Linked Urban Data:

Leveraging the capabilities of urban data analytics for urban data analytics

Xi Peng[1], Diellza Elshani[2], Hana Svatoš-Ražnjević[3], Achim Menges[4], Thomas Wortmann[5]
[1]Institute for Computational Design and Construction (ICD), University of Stuttgart|Achitecture and Sustainable Design (ASD) Pillar - Singapore University of Technology & Design
[2,5]Institute for Computational Design and Construction, Department for Computing in Architecture (ICD/CA), University of Stuttgart.
[3,4]Institute for Computational Design and Construction (ICD), University of Stuttgart
[1]1009895@mymail.sutd.edu.sg
[2,3,4,5]{diellza.elshani|hana.svatos-raznjevic|achim.menge|thomas.wortman}@icd.uni-stuttgart.de

Urban planning often suffers from inefficiencies due to the fragmented static and dynamic datasets. This study proposes a Linked Urban Data Framework that leverages Knowledge Graphs (KGs) to integrate analysis of morphological and mobility data. By incorporating established ontologies - OpenStreetMap Ontology with the newly developed Urban Dynamic Data Ontology (UDO)- the framework facilitates cross-domain querying and semantic reasoning. A case study in Stuttgart demonstrates the framework's capabilities through SPARQL-based analysis across heterogeneous datasets, including UTD 19 traffic data and VVS public transport records, and OpenStreetMap building and road data. Additionally, a prototype web-based interface is developed to support interative exploration of the integrated urban data for diverse stakeholders. The finding highlight the framework's potential to enhance data interoperability, accessibility,transparency, and decision-making process in urban planning, particularly in contexts requiring multisource, multi-scale spatial data integration and complex knowledge-based assessments.

Keywords: *Urban Data Modelling, Linked static and dynamic data, Knowledge Graphs, Ontology, Semantic Web Technologies.*

INTRODUCTION

Urban planning has increasingly embraced data-driven approaches, leveraging diverse datasets to enhance design and decision-making. Data infrastructure is a key factor in the development of smart cities. Traditional methods primarily rely on predefined spatial data, making them incapable of real-time responsiveness to changes in urban mobility and infrastructure usage. (Romualdo-Suzuki and Finkelstein, 2020) As a result, critical urban design decisions—such as site selection, infrastructure planning, and transportation optimization—often depend on incomplete or unlinked datasets, reducing the accuracy and efficiency of planning processes. (Šoštarić et al., 2021)

To address these challenges, this study proposes a Linked Urban Data Framework based on Knowledge Graphs (KGs) to integrate and structure heterogeneous urban data. Built upon existing

ontologies such as OpenStreetMap Ontology and CityGML, this framework develops an Urban Dynamic Data Ontology (UDO) to bridge the gap between static and dynamic datasets. "Semantic 3D city inference provides a novel approach to support urban planning and management."(Chadzynski et al., 2023) By leveraging the semantic linking capabilities of knowledge graphs, the proposed method enables querying and cross-domain reasoning, overcoming the limitations of traditional database models.

The methodology follows a three-phase approach: (1.) Data Preparation and Ontology Development, (2.) Knowledge Graph Construction and Integration, and (3.) Framework Validation and Application Development—as illustrated in Figure 1. The framework is validated through a case study, incorporating datasets from OpenStreetMap, UTD19 traffic records (a publicly available dataset containing historical traffic detector data for the Stuttgart region), and Stuttgart's VVS public transport system. (Verkehrs- und Tarifverbund Stuttgart, the city's official transit authority providing schedules and stop data for bus, tram, and subway services). Through SPARQL-based queries, the study explores practical urban planning applications, such as identifying optimal park locations, assessing public transport efficiency, and determining real estate development sites.

By integrating ontology modeling, and knowledge graphs, this research proposes a solution to mitigate urban data fragmentation and provides a flexible, extensible framework

To enhance urban data analytics and planning decisions. The remainder of this paper explores the state of the art in urban knowledge representation, details the methodology for ontology development and data integration, presents the case study, discusses findings and limitations, and concludes with insights on the potential applications of knowledge graphs in urban planning.

