor interoperability. These interoperability challenges hinder cross-domain collaboration to manage the growing complexities of urban environments. Given the urgency in breaking down these data silos, this paper proposes the use of Semantic Web technologies such as knowledge graphs as a potential solution. By semantically annotating data and their relationships in a standardised machine-readable format, these technologies support knowledge discovery, integration, transfer, derivation, and forecasting processes across different scales. Specifically, we present an implementation of a dynamic knowledge graph approach using domain ontologies, autonomous agents, and a visualisation interface. The integration of cross-domain, dynamic, and multi-scale data from the BIM, GIS, building management and further domains is demonstrated to foster interoperability for two applications, namely laboratory automation and smart cities.

The distinguishing features of dynamic knowledge graphs lie in their dynamism, interoperability, modularity, scalability, and transparency. Semantic Web technologies revolve around open standards and protocols that are publicly accessible and extensible on the Internet. Given the availability of ontologies in the public domain, domain experts can be consulted at any stage to establish a consistent, aligned definition of concepts and schemas. This process addresses data ambiguities and accommodates user requirements across stakeholders from individuals, local communities, organisations, and the government. In addition, the scalability of dynamic knowledge graphs is derived from their ease of extension to different domains and scales. Domain ontologies can be directly linked through critical nodes, enabling new data to be extended from these nodes. As a result, the applications presented have connected diverse domains and scales such as chemistry, energy, laboratory, buildings, and cities. This differs from non-Semantic Web workflows, which often require significant modifications to their database schemas and data processing processes to integrate new data. Moreover, the approach showcased in this paper is flexible and distributable, as evident by the presented applications across diverse geographic locations in Germany and Singapore. The use of containerisation technologies to standardise deployment workflows has enabled the flexibility of this technology for both local servers and cloud environments. By leveraging existing ontologies and workflows, integration barriers are also lowered, granting users access to a wider range of private and public data sources and technologies. This is especially relevant for smaller and less influential players such as municipal and local governments with limited resources.

Acknowledgements

This research was supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. M. Hofmeister acknowledges financial support provided by the Cambridge Trust and CMCL. M. Kraft gratefully acknowledges the support of the Alexander von Humboldt Foundation.

A part of this study has been undertaken in the context of DOME 4.0 project, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 953163.

Nomenclature

3D Three-Dimensional

ABox Assertional Component (of an ontology)

BIM Building Information Model

CARES Cambridge Centre for Advanced Research and Education in Singapore

CDE Common Data Environment

CREATE Campus for Research Excellence and Technological Enterprise

GIS Geographic Information System

IFC Industry Foundation Classes

IRI Internationalised Resource Identifier

KG Knowledge Graph

RDF Resource Description Framework

TBox Terminological Component (of an ontology)

TWA 'The World Avatar' (project)

A Namespace definitions

bot: <https://w3id.org/bot#>

icontact: <http://ontology.eil.utoronto.ca/icontact.owl#>

om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>

ontobim: <https://www.theworldavatar.com/kg/ontobim/>

ontobuildingstructure: <https://www.theworldavatar.com/kg/ontobuildingstructure/>

ontobuiltenv: <https://www.theworldavatar.com/kg/ontobuiltenv/>

ontocitygml: <https://www.theworldavatar.com/kg/ontocitygml/>

References

[1] F. Abastante, I. M. Lami, and P. Lombardi. An integrated participative spatial decision support system for smart energy urban scenarios: A financial and economic approach. *Buildings*, 7(4), 2017. 10.3390/buildings7040103.

[2] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal digital twin-a dynamic knowledge graph. *Data-Centric Engineering*, 2:e14, 2021. 10.1017/dce.2021.10.

[3] G. Antoniou, E. Franconi, and F. van Harmelen. Introduction to semantic web ontology languages. In N. Eisinger and J. Małuszyński, editors, *Reasoning Web: First International Summer School 2005, Msida, Malta, July 25-29, 2005, Revised Lectures*, volume 3564 of *Lecture Notes in Computer Science*, pages 1–21. Springer Berlin Heidelberg, 2005. 10.1007/11526988_1.

[4] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787–2805, 2010. 10.1016/j.comnet.2010.05.010.

[5] J. Bai, L. Cao, S. Mosbach, J. Akroyd, A. A. Lapkin, and M. Kraft. From Platform to Knowledge Graph: Evolution of Laboratory Automation. *JACS Au*, 2(2):292–309, 2022. 10.1021/jacsau.1c00438.

