An Ontology-Driven Approach for Integrating Heterogeneous Data Sources to Enhance Urban Biodiversity and Sustainable Building Design

Albin Ahmeti[1], Defne Sunguroglu Hensel[2], Cédric Pruski[3], Jakub Tyc[1] and Michael Hensel[1]

[1]Department of Digital Architecture and Planning, Faculty of Architecture and Planning, Technical University of Vienna, Karlsplatz 13, 1040 Wien, Austria

[2]Internationalisation Demonstration School, Southeast University, Si-Pai-Lou 2, Nanjing 211102, China

[3]Luxembourg Institute of Science and Technology, 5, avenue des hauts-fourneaux, L-4362, Esch-sur-Alzette, Luxembourg

Abstract

Urban biodiversity is essential for sustainable cities, as it helps address the challenges of environmental degradation, ecosystem loss, species decline, and increased vulnerability to climate hazards, which negatively affect human health and well-being. ECOLOPES (ECOlogical building enveloPES) aims to develop a design approach for multi-species as stakeholders to achieve regenerative urban ecosystems. Integrating the diverse data required for stakeholders and beyond—spanning life sciences, geography, and architecture—and utilizing it for design presents a significant challenge. This paper introduces an innovative ontology-driven approach that integrates diverse data sources, enabling ecologists and architects to design sites and buildings that foster urban biodiversity. The proposed ontology, developed in collaboration with domain experts and adhering to Semantic Web best practices, serves as a mediator between life sciences data (e.g., species distribution and habitats) and geometric information (e.g., maps, voxel models of building structures). This integration enables the adaptation of site, buildings and geometries respectively to create habitats that attract and support urban wildlife, contributing to ecological sustainability. The paper illustrates the practical utility of the ontology through a case study, highlighting its role in guiding building designs that promote species attractiveness and urban biodiversity.

Keywords

urban biodiversity, sustainable building design, ontology-based data access, knowledge graphs

1. Motivation

The expansion of urban areas, and the resulting decline in natural spaces, has led to new challenges, including ecosystem degradation, species loss, and increased vulnerability to climate hazards among others. Urban biodiversity is essential for sustainable cities, as it helps address the challenges, which negatively affect human health and well-being. ECOLOPES (ECOlogical building enveloPES), an H2020 FET Project, aims to develop a design approach for multi-species—plants, animals, microbiota and humans—as stakeholders to achieve regenerative urban ecosystems [1].

Designing with multi-species in mind and implementing the requirements from a design brief, from both architecture and ecological point of view, presents a significant challenge. The problem is exacerbated considering that species have their unique requirements (e.g. solar), distance measures in respect to other species (e.g. prey areas), as well as distances in relation to other architectural objects (e.g. nesting). In order to reduce the complexity level for the designer, the translation of the design brief need to be represented in a more abstract form as a layer of translation that can be easily changed or guided. In order to tackle this problem, we use a network of nodes, i.e., *network configuration*, representing the requirements comprised of predefined 'circle shapes' of *EcoNodes* (denoting species) and *ArchiNodes* (denoting buildings or other architectural objects), which can be placed in the CAD environment by the designer (cf. Fig. 1).

LDAC2025 - 13th Linked Data in Architecture and Construction Workshop, July 09–11, 2025, Porto, Portugal

aljbin.ahmeti@tuwien.ac.at (A. Ahmeti); defnesunguroglu@gmail.com (D. S. Hensel); cedric.pruski@list.lu (C. Pruski); jakub.tyc@tuwien.ac.at (J. Tyc); hensel@dap.tuwien.ac.at (M. Hensel)

0000-0001-8766-4069 (A. Ahmeti); 0000-0003-2968-6432 (D. S. Hensel); 0000-0002-2103-0431 (C. Pruski); 0000-0002-3841-8360 (J. Tyc); 0000-0001-5899-5643 (M. Hensel)

The placement of nodes need to be well-informed by encapsulating and integrating diverse data required for stakeholders and beyond—spanning life sciences, geography, and architecture—and utilizing it for design. Data sources range from publicly available datasets, which can be easily accessed and retrieved via web APIs—such as GLoBI (Global Biotic Interactions for species) [2]—to private datasets generated within ECOLOPES, including Plant Functional Groups (PFGs) [3], and voxel data [4] that capture site-specific variables like coordinates, aspect, slope, or solar radiation. Ontologies and knowledge graphs provide a means of integrating data sources in a machine-readable form, complying with FAIR principles, that can be leveraged by a decision support system for the placement of nodes. By applying Semantic Web and Linked Open Data (LOD) principles data from different datasets can be linked, e.g. when species in different datasets referring to the same entity.

The resulting heterogeneous data requires integration, alignment, and consolidation into a unified RDF-based graph data model, with ontological terms used to define the mappings. We note that the *network configuration* can be easily represented as a graph data model. The outcome is a knowledge graph [5, 6] that contains curated and contextualized knowledge, enabling holistic querying and answer retrieval through the use of domain ontologies. Considering the high volume of voxel data, we have employed Ontology-based Data Access (OBDA) [7] as a data integration framework. This approach led to some data being virtualized (e.g., voxel model), while the remaining data was materialized, i.e., stored and consequently indexed by the triple store. This approach is pragmatic as the data is not moved as well as no data duplication is created, albeit with downside of performance in query evaluation.

The scope of the knowledge graph is often determined by the set of competency questions (CQs)[8] it is designed to answer, e.g.: *"Provide me with the local species that are known to have the protection status threatened?"*. We can contextualise a CQ in our setting: *"Provide me the locations in site where I can place the plant Abies alba (EcoNode) such that its solar requirements are met?"*. To answer the CQs, new data need to be ingested and harmonised, such as public data on local species provided by municipalities, citizen science (e.g. GBIF[1]), and threatened species data from International Union for Conservation of Nature (IUCN)[2].

The computational model that we have employed to generate design outcomes and their validity, based on the designer's input is rule-based. This approach computes the design outcomes based on the selected rules that can be run in a designated order, namely solar radiation and proximity constraints. Our implementation mimics a rule-based reasoning by employing a set of SPARQL queries [9] for the (selected) rules that are run in a sequence and are orchestrated by the Grasshopper environment (cf. Fig. 1).

To summarise, this paper introduces an innovative ontology-driven approach that integrates diverse data sources, enabling ecologists and architects to design sites and buildings that foster urban biodiversity. Proposed ontology-aided design methodology builds up on the concept of performance-oriented design [10, 11] and the conceptual framework focused on data-driven approaches to understanding and designing environments [12]. The paper illustrates the practical utility of the ontology through a case study, highlighting its role in guiding building designs that promote species attractiveness and urban biodiversity.