BACKGROUND

Traditional urban data management segregates static and dynamic datasets, restricting interoperability and comprehensive analysis (Bellini et al., 2014). This disconnect hinders comprehensive

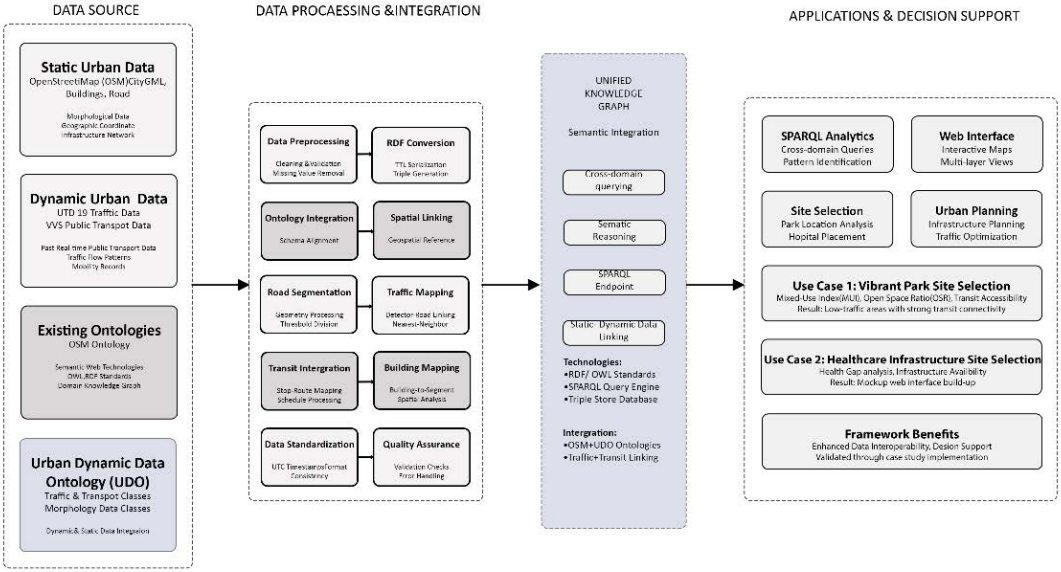


Figure 1: Framework of Linked Urban Data

urban analysis and adaptive decision-making.(Rantanen et al., 2023)

While ontologies such as CityGML and OpenStreetMap provide structured representations of static urban elements, they lack integration with real-time mobility data.

Integrating Semantic Web technologies with federated data schemas can enhance co-design processes in the building industry. Elshani et al., 2023) Knowledge graphs (KGs) offer a solution by enabling semantic linking across heterogeneous datasets. (Huang and Harrie, 2019)

However, existing urban KGs are often domain-specific rather than adopting unified approaches. (Sobral, Galvao and Borges, 2019)This study extends previous work by incorporating urban dynamics into a unified framework.

KNOWLEDGE GRAPH TECHNOLOGIES

In the field of computer science, ontologies provide a formalized representation of domain-specific knowledge by defining key concepts and their interrelationships.(Husáková and Bureš, 2020) A crucial aspect of ontology design is taxonomy, which classifies objects within a domain, enabling hierarchical structuring and contextual associations. These structured relationships enhance semantic interoperability, allowing diverse datasets to be meaningfully linked. (Elshani, Wortmann and Staab, 2022)

One of the primary applications of ontologies is in the development of knowledge graphs, which consist of ontologies—providing a reusable schema—and assertions, representing instance-specific data. By integrating static and dynamic urban datasets, they help urban planners identify relationships between land use, transportation, and infrastructure in a data-driven manner.

METHODS

This research follows a three-phase methodology:

Phase 1: Data Preparation and Ontology Development

In the initial phase of the project, we focused on developing a structured approach to integrate heterogeneous urban datasets. The primary goal was to establish a framework that enables seamless cross-platform data retrieval and querying in urban planning. This approach involved several key components, including background research, the study of semantic web technologies, data extraction techniques, knowledge graph generation methodologies, and the creation of an urban data framework.