[6] T. Berners-Lee. Linked data, 2006. URL `http://www.w3.org/DesignIssues/LinkedData.html`. Accessed June 2023.

[7] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific american*, 284(5):34–43, 2001. 10.1038/scientificamerican0501-34.

[8] C. Boje, A. Guerriero, S. Kubicki, and Y. Rezgui. Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114:103179, 2020. 10.1016/j.autcon.2020.103179.

[9] A. Bolton, L. Butler, I. Dabson, M. Enzer, M. Evans, T. Fenemore, F. Harradence, E. Keaney, A. Kemp, A. Luck, N. Pawsey, S. Saville, J. Schooling, M. Sharp, T. Smith, J. Tennison, J. Whyte, A. Wilson, and C. Makri. Gemini principles, 2018. doi:10.17863/CAM.32260.

[10] H. L. Boschken. Scale, the silo effect and intergovernmental cooperation: Institutional analysis of global cities and ecological sustainability. American Political Science Association 2013 Annual Meeting, 2013. Available at SSRN: `https://ssrn.com/abstract=2300687`.

[11] G. Celeste, M. Lazoi, M. Mangia, and G. Mangialardi. Innovating the construction life cycle through BIM/GIS integration: A review. *Sustainability*, 14(2):766, 2022. 10.3390/su14020766.

[12] Cesium. 3d tiles next, 2023. URL `https://github.com/CesiumGS/3d-tiles/tree/main/next`. Accessed June 2023.

[13] A. Chadzynski, N. Krdzavac, F. Farazi, M. Q. Lim, S. Li, A. Grisiute, P. Herthogs, A. von Richthofen, S. Cairns, and M. Kraft. Semantic 3D city database—an enabler for a dynamic geospatial knowledge graph. *Energy and AI*, 6:100106, 2021. 10.1016/j.egyai.2021.100106.

[14] S. Chung, C.-S. Cho, J. Song, K. Lee, S. Lee, and S. Kwon. Smart facility management system based on open BIM and augmented reality technology. *Applied Sciences*, 11(21):10283, 2021. 10.3390/app112110283.

[15] A. de Waal, M. Weaver, T. Day, and B. van der Heijden. Silo-busting: Overcoming the greatest threat to organizational performance. *Sustainability*, 11(23):6860, 2019. 10.3390/su11236860.

[16] Y. Deng, J. C. Cheng, and C. Anumba. Mapping between BIM and 3D GIS in different levels of detail using schema mediation and instance comparison. *Automation in Construction*, 67:1–21, 2016. 10.1016/j.autcon.2016.03.006.

[17] F. M. Dinis, J. Poças Martins, A. S. Guimarães, and B. Rangel. BIM and semantic enrichment methods and applications: a review of recent developments. *Archives of Computational Methods in Engineering*, 29(2):879–895, 2022.

[18] U. Eicker, V. Weiler, J. Schumacher, and R. Braun. On the design of an urban data and modeling platform and its application to urban district analyses. *Energy and Buildings*, 217:109954, 2020. 10.1016/j.enbuild.2020.109954.

[19] J. A. Fonseca, T.-A. Nguyen, A. Schlueter, and F. Marechal. City energy analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts. *Energy and Buildings*, 113:202–226, 2016. 10.1016/j.enbuild.2015.11.055.

[20] G. Gröger and L. Plümer. CityGML – interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71:12–33, 2012. 10.1016/j.isprsjprs.2012.04.004.

[21] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. Ngonga Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, and A. Zimmermann. Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4):1–37, 2021. 10.1145/3447772.

[22] M. D. Kaplowitz, L. Thorp, K. Coleman, and F. Kwame Yeboah. Energy conservation attitudes, knowledge, and behaviors in science laboratories. *Energy Policy*, 50: 581–591, 2012. ISSN 03014215. 10.1016/j.enpol.2012.07.060.

[23] H. Kitano. Nobel Turing Challenge: creating the engine for scientific discovery. *npj Systems Biology and Applications*, 7(1):1–12, 2021. 10.1038/s41540-021-00189-3.

[24] R. Kitchin. The real-time city? Big data and smart urbanism. *GeoJournal*, 79:1–14, 2014. 10.1007/s10708-013-9516-8.