Our contributions are:

- A new EIM (Ecolopes Information Model) ontology and knowledge graph in the domains of architecture and ecology.
- A method based on OBDA to map and integrate different data sources stemming from different environments from the respective domains.
- A designer workflow in which the designer is getting feedback (e.g. fulfillment of solar radiation or proximity constraints) or ask CQs against the KG.

The rest of the paper is structured as follows. In Sec. 2 we provide the necessary background, preliminaries and related work. In Sec. 3 we explain the ontology design, reuse and alignment. In Sec. 4

[1]https://www.gbif.org
[2]https://iucn.org

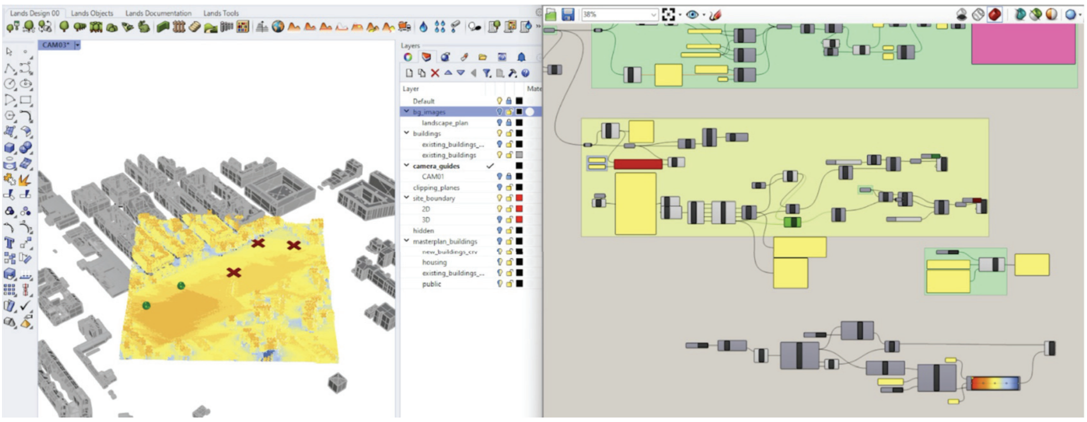

Figure 1: The placement of Nodes in CAD environment and their validation against the knowledge graph (left). Nodes that fulfil the constraints are depicted as green circles, whereas the ones that do not are depicted as red crosses. In this case constraints are subject to proximity distance between nodes or/and solar radiation, which has the information encoded in the voxel model, and therefore both can be queried via the knowledge graph.

we discuss the data integration framework used for integrating the heterogeneous data sources. In Sec. 5 we describe the demonstrator and we provide the results achieved in that context. And, finally in Sec. 6 we conclude the paper with conclusions and future work.

2. Background

The *ontology-aided generative computational design process* facilitates design generation, namely of the search space populated with alternative design solutions that can be analysed and evaluated. The overall process of design generation entails synthesising heterogeneous ecological, environmental, and architectural data, voxel-based data structuring, context-information at regional/urban and local/architectural scales, and modeling information according to decision needs.

The role of the ontologies is to model information regarding the entities and relationships that need to be represented and to aid the design of ecological building envelopes. We distinguish between ontologies as schema TBox and instance data as ABox. More formally, for a given data source and its variables x we create ABox by applying a set of mappings: $x \rightsquigarrow E(f(x))$, where E is a TBox class or property, $f(x)$ is a set of IRI templates, e.g. `https://resource.dap.tuwien.ac.at/network/{ID}`, each applied to the variable x in order to create a Literal or IRI. Knowledge graph KG is union of ABox and TBox[3], and typically stored (materialised) in a triple store. A (part of) ABox can be virtualised instead of being materialised depending on the use case requirements. In this particular case, the mappings $\rightsquigarrow$ that are represented in a specific mapping language are used to translate the queries on the fly back to the original data sources. TBox axioms are used to infer new implicit triples in ABox based on asserted (ABox) triples, which are either stored in the triple store, or alternatively query is rewritten during the query time with respect to TBox axioms to account for the same (query) answers. The latter does not hold for very expressive OWL ontology languages, but rather for (minimal) RDFS, consisting of domain/range, subClassOf and subPropertyOf axioms [13].

The role of the voxel model within the *ontology-aided design process* is to structure and spatialize datasets describing the real-world location in which the design process is taking place. Based on an extensive literature review of the voxel models' role in the engineering fields [4], an early definition of voxel models as "knowledge representation schemata" [14] was established in the context of computational design. Voxel model implementation developed for this study extends the Composite Voxel Model methodology [15] to incorporate the explicit link with the knowledge representation through the ontology-aided design process. The implemented ECOLOPES Voxel Model contains data describ-

[3]Herein, we also distinguish taxonomies or controlled vocabularies as a part of KG, sometimes referred as CBox. They are essential to capture hierarchical representations, e.g., animal ranks.

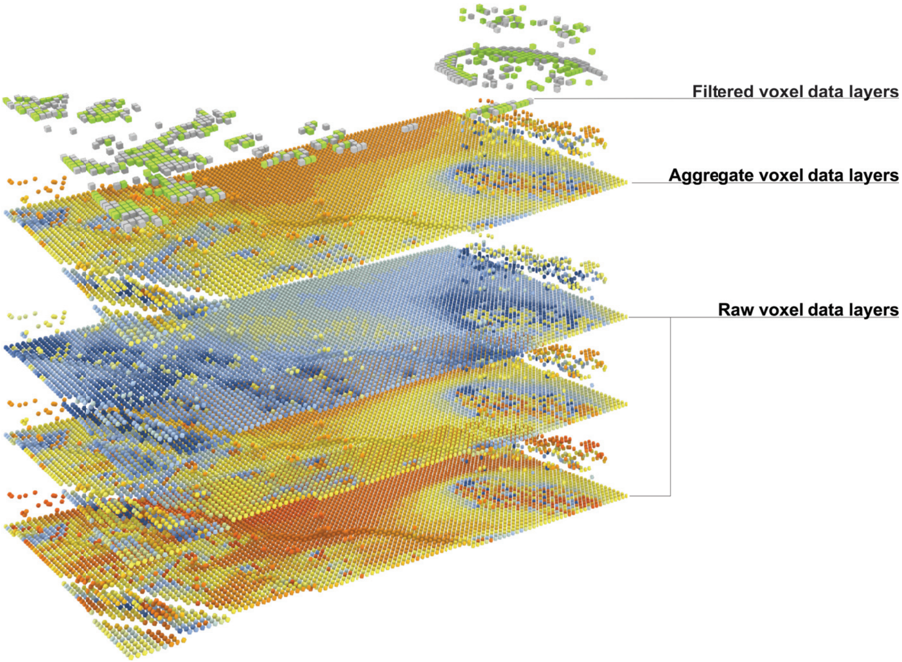

Figure 2: Three types of voxel data layers and their 3D visualisations provided by the current selection of temporal datasets [18].

ing the real-world site's physical geometry and environmental performance. Currently, voxel-based design approaches addressing anthropogenic landscape adaptations can be found [16], and studies aiming to integrate voxel- and ontological modeling are ongoing [17]. The presented implementation of the ECOLOPES Voxel Model is unique since it operates in the relational database environment (PostgreSQL), providing the implemented EIM Ontology with direct access to voxel data through the OBDA bindings. As indicated in Fig. 2, datasets provided by the ECOLOPES Voxel Model are structured as voxel data layers. Currently raw, aggregate and filtered data layers are exposed in the EIM Ontologies, enabling interactive temporal resolution change of the environmental simulation data contained in the ECOLOPES Voxel Model. For example, existing data describing monthly average sunlight exposure can be aggregated to create five additional data layers. Those layers represent average sunlight exposure (sunlight hours) in summer, autumn, winter, spring and in the growing season. Designers can retrieve voxel data, including geometry and all available parameters, stored in different levels that represent different scales and resolutions, which are exposed as a SPARQL endpoint (see Sec. 4).

