# Dynamic Control of District Heating Networks with Integrated Emission Modelling: A Dynamic Knowledge Graph Approach

Markus Hofmeister[1,2,3], Kok Foong Lee[2], Yi-Kai Tsai[2], Magnus Müller[2], Karthik Nagarajan[2], Sebastian Mosbach[1,2,3], Jethro Akroyd[1,2,3], Markus Kraft[1,2,3,5]

released: February 7, 2024

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

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

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

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

Preprint No. 319

UNIVERSITY OF CAMBRIDGE

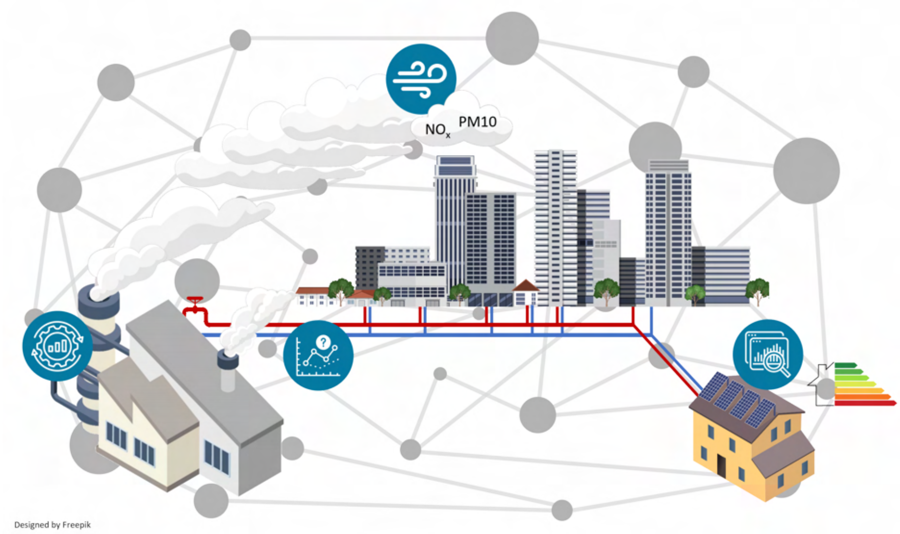
*Keywords:* knowledge graph, digital twin, interoperability, energy modelling, emission dispersion

**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**

This paper presents a knowledge graph-based approach for the dynamic control of a district heating network with integrated emission dispersion modelling. We propose an interoperable and extensible implementation to forecast the anticipated heat demand of a municipal heating network, minimise associated total generation cost based on a set of available heat sources, and couple it with dispersion modelling of corresponding emissions to provide automatic insights into air quality implications of various heat sourcing strategies. We achieve cross-domain insights in the nexus of energy and air quality via a set of developed ontologies and autonomous software agents, which can be chained together via the World Avatar dynamic knowledge graph to resemble the behaviour of complex systems. Furthermore, we have integrated the City Energy Analyst into this ecosystem to provide building-level insights into energy demand and generation potential to foster strategic analyses and scenario planning. Utilising actual instantiated building and weather data, this enhanced bottom-up version addresses inherent assumptions in the official software release, facilitating a more data-driven approach. All use cases are implemented for a mid-size town in Germany as a proof-of-concept, and a unified visualisation interface is provided, allowing for the examination of 3D buildings alongside their corresponding energy demand and supply time series, as well as emission dispersion data. With this work, we outline the potential of Semantic Web technologies to connect digital twins for holistic energy modelling in smart cities, thereby addressing the increasing complexity of interconnected energy systems.




## Highlights

- Implement knowledge graph native model predictive control-style optimisation.
- Connect energy and air quality domain with integrated emission dispersion modelling.
- Embed general forecasting capabilities within a dynamic knowledge graph.
- Develop domain ontologies to represent district heating operations data.

# Contents

# 1 Introduction

Climate change arguably poses humanity's most formidable challenge, impacting almost every aspect of our lives including public health, food safety, water supply, biodiversity, and increasing the frequency of extreme weather events [49, 55]. Recognising greenhouse gas emissions as key driver of climate change, the transition towards a low-carbon future is widely acknowledged as a crucial imperative [11, 41]. There is consensus that the decarbonisation of the energy sector requires significant changes, such as increased sector-coupling and greater penetration of distributed renewable resources like wind and solar as well as the development of intelligent infrastructure and modelling approaches [3, 41, 94]. Solutions for this inherently interdisciplinary transition require holistic consideration of social, economic, environmental, and engineering factors across various geographic and temporal scales [87].