The datasets collected for this study are categorized into static and dynamic urban data. Static data include road networks and building attributes derived from OpenStreetMap (OSM), such as road type, number of lanes, sidewalk presence, building height, and address information(see Table 1 for selected attributes). Dynamic data consist of two main components:

1. Traffic detector data from the UTD19 Stuttgart dataset, which provides information on traffic flow at specific locations, including detector ID (decID), timestamp (UTC time), flow volume, and average speed.
2. Public transportation records from the Verkehrs- und Tarifverbund Stuttgart (VVS) system, comprising multiple GTFS-derived files (e.g., stops.csv.gz, stop_times.csv.gz, trips.csv.gz). These contain stop coordinates, stop sequences, route IDs, trip schedules, and mode-specific identifiers (e.g., bus, tram, subway).

These data were collected and transformed into RDF format to support integration into the knowledge graph. The combined dataset enables multi-source spatial-temporal analysis for urban planning applications

To ensure efficient data connectivity, it was crucial to maintain data clarity and organization. This required an analysis of existing urban ontologies to verify that the data schema aligned with the specific datasets incorporated in the project. The process involved the collection of both static and dynamic urban data, forming the basis for constructing domain-specific ontology tailored to the project's needs.

Table 1: OpenStreetMap Selected Road Data Attribute Table

Category	Attribute	Data Type
Highway	max speed	xsd:integer
Highway	cycleway:both	xsd:boolean
Highway	lanes	Xsd:integer
Highway	zone:traffic	xsd:string
Highway	priority road	xsd:string
Highway	highway	xsd:string
Highway	sidewalk	xsd:string
Building	building: level	xsd:string
Building	addr:housenumber	xsd:string
Building	roof:level	xsd:string
Building	building	xsd:string
Building	Height	xsd:string
Building	addr: postcode	xsd:string
Building	Addr: city	xsd:string
Building	Addr: street	xsd:string

Ontology Extension and Formalization

Managing and processing dynamic urban data presents a significant challenge due to its complexity and variability. To address this issue, we developed the UDO, which provides a structured framework for classifying and linking dynamic urban data. This ensures geospatial query support while maintaining semantic consistency across the knowledge graph through standardized TTL serialization.

This ontology offers a clear and systematic representation, making data analysis more intuitive and facilitating better comprehension of the relationships between different data components.

UDO is structured into six primary categories, with two categories specifically related to traffic volume data and four categories representing historical public transportation data. This classification ensures a well-organized data structure, enabling seamless integration and analysis of urban mobility patterns.

Within the traffic volume data section, the "Road" entity plays a crucial role in representing location-based information. It is closely linked to the 'Measuring_Point' entity, which contains essential attributes such as geographic coordinates, a unique identifier (decID), and traffic flow data (flow). This structured linkage enhances data accuracy and consistency, allowing for precise tracking of urban traffic dynamics. Additionally, temporal attributes, including "date" and "interval", further enrich the dataset by capturing when and how frequently measurements are recorded, supporting time-series analysis and traffic trend forecasting.

Phase 2: Knowledge Graph Construction and Integration

designed a three-step approach:

1. **Data Parsing and Ontology Integration**: Converting raw data into RDF triples and serializing them into Turtle (TTL) format, aligning with established ontologies, and ensuring consistency.
2. **Linking Static and Dynamic Data**: Connecting datasets through geospatial references and semantic linkages for cross-domain integration.

Phase 3: Framework Validation and Application Development

By following this structured methodology, we developed a framework that integrates heterogeneous urban datasets into a unified knowledge graph. This facilitates complex data queries, enhances interoperability, and supports informed decision-making in urban planning and development.