Two processes are combined in the ontology-aided generative computational design process: (i) the *translational*, and (ii) the *generative process*.

In the *translational process* the designer analyses, correlates, and locates spatially architectural and ecological requirements contained in the design brief, and design-specific determinations, while taking into consideration relevant constraints (i.e. as given by planning regulations). In order to do so, the designer uses a set of predefined Nodes and places them in the CAD environment in order to satisfy ecological and architectural requirements. Herein, the designer can query the knowledge graph in order to get feedback regarding the solution space regarding permissible design outputs, in terms of satisfying the proximity distances between Nodes (e.g. species) or solar radiation requirements. Knowledge graph is also a mediator with the voxel model (cf. Fig. 2) that contains prepared datasets regarding environment variables (including geometric). In this paper we focus on the translation process only. The intended goal of the translational process is shown in Fig. 1.

The *generative process* consists of two distinct stages. In the first stage, the variations of spatial organisation are generated and evaluated. In the second one, specific surface geometries are generated for selected outputs of the previous stage and evaluated according to defined criteria, i.e. key performance indicators. The generative process falls outside the scope of this work and is planned for future exploration.

Related Work The use of Semantic Web technologies and OWL to represent ecological network specifications is explored in [19]. Ecological networks capture the structure of existing ecosystems and provide a basis for planning their expansion, conservation, and enhancement. An OWL ontology is employed to model these networks, enabling reasoning about potential improvements and identifying violations in proposed network expansions.

Global Biotic Interactions (GLoBI) [2] is an open-access platform designed for researchers, aggregating diverse datasets to provide comprehensive information on species interactions, such as predator-prey, pollinator-plant, and parasite-host relationships, among others. GLoBI offers both an API and a SPARQL endpoint (with partial data coverage) for accessing species interaction data in CSV or RDF formats. Species data is structured using the Darwin Core[4] vocabulary, with equivalence between species established through `owl:sameAs` relations. Additionally, the platform features a web-based frontend application that supports species searches via text and/or geographical boundaries. The OBO Relations[5] ontology can be leveraged for reasoning to infer implicit interactions based on explicit triples.

Nature FIRST KG [20] is a knowledge graph consisting of taxonomies that connect habitats, species and Natura 2000 areas by using so-called "crossovers" via SKOS or bespoke OWL properties. The KG describes different aspects of ecologically-related knowledge obtained from European Environment Agency (EUNIS)[6], IUCN, where different versions of habitats are connected each containing different contextual knowledge and crossover relations to species.

The integration of ontologies and knowledge graphs has gained prominence in urban ecology for their ability to model complex interactions between built environments and natural ecosystems [21]. These semantic technologies are widely applied across Architecture, Engineering, and Construction (AEC(O)) and Urban Planning to enhance data interoperability, life-cycle management, and sustainability [22, 23, 24]. In the AEC(O) domain, ontologies support structured data exchange in building modeling, optimize construction processes, enable intelligent building automation, and facilitate sustainability assessments by integrating certification standards with performance data [25]. At the urban scale, they enhance digital twins and infrastructure management through semantic city models, support environmental planning via sustainability ontologies, and model green and blue infrastructure for biodiversity and ecosystem services assessment [26, 27, 28]. Further applications include enabling multi-domain simulations, integrating vegetation and biodiversity data for urban green space management, and structuring urban sustainability indicators for performance monitoring [29, 30]. Despite these advancements, challenges remain in harmonizing ontologies across domains, ensuring scalability, and developing robust reasoning mechanisms for automated decision support.

In contrast to the approaches reviewed in this section, our proposed ontology is specifically designed to bridge the gap between the distinct domains of ecology and architecture, enabling the seamless integration of heterogeneous data sources, complying with FAIR principles. Beyond serving as a conceptual model, the ontology also supports semantic reasoning and decision-making. This empowers stakeholders to uncover new insights, infer implicit knowledge, and make informed choices in the context of sustainable urban development and biodiversity-aware architectural design.

3. Ontology Design and Development

Ontologies, particularly those built using OWL[7] (Web Ontology Language), not only provide a comprehensive description of a domain but also enable the inference of new implicit facts (i.e., RDF triples) based on the explicit facts asserted in the knowledge graph. In our case, we have modeled EIM Ontologies using a top-down conceptualization, a bottom-up approach informed by the data, and by reusing

[4]https://dwc.tdwg.org
[5]http://obofoundry.org/ontology/ro.html
[6]https://eunis.eea.europa.eu
[7]https://www.w3.org/TR/owl2-overview/

structures from well-established ontologies, namely OBO Relations [31] or Darwin Core[8] following ontology construction best practice.

EIM Ontologies developed in collaboration with domain experts and adhering to Semantic Web best practices, serves as a mediator between life sciences data (e.g., species distribution and habitats) and geometric information (e.g., maps, voxel models of building structures). They are an interface with CAD and aids the design of ecological building envelopes addressing the critical "data-to-design gap" [32][21] in computational architecture and multi-species architectural design.

3.1. Defining the Ontology Requirements

We organized a collaborative workshop with experts from the Ecolopes consortium, comprising architects and ecologists, to define the ontology and knowledge graph requirements. The participants were divided into four interdisciplinary teams, each consisting of six people, to encourage "cross-pollination" of ideas. Each team was tasked with defining a comprehensive set of CQs spanning both domains, which were written on post-it notes to foster discussion and iterative refinement.

In parallel, participants identified key concepts and relationships relevant to answering the CQs, writing them on post-it notes to form the foundation of the ontology. Where possible, they proposed synonyms, grouped concepts into categories, and outlined hierarchical relationships, effectively constructing initial ontological classes and subclasses. Teams were encouraged to identify potential properties linking concepts and to define attributes that could enrich the representation of data.