Digital technologies like advanced metering infrastructure, big data, machine learning, and the internet of things are increasingly recognised for facilitating cost-effective decarbonisation [69]. Orchestrating these technologies into cyber-physical systems yields synergies that enhance energy and industrial efficiency, optimising both economic feasibility and environmental impact [55]. The application of cyber-physical systems in energy research has grown significantly in recent years, often in the form of digital twins to explore optimal solutions for real-world problems through the study of fully digital replicas [105]. A digital twin can be understood as a realistic digital representation of assets, processes, or entire systems describing their current state and how they behave over time and under different conditions and constraints, offering augmented opportunities for interaction with the physical counterpart, including analyses, optimisation, simulation or active feedback into the physical twin [1]. Digital twins have effectively addressed numerous real-world problems [114]; however, the majority remains isolated and lacks interoperability due to differences in set-up, hardware or software, often stemming from individual funding initiatives or business interests [91]. Interoperability is defined as the ability of tools, systems, and data to understand and use each other's functionalities, which is essential to foster reusability and address cross-domain questions comprehensively and collectively [110]. Challenges arise in sectors like energy with legacy structures, traditional silo-thinking, and the use of proprietary instead of open standards [94]. For future energy systems, however, effective cooperation and coordination beyond the 'traditional' energy sector are essential to maximise synergies of increasingly intertwined systems, encompassing the built environment, transportation, water, *etc*. But a holistic optimisation framework requires more than just data assimilation. Instead, interoperability between tools and models will be necessary [55, 89, 94].

Integrated energy system modelling represents a promising direction for digital twins in the energy transition [114]. Although it is anticipated that energy modelling will transition from single-institution models to distributed, collaborative approaches, allowing multiple domain experts to contribute [108], integrating data across domains and resolving ambiguities while ensuring openness and transparency remains a widespread problem [88, 89]. Data are highly heterogeneous in both format and semantics, as different sources (*i.e.*, sensors, texts, web, *etc.*) use individual formats (*e.g.*, tabular data, geospatial data, natural language, *etc.*) [92]. Moreover, a lack of semantic interoperability can arise when certain information is only known implicitly by domain experts or when the same concept might possess different meanings in different domains; however, an aligned understanding and transparency of models, assumptions, and data is pivotal [55]. O'Dwyer et al. [84] demonstrate a sustainable energy management system to manage the flow of data between machine learning models, cities and districts; however a general and scalable solution for the construction of cross-domain models remains unrealised, impeding the ability to reproduce results as well as to adapt and combine existing models [88]. A potential solution to this theme of problems can be generalised in the form of connected digital twins — distributed collaborative entities that share data and computational capabilities to efficiently and effectively address complex questions [1]. As the value of providing meaningful decision-making support relies on the availability, accessibility, and compatibility of data and computational capabilities [105], a comprehensive energy digital twin for a smart city, for instance, must integrate diverse data formats, existing applications, and even other digital twins to enable holistic analyses across different levels of aggregation, from individual building to city scale.

The World Avatar project [1] creates an ecosystem that enables the transparent integration of heterogeneous models and data, improving interoperability between various formats and software [99]. It leverages technologies from the Semantic Web stack to create a distributed dynamic knowledge graph, which by design is well suited to effectively address cross-domain questions. The World Avatar combines ontologies (*i.e.*, data definitions) with actual data instances and computational agents, which provide the dynamic nature of the knowledge graph. These autonomous agents act as executable knowledge components and accomplish tasks such as updating the graph to ensure it remains current in time, simulating systems, or transmitting responses to the physical world. As all data are instantiated based on aligned modular ontologies with explicit descriptions of notions for different domains, all agents share a common world view and ensure self-consistent analyses. Agents can represent black box, grey box or physics-based models and also wrap around existing software or third party application programming interface (APIs) to make them available semantically. With an initial focus on chemical and process engineering [33, 96, 117], the World Avatar has evolved into a versatile tool to address decarbonisation questions in the energy sector [6, 27, 99, 100], overcome cross-domain interoperability challenges in smart cities and city planning [20, 52], and improve the resilience of complex systems [28]. Akroyd et al. [1] showed how a dynamic general-purpose knowledge graph-based on ontologies and autonomous semantic agents is ideally suited to realising connected digital twins, *e.g.*, to control real-world assets, perform cross-domain simulations, or conducting geospatial and scenario analyses.