[25] M. Q. Lim, X. Wang, O. Inderwildi, and M. Kraft. The World Avatar – a world model for facilitating interoperability. In *Intelligent Decarbonisation: Can Artificial Intelligence and Cyber-Physical Systems Help Achieve Climate Mitigation Targets?*, pages 39–53. Springer International Publishing, 2022. 10.1007/978-3-030-86215-2_4.

[26] Y. Lu, Z. A. Khan, M. S. Alvarez-Alvarado, Y. Zhang, Z. Huang, and M. Imran. A critical review of sustainable energy policies for the promotion of renewable energy sources. *Sustainability*, 12(12):5078, 2020. 10.3390/su12125078.

[27] N. Moretti, C. Ellul, F. Re Cecconi, N. Papapesios, and M. C. Dejaco. GeoBIM for built environment condition assessment supporting asset management decision making. *Automation in Construction*, 130:103859, 2021. 10.1016/j.autcon.2021.103859.

[28] J. Opara-Martins, R. Sahandi, and F. Tian. Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. *Journal of Cloud Computing*, 5:1–18, 2016. 10.1186/s13677-016-0054-z.

[29] P. Pauwels, S. Zhang, and Y.-C. Lee. Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, 73:145–165, 2017. 10.1016/j.autcon.2016.10.003.

[30] H. Y. Quek, F. Sielker, J. Akroyd, A. N. Bhave, A. von Richthofen, P. Herthogs, C. van der Laag Yamu, L. Wan, T. Nochta, G. Burgess, M. Q. Lim, S. Mosbach, and M. Kraft. The conundrum in smart city governance: Interoperability and compatibility in an ever-growing ecosystem of digital twins. *Data & Policy*, 5:e6, 2023. 10.1017/dap.2023.1.

[31] M. H. Rasmussen, P. Pauwels, M. Lefrançois, and G. F. Schneider. Building topology ontology, 2021. URL `https://w3c-lbd-cg.github.io/bot/`. Accessed June 2023.

[32] J. Robert, S. Kubler, N. Kolbe, A. Cerioni, E. Gastaud, and K. Främling. Open IoT ecosystem for enhanced interoperability in smart cities—example of Métropole De Lyon. *Sensors*, 17(12):2849, 2017. 10.3390/s17122849.

[33] D. Shkundalov and T. Vilutienė. Bibliometric analysis of building information modeling, geographic information systems and web environment integration. *Automation in Construction*, 128:103757, 2021. 10.1016/j.autcon.2021.103757.

[34] Y. Song, X. Wang, Y. Tan, P. Wu, M. Sutrisna, J. C. Cheng, and K. Hampson. Trends and opportunities of BIM-GIS integration in the architecture, engineering and construction industry: a review from a spatio-temporal statistical perspective. *ISPRS International Journal of Geo-Information*, 6(12):397, 2017. 10.3390/ijgi6120397.

[35] A. M. Townsend. *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company, New York, 2013.

[36] W3C. Semantic web, 2015. URL `https://www.w3.org/standards/semanticweb/`. Accessed June 2023.

[37] H. Wang, Y. Pan, and X. Luo. Integration of BIM and GIS in sustainable built environment: A review and bibliometric analysis. *Automation in Construction*, 103: 41–52, 2019. 10.1016/j.autcon.2019.03.005.

[38] D. Wen, W. Gao, F. Qian, Q. Gu, and J. Ren. Development of solar photovoltaic industry and market in China, Germany, Japan and the United States of America using incentive policies. *Energy Exploration & Exploitation*, 39(5):1429–1456, 2021. 10.1177/0144598720979256.

[39] F. Xue, L. Wu, and W. Lu. Semantic enrichment of building and city information models: A ten-year review. *Advanced Engineering Informatics*, 47:101245, 2021. 10.1016/j.aei.2020.101245.

[40] S. Zhang, D. Hou, C. Wang, F. Pan, and L. Yan. Integrating and managing BIM in 3D web-based GIS for hydraulic and hydropower engineering projects. *Automation in Construction*, 112(January):103114, 2020. 10.1016/j.autcon.2020.103114.

[41] J. Zhu and P. Wu. BIM/GIS data integration from the perspective of information flow. *Automation in Construction*, 136:104166, 2022. 10.1016/j.autcon.2022.104166.