When participants could not fully define the ontology structure, they contributed by suggesting datasets or external sources that could address specific CQs, ensuring the ontology's relevance and grounding in real-world data. This bottom-up approach resulted in a draft *skeleton* of the ontology, comprising initial classes, properties, and attributes. The output serves as a robust starting point for iterative ontology development, ensuring it aligns with the goals of integrating ecological and architectural knowledge.

Requirements were captured through CQs, to name a few: *Which species we want/don't want to attract close to our building? Which species are in PFG* herbs_3*? Which species that colonize the building are protected? Which species are invasive in this location? Which species can reach or live on sloped surfaces? What is the soil depth required for PFG* herbs_3*?*

3.2. Ontology Requirements and Scope

After the CQs have been defined they were grouped based on similarity and prioritized based on importance. After the screening and processing of the CQs, we could extract general (i.e. upper level) requirements from them. In this paper, considering that processing of all CQs is out of scope, we will focus only on general constraints and requirements that are pertaining to:

- Proximity between architectural objects and species (and vice versa), or between species only.
- Solar radiation (regarding species).
- Prey areas in CAD environment, which denote areas that species are allowed to prey on.

3.3. Ontology Creation Process

The ontology[9] was developed following Semantic Web best practices, ensuring alignment with the FAIR principles. The methodology included the following core activities:

- Iterative design and refinement with input from domain experts and ecological datasets.
- Modular construction to enhance scalability and reuse.
- Leveraging established vocabularies and ontologies, such as GeoSPARQL for geographical information and Darwin Core for biodiversity data.

[8]https://dwc.tdwg.org
[9]https://github.com/aahmeti/Ecolopes

Protégé is used as a tool to manage ontologies and taxonomies, while GraphDB to manage the KG instance data. The consistency checks provided internally by Protégé have been leveraged to ensure that the ontology is consistent in respect to modeling. From the modeling perspective, we have kept the approach lean by relying mainly on RDFS, with only few OWL axioms incorporated. For ontologies, modularisation has been applied, and accordingly a prefix scheme for ontologies, taxonomies and instance data has been defined facilitating accessibility and reusability. In the end, the ontologies are also imported in GraphDB and stored in separate named graphs. The repositories are configured with reasoning enabled, and for governance reasons development and production environments are distinguished. GraphDB allows for querying of derived implicit triples based on querying of the graph `FROM <http://www.ontotext.com/implicit>` [10].

In the following are given the set of prefixes used in describing instance, taxonomy and ontological data, implying the same set of prefixes for describing and CRUD operations. The modularisation `networks-schema:` vs `ecolopes-schema:` is used to delineate the respective ontologies that are used to represent different environments, i.e. networks vs ecolopes (cf. Fig. 5).

```
PREFIX networks-schema: <https://schema.dap.tuwien.ac.at/network#>
PREFIX networks-inst: <https://resource.dap.tuwien.ac.at/network/>
PREFIX networks-taxo: <https://taxonomy.dap.tuwien.ac.at/network/>
PREFIX ecolopes-inst: <https://resource.ecolopes.org/>
PREFIX ecolopes-schema: <https://schema.ecolopes.org#>
PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
PREFIX obo-ro: <http://purl.obolibrary.org/obo/>
PREFIX dwc: <http://rs.tdwg.org/dwc/terms/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
```

3.4. Core Concepts and Structure

The ontology's core concepts address the integration of ecological, architectural, and geographical data. Key classes include:

- **Classes:** Archi- and EchoNode, Voxel, PlaneConstraint, SolarConstraint, and more.
- **Relationships:** nodeHasDistanceToNode, nodeHasSpecies, nodeHasArch, nodeOverlapsPlane, and more.
- **Attributes:** x, y, z, proximityRequirementMin, proximityRequirementMax, solarRequirementMin, solarRequirementMax, and more.

3.5. Alignment with Existing Standards

Reusing standards enhances the ontology's integration with external datasets and systems, supporting its applicability across diverse domains. To ensure interoperability, the ontology aligns with established standards and vocabularies:

- **OBO Relations:** For describing species interaction, i.e., predator-prey.
- **Darwin Core:** For species metadata, ensuring compatibility with life sciences data.
- **GeoSPARQL:** For representing geographical data, enabling spatial queries.

GLoBI employs OBO Relations to describe species interactions and their hierarchical order based on `rdfs:subPropertyOf` axioms, which can be leveraged for reasoning. Darwin Core is a widely recognized standard for capturing species metadata and species taxonomic rank (`dwc:genus`, `dwc:order`, and so forth), while the GeoSPARQL vocabulary provides a framework for representing spatial datasets in RDF. GeoSPARQL query language can be used as a standard for querying the RDF spatial datasets.

The alignment of OBO Relations has been done by adding `owl:equivalentProperty` relations, while keeping the existing GLoBI data not subject to data wrangling (cf. Fig. 3). As shown in the figure, given the relatively small number of properties, ontology matching was performed manually. Using SPARQL queries, we verified that all properties occurring in the instance data have resp. matches.

[10]https://graphdb.ontotext.com/documentation/10.8/query-behavior.html#tuning-query-behavior

	subject	predicate	object	context
1	ecolopes-schema:adjacentTo	owl:equivalentProperty	obo-ro:RO_0002220	https://tbox.crossovers.dap.tuwien.ac.at
2	ecolopes-schema:eatenBy	owl:equivalentProperty	obo-ro:RO_0002471	https://tbox.crossovers.dap.tuwien.ac.at
3	ecolopes-schema:eats	owl:equivalentProperty	obo-ro:RO_0002470	https://tbox.crossovers.dap.tuwien.ac.at
4	ecolopes-schema:hasEctoparasite	owl:equivalentProperty	obo-ro:RO_0002633	https://tbox.crossovers.dap.tuwien.ac.at
5	ecolopes-schema:hasEndoparasite	owl:equivalentProperty	obo-ro:RO_0002635	https://tbox.crossovers.dap.tuwien.ac.at
6	ecolopes-schema:hasHost	owl:equivalentProperty	obo-ro:RO_0002454	https://tbox.crossovers.dap.tuwien.ac.at
7	ecolopes-schema:hasParasite	owl:equivalentProperty	obo-ro:RO_0002445	https://tbox.crossovers.dap.tuwien.ac.at
8	ecolopes-schema:hasPathogen	owl:equivalentProperty	obo-ro:RO_0002557	https://tbox.crossovers.dap.tuwien.ac.at
9	ecolopes-schema:hostOf	owl:equivalentProperty	obo-ro:RO_0002453	https://tbox.crossovers.dap.tuwien.ac.at
10	ecolopes-schema:interactsWith	owl:equivalentProperty	obo-ro:RO_0002437	https://tbox.crossovers.dap.tuwien.ac.at
11	ecolopes-schema:parasiteOf	owl:equivalentProperty	obo-ro:RO_0002444	https://tbox.crossovers.dap.tuwien.ac.at
12	ecolopes-schema:preyedUponBy	owl:equivalentProperty	obo-ro:RO_0002458	https://tbox.crossovers.dap.tuwien.ac.at

Figure 3: OBO relations alignment via `owl:equivalentProperty` stored in a separate named graph.