The **purpose of this paper** is to provide an example of dynamic knowledge graph capabilities to realise connected digital twins, illustrated using the World Avatar as versatile tool to provide a holistic energy perspective for smart cities: We present a proof-of-concept for a model predictive control-style optimisation of heat generation in a district heating network, executed entirely within the knowledge graph. Optimisation outputs are directly used for integrated emission dispersion modelling to understand the impact of various heat generation and sourcing strategies on air pollution. Beyond this dynamic and supply-side-focused optimisation, the City Energy Analyst is made available as part of the World Avatar to provide strategic demand-side insights into buildings' energy profiles and own generation potentials based on latest instantiated building stock data. With this we provide concrete implementation examples that showcase capabilities previously introduced conceptually by Akroyd et al. [1].

The structure of this paper is as follows: Section 2 provides an overview of the current energy modelling landscape for cities, together with its challenges, as well as an introduction to the World Avatar dynamic knowledge graph; section 3 details newly developed ontologies and software capabilities to address identified interoperability gaps using the World Avatar; section 4 highlights the results from the connected digital twin implementation; and section 5 concludes the work.

# 2 Background

This section provides an overview of previous research and the status quo in several fields relevant to this work, namely energy system and dispersion modelling as well as prevalent interoperability gaps within the energy domain. Each topic is introduced independently, following conventional community practices; however, these silos are resolved in the next chapter using a dynamic knowledge graph approach. The World Avatar dynamic knowledge graph, which enables this integration, is also introduced here, alongside existing and reused ontology efforts.

## 2.1 Energy system modelling for smart cities

Energy systems are increasingly intertwined and demand a comprehensive approach to drive overall resource efficiency and decrease emissions [3, 11]. Moreover, they are closely tied to numerous key challenges of the twenty-first century, including security, affordability, and resilience of energy supply, as well as political, socio-economic, and environmental concerns, ranging from local air and water pollution to, most importantly, climate change and global sustainability [87, 113]. Initially focused on security of supply and costs, energy system modelling has pivoted to also study transition pathways and strategies for a carbon-neutral future. An integrated energy system approach is imperative to provide an efficient, low-carbon, and reliable energy supply, consolidating the planning and scheduling of diverse energy carriers, such as electricity, gas, heating, and cooling [114].

In this context, the established methods to model energy systems are being challenged by several emerging themes, as extensively and consistently discussed in the literature [41, 49, 108, 113]: increased sector-coupling and interactions between energy vectors (*i.e.*, electricity, natural gas, hydrogen, heating and cooling) at various scales (*e.g.*, from multi-national, national, community scale down to building level); rising flexibility of demand driven by new technologies such as smart meters and load shifting; enhanced integration of intermittent renewable resources, with the resulting need for more temporal detail; distributed generation and an increasing share of prosumers, with the resulting need for higher spatial granularity. To address the increased complexity of multi-carrier energy systems, modelling frameworks need to balance uncertainty and transparency while optimising across scales, considering multiple spatial and temporal resolutions [41, 87].

A strong push towards open and transparent simulation and planning tools has been observed in recent years as vital building blocks for modelling approaches that bridge these scales and domains [49, 114]. While it has been demonstrated that open-source modelling frameworks and data platforms are often on par with proprietary or commercial models [41], impediments to interoperability persist due to technical and market barriers. Diverse requirements and limitations (*e.g.*, regional scope) of individual tools, coupled with variations in applicability to specific problems and scenarios, pose a real challenge in integrating data and models [66, 114]. To address these challenges, semantic approaches have been proposed, such as by Li and Hong [66], who developed a framework for grid-interactive efficient buildings which are responsive to grid pricing or carbon signals to achieve energy and carbon neutrality.

## 2.2 The City Energy Analyst

The City Energy Analyst (CEA) is an established open-source computational framework for urban energy system analysis [42], offering insights into buildings' overall energy demand, heating and cooling requirements, *etc.* as well as on-site renewable energy generation potentials. It has a global user base and has been applied to numerous case studies across the world. In Switzerland, CEA has been used to study the impact of seasonal effects [75] and air infiltration rates [45] on building energy demands, as well as for the analysis of photovoltaic (PV) system adoption [71]. Moreover, CEA has been used to assess the energetic implications of proposed master plans for residential districts in Almere [76], the Netherlands, as well as the analysis of building integrated PV installations in Singapore [46].