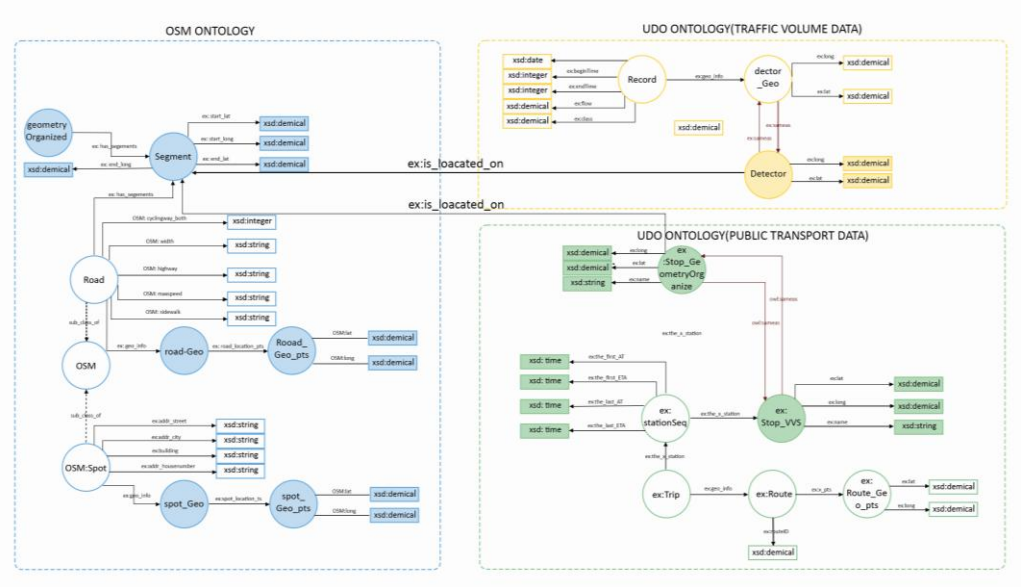


Figure 2:
OSM and UDO extension Ontologies

Conversion of Morphological Data into TTL format

OSM and morphological data transformation follows a structured pipeline: file initialization with gzip compression and namespace declarations, entity definition for 'osm', 'road', and 'spot' classes, and iterative processing of road ('df_road') and landmark ('df_spot') datasets. Each entity includes attributes such as name, type, length, and geographic coordinates, with missing values removed during preprocessing.

Geographic information processing extracts latitude and longitude coordinates from location fields, converting them into RDF triples with explicit spatial relationships. This ensures geospatial query support while maintaining semantic consistency across the knowledge graph through standardized TTL serialization. This combine ontologiey structure integrating OSM and UDO is illustrated in Figure 2.

Transformation of Dynamic Data into RDF Format

Traffic and public transportation data transformation involves extraction, preprocessing, and RDF conversion. Traffic detector data from *detector_stuttgart.csv* creates a mapping dictionary ('detector_dict') linking detector IDs to coordinates. Traffic flow data undergoes preprocessing to remove invalid values, normalize UTC timestamps, and convert numerical fields before Turtle (TTL) serialization.

Public transportation data extraction processes six compressed files (*curveName.csv.gz*, *stops.csv.gz*, *stopTime.csv.gz*, etc.) into two dictionaries: 'dict_stops' (stop attributes and locations) and 'dict_trip' (trip IDs, routes, sequences).

RDF triple generation uses Python f-strings to create structured representations where 'ex:stopID_{stop}' represents unique identifiers and 'a ex:Stop_VVS' assigns entities to the 'Stop_VVS' class for public transport stops. Bus routes structure as RDF triples containing Trip ID, Route ID, Stop

Sequence, Arrival/Departure Times, and Geographic Coordinates through systematic matching between 'dict_stops' and 'dict_trip' dictionaries. Final outputs serialize to *utd19_stuttgart.ttl.gz* (traffic) and *busttl.ttl.gz* (transit), with validation ensuring ontology compliance and cross-domain interoperability.