4. Data Integration and Ontology-based Data Access

Currently a number of datasets have been identified and integrated in the KG using the defined ontological classes, properties or attributes; namely:

- **Vienna local plants (CSV)**: plants, metadata and solar requirements.
- **Vienna local species (shapefile)**: primarily, birds metadata (nesting boxes)[11].
- **GBIF citizen science (API)**: local species filtered based on geographical coordinates.
- **Plant Functional Groups (CSV)**: internal dataset generated via the *Ecological Model* [3].
- **Animal Functional Groups (JSON)**: internal dataset generated via the *Ecological Model*.
- **Animal-aided design (PDF)**: specifying the constraints between species and various architectural objects, used as a placeholder.
- **GLoBI (API)**: Global biotic interactions between species described in OBO Relations.
- **GBIF species (RDF)**: query federation of species including metadata details such as taxonomic rank described in Darwin Core[12][33].
- **Wikidata (RDF)**: query federation of species including metadata details, synonyms, common names, crossovers to many datasets including DBpedia, Catalogue of Life and so forth[13].
- **CAD environment**: Rhino CAD and Grasshopper environments.
- **Voxel model (RDB-RDF)**: query federation of voxel model (cf. Sec. 2).

Heterogeneous data is mapped and converted into a graph representation in KG (ABox) by using different mapping languages such as RML, Ontotext Refine and OBDA[14]. Shapefiles are converted to GeoSPARQL vocabulary in RDF using RML mappings that are generated by GeoTriples [34]. Some of the data is virtualized meaning data is not materialised (voxel model), while the remaining data is materialised. The user can query the KG using SPARQL as a unified interface that encompasses and integrates different datasets combining the results, and, if required, reasoning on top using TBox statements. Therefore, the KG integrates heterogeneous data from different sources, providing unified access to the data by using the SPARQL query language.

For the implementation of the virtualisation we use Ontop [35] module that is available within GraphDB triple store. The system allows to define the sources (voxel model stored and managed in the relational database) and mapping definitions (OBDA or R2RML[15]), thereby exposing it as separate repository in which SPARQL queries (via `SERVICE` keyword) are translated to SQL queries respectively on the fly. For the time being, we have used the native OBDA mappings, and in the future we plan to replace with the more future proof R2RML mapping language.

[11] Provided by city of Vienna https://www.wien.gv.at/umweltschutz/umweltgut/
[12] http://graph.openbiodiv.net
[13] https://query.wikidata.org
[14] https://graphdb.ontotext.com/documentation/10.8/virtualization.html
[15] https://www.w3.org/TR/r2rml/

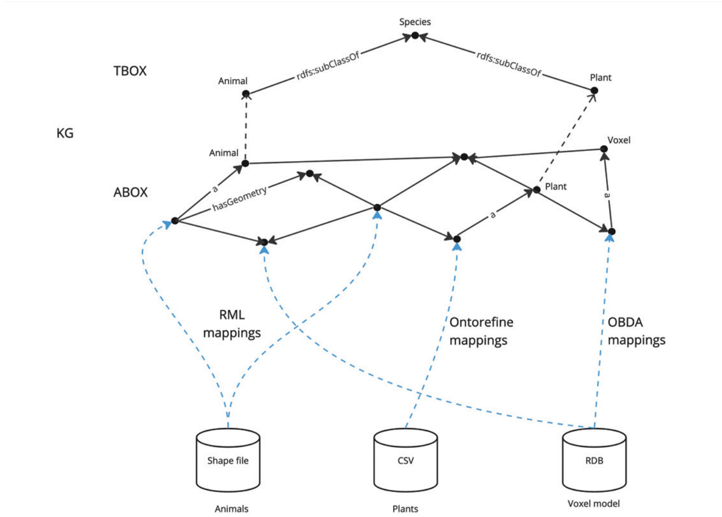

Figure 4: Ontology-based data access - data integration framework of heterogeneous datasets. Some of the datasets are materialised in KG (RML and Ontotext refine mappings), with the aim of achieving better query performance.

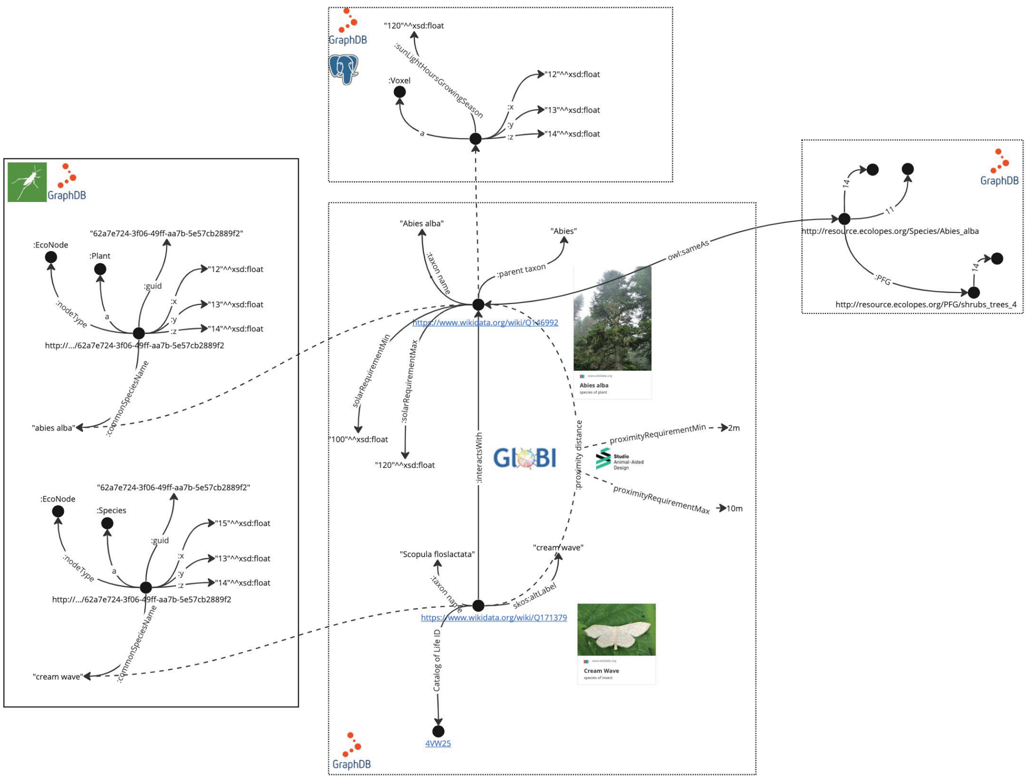

Figure 5: Knowledge graph snapshot connecting data from different environments. The resulting graph data is managed in the GraphDB triple store. The dashed lines represent equality (computed via query on run time) between entities in KG, whereas in other cases they are explicitly asserted using `owl:sameAs` relations.