Knowledge Graph: Linking

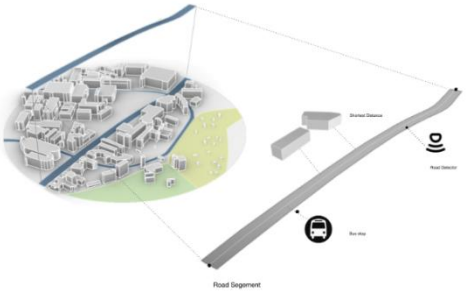


Figure 3: Method of Integrating Morphology, Historical Public Transport, and Traffic Volume Data – Road Segmentation and Bus Stop Mapping

Heterogeneous Datasets

This study employs a spatial integration approach to unify road networks, traffic flow, and public transportation data. The data processing workflow includes road segmentation, traffic detector mapping, public transportation stop matching, building-to-road spatial mapping, and data standardization, ensuring interoperability across different data sources(see Figure 3).

Road segmentation treats roads as segmented entities rather than continuous structures, extracting road geometry information including road ID, start and end coordinates, road length, and road type. A segmentation threshold divides longer roads into smaller segments while inheriting the original road ID and recalculating coordinates. Segmented roads are stored in 'Segment' entities and managed through 'geometryOrganized,' retaining attributes such as start and end latitude/longitude, road type, and metadata for analysis and matching.

Traffic detector integration maps geographic coordinates to corresponding road segments using a nearest-neighbor algorithm computing Euclidean distances between detectors and road segments.

Mapping relationships are stored in 'detector_Geo' entities, while traffic flow data (timestamps, volume, average speed) link to detectors and store in 'Record' entities associated with specific road segments. Public transportation integration processes bus routes, stop information, and trip schedules. Bus stops ('ex:Stop_VVS') link to 'ex:Stop_GeometryOrganize' for spatial consistency and map to road segments using nearest-neighbor approaches. Route data stores in 'ex:Route' and 'ex:Route_Geo_pts' entities, while trip schedules manage within 'ex:Trip' entities with chronological stop sequences through 'ex:stationSeq' entities.

Data standardization ensures consistency: temporal data converts to UTC 24-hour format ('xsd:time'), geographic coordinates conform to WGS84 standard, numerical fields (traffic volume, speed) store as 'xsd:decimal' for precision, and textual fields (stop names, road types) store as 'xsd:string' to facilitate querying. Building-to-road spatial mapping uses nearest-neighbor analysis, storing relationships in 'building_to_segment' tables.

Data-Driven Site Selection Based on OSM-UDO Ontology

As a case study to validate the OSM-UDO framework, we conducted data-driven site selection analysis across multiple urban planning scenarios in Stuttgart. Two distinct use cases were implemented to demonstrate the framework's versatility: (1.) identifying optimal locations for a vibrant public park, (2.) strategic hospital site selection. Each use case employed SPARQL-based queries to integrate

Figure 4:
Shows SPARQL queries for potential site locations in Stuttgart, depicting traffic volume and public transport accessibility based on proximity to road sensors and transit stops

static and dynamic urban data, enabling spatial analysis and decision support.

Use Case 1: Vibrant Park Site Selection

The first case focused on identifying a vibrant park location capable of accommodating 200 people while ensuring efficient traffic flow and seamless accessibility. The analysis integrated static and dynamic urban data through OSM-UDO ontology using three core selection criteria: Vibrancy, Accessibility, and Spatial Suitability.

Vibrancy was assessed using the Mixed-Use Index (MUI), which measures land-use diversity within a given area. Accessibility was evaluated based on public transportation data from OSM-UDO, calculating the distance to bus, subway, and tram stops, along with transit frequency analysis to assess service quality. Spatial Suitability was determined through the Open Space Ratio (OSR), ensuring that the selected park location provides sufficient open space for public activities.

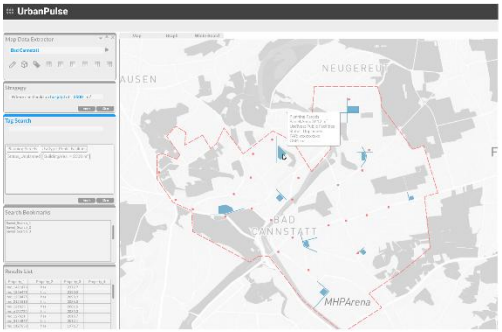
The framework systematically examined traffic patterns, land availability, and spatial utilization through SPARQL queries that identified low-traffic areas within a 600-meter radius, filtering potential locations(see Figure 4). Semantic inference was applied to assess building functions, road types, and land use, followed by the calculation of a public transit accessibility index to ensure well-connected park sites. The analysis successfully identified several locations on Stuttgart's outskirts that best met the selection criteria, offering low traffic congestion, abundant open space, and strong connectivity to transit networks, making them highly viable for development.

Use Case 2: Healthcare Infrastructure Site Selection

The framework was applied to hospital site selection to meet growing healthcare demands, prioritizing areas with limited healthcare resources for equitable service distribution. Evaluation criteria included proximity to existing medical facilities (preventing redundant infrastructure while optimizing accessibility), land use classification, infrastructure availability, and expansion potential.

An interactive web application prototype simulated hospital site selection decisions through land suitability analysis, dual-view location comparison, and examination of infrastructure layers (existing hospitals, road networks, accessibility indicators). The prototype demonstrated the framework's potential for balancing healthcare accessibility with resource optimization while

integrating complex spatial and infrastructure data for planning decisions.



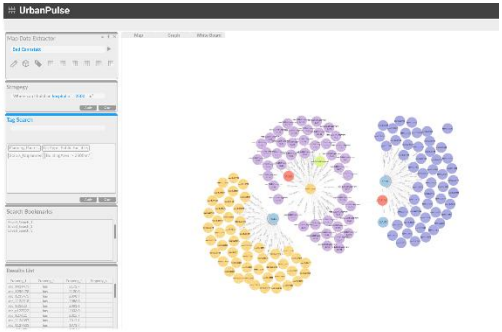


Figure 5:
Mock up web interface Left: Geospatial Query Visualization Right: Knowledge Graph Data Network

Web Application Prototype Performance

Healthcare web application prototypes demonstrated the practical implementation of OSM-UDO framework results(see figure 5). The prototypes validated the framework's capability to:

- Integrate real-time static and dynamic urban data.
- Provide interactive visualization interfaces for complex spatial analysis.
- Support comparative analysis across multiple potential sites.
- Enable examination of multiple infrastructure layers simultaneously
- Facilitate data-driven decision-making through intuitive user interfaces.

DISCUSSION AND OUTLOOK

In this research, the Urban Dynamic Data Ontology (UDO) was developed as a novel structured framework specifically designed to integrate static urban morphology data with dynamic traffic information. This new ontology extends existing urban data models by leveraging knowledge graph technology and aligning with established OSM and UDO ontologies. The developed approach facilitates cross-domain data integration, enhancing data interoperability in urban applications that was previously challenging to achieve.

The ontology extension offers significant advantages in handling and analyzing multi-source urban data, enabling the integration of various formats and sources, such as transportation systems, energy consumption, social media, and weather patterns. This integration capability helps urban planners gain a more holistic understanding of the urban environment and make more informed decisions than traditional single-domain approaches.

From an analytical perspective, the graph-based representation method developed in this research makes it easier to identify patterns and relationships that may not be immediately apparent in traditional tabular data formats. Furthermore, this new method supports predictive modeling applications - for instance, by integrating traffic patterns, weather data, and energy consumption through the UDO framework, it becomes possible to develop models that forecast future energy demand and identify areas needing efficiency improvements, ensuring the sustainability of urban infrastructure.

However, the web application prototypes remain in a proof-of-concept stage, requiring further development to support data processing, and adaptation to different urban contexts. Additionally, challenges related to data availability, standardization, and computational efficiency must be addressed for broader implementation of the

UDO framework. Future research should focus on automated data enrichment, machine learning-driven pattern recognition, and integration with real-time sensor networks to enhance the framework's applicability in urban planning, transportation modeling, and city management.