Figure 5 shows a snapshot of the KG, which links knowledge from datasets originated in Grasshopper (left), global biotic interactions and proximity constraints (centre), voxel model (top) and PFGs (right). One can traverse from one context to another given that the relations between datasets are explicitly encoded (e.g. `owl:sameAs`), or they can be inferred by SPARQL queries (e.g. if x,y,z coordinates in-between datasets coincide). Whenever possible for URIs of species we treat Wikidata identifiers as canonical URIs, and provide crossovers via `owl:sameAs` to other linked datasets.

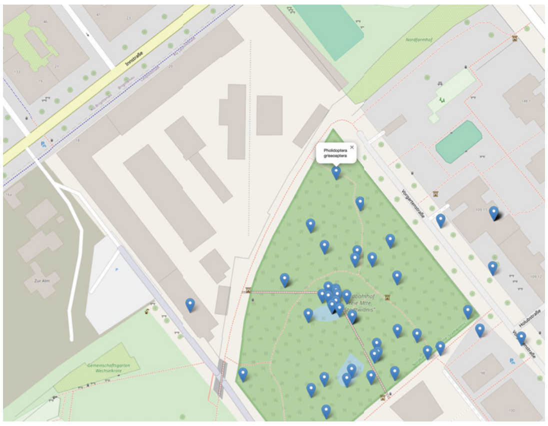

Figure 6: The citizen science data on species obtained from GBIF in the proximity of Nordbahnhof.

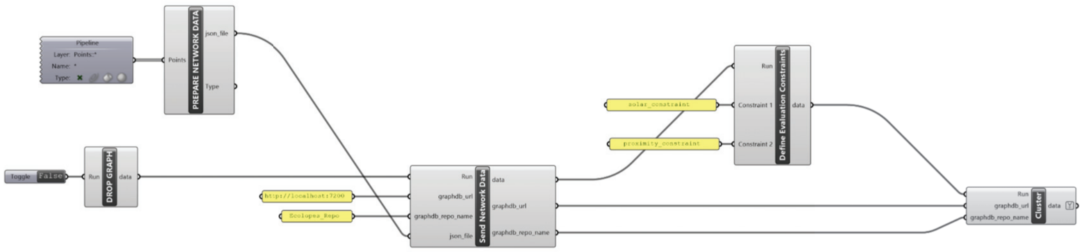

Figure 7: The complete Grasshopper pipeline, which prepares points as JSON-LD, uploads them to GraphDB. Afterwards, validates the chosen constraints and returns a boolean value.

5. Demonstrator

In the following we demonstrate the application and utility of the ontology and KG in a specific context in Vienna, in the site of the Nordbahnhof Freie Mitte location. Local species around the site are curated and provided by Vienna municipality, and in addition can be obtained from GBIF API (cf. Fig. 6) - albeit with the common data quality issues prone in the common citizen science approaches.

The pipeline (cf. Fig. 7) that we will describe in the following, consists of various steps, starting with cleaning, mapping of the network consisted of Nodes in JSON-LD serialisation[16], preparing and pushing data to GraphDB, and finally the computation of the validation. Before every run, the data is cleaned (DROP GRAPH via curl command) using Grasshopper, in order to account for the changes in the CAD environment. The rationale for choosing JSON-LD lies on the fact that JSON-LD is compatible with JSON, and the Grasshopper components could operate using JSON data as a well-established standard.

Mapping of Voxel Data Layers in EIM Ontology Voxel model contains environmental data on the specific site of Nordbahnhof. `sunlightHoursSummer` and `sunlightHoursGrowingSeason` properties map voxel data layers in the respective ontology attributes. For instance, designer places an EcoNode, defines its `subType` as `Plant`, and sets `commonSpeciesName` as "Silver fir". Ontology stores mappings between `commonSpeciesName` ("Silver fir") and `latinSpeciesName` (e.g., *Abies alba*), along with additional plant datasets and respective metadata. Consequently, sunlight requirements (`solarRequirementMin` and `solarRequirementMax`) are retrieved for the given EcoNode.

Validation Process for EcoNodes and ArchiNodes Compares `sunlightHoursGrowingSeason` from the ECOLOPES Voxel Model with `solarRequirementMin` and `solarRequirementMax`. Returns a boolean value (`?isValid`) as the result of the comparison. Designer places an ArchiNode and

[16]https://www.w3.org/TR/json-ld11/

```
INSERT {
        GRAPH <https://data.validation.dap.tuwien.ac.at> {
                [] a networks-schema:ValidationResult ;
                   networks-schema:validationResultHasConstraint ?constraint ;
                   networks-schema:validationResultHasNode ?node1 ;
                   networks-schema:validationResultHasNode ?node2 ;
                   networks-schema:isValid ?final .
        }
}
WHERE
{
        ?node1 networks-schema:x ?x1 ;
                networks-schema:y ?y1 ;
                networks-schema:z ?z1 .
        { ?node1 networks-schema:nodeHasSpecies ?species1 . }

        ?node2 networks-schema:x ?x2 ;
                networks-schema:y ?y2 ;
                networks-schema:z ?z2 .
        { ?node2 networks-schema:nodeHasSpecies ?specOrArch2 . }
                UNION { ?node2 networks-schema:nodeHasArch ?specOrArch2  }


        FILTER (str(?node1) < str(?node2))
        BIND (ofn:sqrt(ofn:pow(?x1 - ?x2, 2) + ofn:pow(?y1 - ?y2, 2) + ofn:pow(?z1 - ?z2, 2)) as ?res)

        <<?species1 ecolopes-schema:proximityDistance ?specOrArch2>>
                                ecolopes-schema:speciesHasRequirement ?constraint .
        ?constraint ecolopes-schema:proximityRequirementMin ?min ;
                                ecolopes-schema:proximityRequirementMax ?max .

        BIND (IF(?min <= ?res && ?res <= ?max,true,false) AS ?final)
}
```

Figure 8: INSERT query – as a counterpart to CONSTRUCT – that inserts the validation results based on the proximity constraints.