Overall, the integration of UDO with graph-based methods presents an effective approach to link and analyze cross-domain urban data, enabling predictive insights and improved decision-making capabilities that were not previously available. Further research and collaboration will be essential to advance this approach, ensuring its broader adoption and effectiveness in data-driven urban development initiatives.

Acknowledgment

This research was realized in 2023 as a Master's thesis within the Integrative Technologies and Architectural Design Research (ITECH) M.Sc. Programme at the University of Stuttgart, Germany, jointly led by the Institute for Computational Design and Construction (ICD) and its Department for Computing in Architecture (ICD/CA). The research was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Excellence Strategy - EXC 2120/1 – 390831618.

REFERENCES

Bellini, P., Benigni, M., Billero, R., Nesi, P. And Rauch, N. (2014) 'Km4City ontology building vs data harvesting and cleaning for smart-city services', *Journal of Visual Languages & Computing,* 25**,** pp.827-839. Available at: https://doi.org/10.1016/j.jvlc.2014.10.023 (Accessed: December 2014).

Chadzynski, A., Silvennoinen, H., Grisiute, A., Farazi, F., Mosbach, S., Raubal, M., Herthogs, P. and Kraft, M. (2023) 'Semantic 3D city inferences: Multi-domain reasoning on dynamic geospatial knowledge graphs', *Data-Centric Engineering,* 4. Available at: https://doi.org/10.1017/dce.2023.14. (Accessed: September 2024).

Elshani, D., et al. (2023) *Building Information Validation and Reasoning Using Semantic Web Technologies*. [Online]. Available at: http://doi.org/10.1007/978-3-031-37189-9_31 (Accessed: July 2023).

Elshani, D., Wortmann, T. & Staab, S. (2022) *Towards Better Co-Design with Disciplinary Ontologies: Review and Evaluation of Data Interoperability in the AEC Industry*. [Online]. Available at: https://www.researchgate.net/publication/360962321_Towards_Better_Co-Design_with_Disciplinary_Ontologies_Review_and_Evaluation_of_Data_Interoperability_in_the_AEC_Industry (Accessed: May 2022).

Huang, W. and Harrie, L. (2019) 'Towards knowledge-based geovisualisation using Semantic Web technologies: a knowledge representation approach coupling ontologies and rules', *International Journal of Digital Earth,* 13**,** pp. 1-22. Available at: http://doi.org/10.1080/17538947.2019.1604835 (Accessed: April 2019).

Husáková, M. & Bureš, V. (2020) 'Formal Ontologies in Information Systems Development: A Systematic Review', *Information,* 11**,** p. 66. Available at: https://doi.org/10.3390/info11020066 (Accessed: January 2020).

Rantanen, T., Julin, A., Virtanen, J.-P., Hyyppä, H. and Vaaja, M. T. (2023) 'Open Geospatial Data Integration in Game Engine for Urban Digital Twin Applications', *ISPRS International Journal of Geo-Information,* 12**,** p. 310. Available at: https://doi.org/10.3390/ijgi12080310 (Accessed: July 2023).

Romualdo-Suzuki, L. and Finkelstein, A. C. W. (2020) 'Data as Infrastructure for Smart Cities: Linking Data Platforms to Business Strategies', *ArXiv,* abs/2005.11414. Available at: https://doi.org/10.48550/arXiv.2005.11414 (Accessed: May 2020).

Sobral, T., Galvão, T. and Borges, J. (2019) 'Visualization of Urban Mobility Data from Intelligent Transportation Systems', *Sensors,* 19**,** 332. Available at: https://doi.org/10.3390/s19020332 (Accessed: January 2019).

Šoštarić, M., Vidović, K., Jakovljević, M. and Lale, O. (2021) 'Data-Driven Methodology for Sustainable Urban Mobility Assessment and Improvement', *Sustainability,* 13, 7162. Available at: https://doi.org/10.3390/su13137162 (Accessed: June 2021).