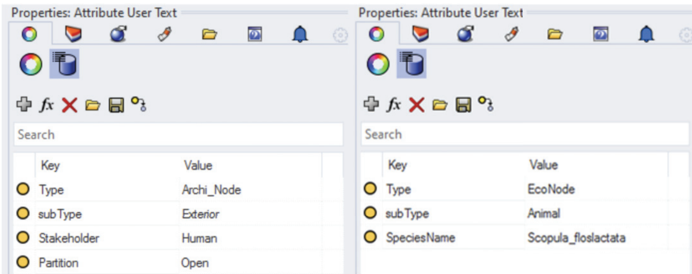

Figure 9: The placed EcoNodes and ArchiNodes, including the respective metadata defined in CAD panel. The metadata are mapped to the resp. attributes of the ontology, creating entities in ABox.

defines its `subType` as `Infrastructure` or `Building`. Similar validation and mapping processes are outlined for ArchiNodes, including the prey area that is predefined by the designer as a rectangular shape. Rule-based process is based on CONSTRUCT queries for the respective solar, proximity and prey areas constraints. In Fig. 8 one can see the query for the proximity constraint. The CONSTRUCT query is transfored to an INSERT query in order the validation data to not only be computed but also materialised. The query takes into account multiple proximity constraints between different nodes n, in the range of $O(n^2)$ computational complexity. Variable ?res computes the *Euclidian distance* between nodes using GraphDB's math functions[17].

Preparing and Sending Network Data Reads data stored as key-value pairs from Rhino Objects representing Network Nodes. Rhino Point geometry created via the `PlaceArchiNode` and `PlaceEcoNode` commands (cf. Fig. 9). These are used to prepopulate Nodes based on a selection. Converts and outputs data to JSON-LD matching the ontology (TBox) schema. Sends JSON-LD formatted Network Node data to the GraphDB environment.

Define Evaluation Constraints Enables the designer to select constraints for validating the *network configuration* created in Rhino. One can choose as input Solar or Proximity constraints to be validated against, cf. Fig. 10 for visualisation of the validation, which is stored in GraphDB. The final output is a boolean answer that is computed by the AND ("join") intersection of constraints using MIN operator.

[17] https://graphdb.ontotext.com/documentation/10.8/sparql-ext-functions-reference.html#mathematical-function-extensions

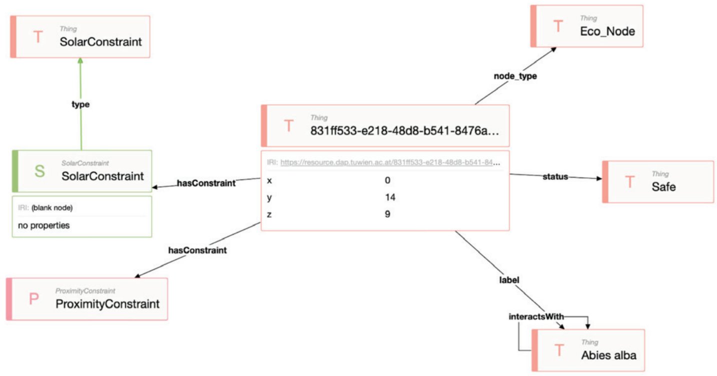

Figure 10: The graph visualisation (in open access version of Ontodia [36]) of the EcoNode that satisfies both solar and proximity constraints.

6. Conclusion and Future Work

This paper presented an ontology-driven approach to integrate diverse datasets from public APIs, private sources, and CAD environments into a unified RDF-based knowledge graph, facilitating ecological building design. By leveraging OBDA, the approach balances virtualization and materialization, ensuring efficient data integration while supporting advanced queries and reasoning. The ECOLOPES ontologies bridges life sciences and geometric data, enabling designers to address ecological constraints like solar radiation and proximity through interactive workflows. A case study highlighted the ontology's practical application in designing habitats that promote species attractiveness and urban biodiversity. For future work, we plan to use SHACL to define constraints, use a rule language instead of SPARQL queries in order to adhere to Semantic Web standards. More importantly to extend the problem setting towards generative process (geometric articulation) of the ontology-aided generative computational design by employing more sophisticated approaches akin to answer set programming.

Acknowledgments

This work is funded by the European Union Horizon 2020 Future and Emerging Technologies research project ECOLOPES: ECOlogical building enveLOPES (grant number: 964414). The funding source is not involved in the conduct of the research or preparation of the article. Furthermore, we extend our gratitude to Mariasole Calbi for kindly providing the PFGs dataset, to Victoria Culshaw for sharing the AFGs dataset, and to Navid Javan for assisting us with the Grasshopper workflow and components.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Grammarly in order to: grammar and spelling check, paraphrase and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] K. Perini, M. Canepa, S. Barath, M. Hensel, A. Mimet, S. Uthaya Selvan, E. Roccotiello, T. Selami, D. Sunguroglu Hensel, J. Tyc, V. Vogler, W. Weisser, F. Mosca, ECOLOPES: A multi-species design approach to building envelope design for regenerative urban ecosystems, in: Responsive Cities:

Design with Nature Symposium Proceedings 2021, Institut d'Arquitectura Avançada de Catalunya, Barcelona, 2021, pp. 368–380.
[2] J. H. Poelen, J. D. Simons, C. J. Mungall, Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets, Ecol. Inform. 24 (2014) 148–159.
[3] M. Calbi, G. Boenisch, I. Boulangeat, D. Bunker, J. A. Catford, A. Changenet, V. Culshaw, A. S. Dias, T. Hauck, J. Joschinski, J. Kattge, A. Mimet, M. Pianta, P. Poschlod, W. W. Weisser, E. Roccotiello, A novel framework to generate plant functional groups for ecological modelling, Ecol. Indic. 166 (2024) 112370.
[4] J. Tyc, T. Selami, D. S. Hensel, M. Hensel, A scoping review of voxel-model applications to enable multi-domain data integration in architectural design and urban planning, Architecture 3 (2023) 137–174.
[5] D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, A. Wahler, Knowledge graphs: Methodology, tools and selected use cases, 1 ed., Springer Nature, Cham, Switzerland, 2020.
[6] A. Hogan, C. Gutierrez, M. Cochez, G. D. Melo, S. Kirrane, A. Polleres, R. Navigli, A.-C. N. Ngomo, S. M. Rashid, L. Schmelzeisen, S. Staab, E. Blomqvist, C. d'Amato, J. E. L. Gayo, S. Neumaier, A. Rula, J. Sequeda, A. Zimmermann, Knowledge Graphs, Synthesis Lectures on Data, Semantics, and Knowledge, Springer International Publishing, Cham, 2022. URL: https://link.springer.com/10.1007/978-3-031-01918-0. doi:10.1007/978-3-031-01918-0.
[7] A. Poggi, D. Lembo, D. Calvanese, G. D. Giacomo, M. Lenzerini, R. Rosati, Linking data to ontologies, J. Data Semant. 10 (2008) 133–173.
[8] M. Grüninger, M. S. Fox, Methodology for the design and evaluation of ontologies, in: IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995.
[9] A. Polleres, From sparql to rules (and back), in: Proceedings of the 16th International Conference on World Wide Web, WWW '07, Association for Computing Machinery, New York, NY, USA, 2007, p. 787–796. URL: https://doi.org/10.1145/1242572.1242679. doi:10.1145/1242572.1242679.
[10] M. Hensel, Performance-Oriented Architecture: Rethinking Architectural Design and the Built Environment, 1 ed., Wiley, 2013. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118640630. doi:10.1002/9781118640630.
[11] M. U. Hensel, D. Sunguroğlu Hensel, Performances of Architectures and Environments: En Route to a Theory and Framework, in: The Routledge Companion to Paradigms of Performativity in Design and Architecture, Routledge, 2019. Num Pages: 12.
[12] D. Sunguroğlu Hensel, J. Tyc, M. Hensel, Data-driven design for Architecture and Environment Integration: Convergence of data-integrated workflows for understanding and designing environments, SPOOL 9 (2022) 19–34. URL: https://www.spool.ac/index.php/spool/article/view/195. doi:10.47982/spool.2022.1.02.
[13] A. Ahmeti, Updates in the context of ontology-based data management, Ph.D. thesis, Technische Universität Wien, Wien, 2020. doi:10.34726/hss.2020.75783.
[14] S. N. Srihari, Representation of Three-Dimensional Digital Images, ACM Computing Surveys 13 (1981) 399–424. URL: https://doi.org/10.1145/356859.356862. doi:10.1145/356859.356862.
[15] J. Tyc, D. Sunguroğlu Hensel, E. I. Parisi, G. Tucci, M. U. Hensel, Integration of Remote Sensing Data into a Composite Voxel Model for Environmental Performance Analysis of Terraced Vineyards in Tuscany, Italy, Remote Sensing 13 (2021) 3483. URL: https://www.mdpi.com/2072-4292/13/17/3483. doi:10.3390/rs13173483, number: 17 Publisher: Multidisciplinary Digital Publishing Institute.
[16] J. Tyc, E. I. Parisi, G. Tucci, D. Sunguroğlu Hensel, M. U. Hensel, A data-integrated and performance-oriented parametric design process for terraced vineyards, Journal of Digital Landscape Architecture (2022) 504–520. URL: https://gispoint.de/fileadmin/user_upload/paper_gis_open/DLA_2022/537724049.pdf. doi:10.14627/537724049.
[17] F. Ludwig, M. Hensel, T. Rötzer, A. Ahmeti, X. Chen, H. I. Erdal, A. Reischel, S. Qiguan, J. Tyc, H. Yazdi, Digital Workflow for Novel Urban Green System Design Derived from a Historical Role Model, Journal of Digital Landscape Architecture (JoDLA) 2024 (2024) 333–345. URL: https://doi.org/10.14627/537752030. doi:10.14627/537752030.

[18] M. U. Hensel, J. M. Tyc, D. S. Hensel, A. Ahmeti, M. A. Cifci, ECOLOPES Voxel Model, Technical Report, TU Wien, 2023.
[19] G. Torta, L. Ardissono, L. La Riccia, A. Savoca, A. Voghera, Representing ecological network specifications with semantic web techniques, in: Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, SCITEPRESS - Science and Technology Publications, 2017.
[20] A. Ahmeti, J.-K. Schakel, R. David, A. Revenko, Towards preserving biodiversity using nature FIRST knowledge graph with crossovers, in: I. Fundulaki, K. Kozaki, D. Garijo, J. M. Gómez-Pérez (Eds.), Proceedings of the ISWC 2023 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, November 6-10, 2023, volume 3632 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023.
[21] C. Pruski, D. S. Hensel, The role of information modelling and computational ontologies to support the design, planning and management of urban environments: Current status and future challenges, in: A. Chokhachian, M. U. Hensel, K. Perini (Eds.), Informed Urban Environments: Data-Integrated Design for Human and Ecology-Centred Perspectives, Springer International Publishing, Cham, 2022, pp. 51–70.
[22] F. Sadeghineko, B. Kumar, Application of semantic web ontologies for the improvement of information exchange in existing buildings, Construction Innovation 22 (2022) 444–464.
[23] F. H. Abanda, B. Kamsu-Foguem, J. H. M. Tah, Bim–new rules of measurement ontology for construction cost estimation, Engineering Science and Technology, an International Journal 20 (2017) 443–459.
[24] L. Liao, Y. Wen, W. Mo, C. Gan, A carbon footprint management framework for prefabricated buildings based on knowledge graph, in: International Symposium on Advancement of Construction Management and Real Estate, Springer, 2023, pp. 1551–1569.
[25] O. I. Olanrewaju, W. I. Enegbuma, M. Donn, N. Chileshe, Building information modelling and green building certification systems: A systematic literature review and gap spotting, Sustainable Cities and Society 81 (2022) 103865.
[26] J. Shi, Z. Pan, L. Jiang, X. Zhai, An ontology-based methodology to establish city information model of digital twin city by merging bim, gis and iot, Advanced Engineering Informatics 57 (2023) 102114.
[27] F. Sica, P. De Paola, F. Tajani, E. Doko, Spatial–temporal ontology of indicators for urban landscapes, Land 14 (2025) 72.
[28] F. Affinito, J. M. Holzer, M.-J. Fortin, A. Gonzalez, Towards a unified ontology for monitoring ecosystem services, EcoEvoRxiv (2024).
[29] R. Ma, Q. Li, B. Zhang, H. Huang, C. Yang, An ontology-driven method for urban building energy modeling, Sustainable Cities and Society 106 (2024) 105394.
[30] M. Al-Ageili, M. Mouhoub, An ontology-based information extraction system for residential land-use suitability analysis, International Journal of Software Engineering and Knowledge Engineering 32 (2022) 1019–1042.
[31] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, C. Rosse, Relations in biomedical ontologies, Genome Biol. 6 (2005) R46.
[32] J. Glick, Ontologies and databases – knowledge engineering for materials informatics, in: Informatics for Materials Science and Engineering, Elsevier, 2013, pp. 147–187.
[33] L. Penev, M. Dimitrova, V. Senderov, G. Zhelezov, T. Georgiev, P. Stoev, K. Simov, OpenBiodiv: A knowledge graph for literature-extracted linked open data in biodiversity science, Publications 7 (2019) 38.
[34] K. Kyzirakos, D. Savva, I. Vlachopoulos, A. Vasileiou, N. Karalis, M. Koubarakis, S. Manegold, Geotriples: Transforming geospatial data into rdf graphs using r2rml and rml mappings, Journal of Web Semantics 52-53 (2018) 16–32. URL: https://www.sciencedirect.com/science/article/pii/S1570826818300428. doi:https://doi.org/10.1016/j.websem.2018.08.003.
[35] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro,

G. Xiao, Ontop: Answering SPARQL queries over relational databases, Semant. Web 8 (2017) 471–487.

[36] D. Mouromtsev, D. Pavlov, Y. Emelyanov, A. Morozov, D. Razdyakonov, M. Galkin, The simple web-based tool for visualization and sharing of semantic data and ontologies, in: International Semantic Web Conference (Posters & Demos), 2015.