The CEA toolkit comes with built-in databases, containing numerous assumptions required to run simulations. The databases include information about building properties (*i.e.*, OpenStreetMap (OSM) [83] building footprint, height and building usage) as well as environmental data such as weather and terrain [42]. Although this approach enables users to conduct simulations without the requirement for specific input data, the built-in assumptions may not always be representative. Prioritising broad applicability over the integration of actual building-specific characteristics is a deliberate design choice, inherent to many top-down energy assessment tools.

## 2.3 Interoperability gaps in current energy modelling landscape

Interoperability is the ability of different systems, devices, or applications to exchange and use information effectively and collectively. While technical interoperability within the energy sector is quite well-established, there is a need for more informational, functional, and business interoperability [92]. Despite numerous initiatives involving both academia and industry, many interoperability gaps remain, such as coordination issues between relevant stakeholders and efforts as well as a lack of practical tools for assessing interoperability capabilities of individual platform solutions [57, 94]. Fragmented platforms dominate the energy modelling landscape, with no unified approach to harmonise data models or knowledge across all domains of the value chain [29]. This poses challenges for data integration, model validation, scenario comparisons, policy evaluation, and often results in biased or subpar overall system performance, as decision-makers lack valuable information to assess certain cross-domain trade-offs or co-benefits of different scenarios.

Just to name a few examples, cross-domain interoperability would allow for the assessment of life cycle environmental impacts of diverse building design alternatives, including materials, appliances, building orientation, shape, *etc.* Furthermore, the effects of extreme weather events, such as heat waves, floods, storms, and earthquakes, on both the built environment and smart grid infrastructure can be studies, aiding in identifying potential weak points and enhancing resilience [28]. Moreover, emission analyses could extend beyond the established assessment of overall amounts to explore detailed dispersion patterns of individual air pollutants as the result of different energy provision strategies, by incorporating location and weather data.

Current interoperability gaps can be addressed by adopting standards and frameworks to facilitate communication and collaboration among different modelling tools, stakeholders, and platforms or enhancing information exchange using common data models, ontologies, and Semantic Web technologies [94]. While the first approach remains focused on the broader energy domain (*e.g.*, by incorporating solar panels, battery storage, heat pumps, boilers and electric vehicles) [29], the latter one is in principle capable to connect seamlessly with any related domain, such as transport, agriculture and industrial production [57]. Eibeck et al. [33] discuss an initial Semantic Web-based attempt to estimate dispersion profiles for emissions of a power plant considering the effects of surrounding buildings and real-time weather conditions. In this work, we expand upon this study and integrate dispersion modelling into a dynamic energy generation dispatch problem to showcase interoperability between the energy and air quality domain.

## 2.4 Dispersion modelling

Dispersion models can broadly be categorised as box models, Gaussian plume models, or advanced physical models (*e.g.*, computational fluid dynamics, Eulerian or Lagrangian) [35, 53, 62, 82]. Box models utilise a rectangular control volume to estimate the average concentrations of pollutants. They assume that all air in the box is well-mixed and pollutants can enter and exit the box freely. Due to their simplicity, they can accommodate detailed chemical reactions. Gaussian plume models assume that the pollutant concentrations follow a Gaussian distribution. Advanced models typically solve detailed transport equations to conserve mass and momentum, and are usually more computationally expensive.

Due to their popularity and ability to incorporate a wide variety of input types [35, 53], *e.g.*, complex terrains and buildings in the dispersion pathway, a Gaussian plume model is selected for our work. Box models are dismissed due to their over-simplicity and since advanced chemical interactions are neglected due to high computational cost. More precisely, AERMOD [21], a steady-state Gaussian plume model, also deployed by the United States Environmental Protection Agency to assess air pollution, is chosen for this study. The key deciding factors are the availability of the source code, good documentation, the support for multiple emission sources, and achievable input data requirements (*i.e.*, to ensure the availability of all required inputs to run the model).

AERMOD has been applied and validated for a wide variety of conditions: flat and complex terrains [17, 18, 85], various time scales [119] and emission sources, such as a cement complex [101] or a coal-fired power plant [74], and many more available in the literature. To account for the effect of buildings on the dispersion of air pollutants, AERMOD incorporates a validated downwash model to capture relevant turbulence effects [86].

## 2.5 The World Avatar dynamic knowledge graph

As introduced by Akroyd et al. [1], the World Avatar (TWA) project aims to create a digital ‘avatar’ of the world. This vision of an all-encompassing world model is currently worked towards using Semantic Web technology, following a general-purpose dynamic knowledge graph (dKG) approach [69].

The Semantic Web [13] is an extension of the World Wide Web with the aim of creating an interoperable "web of data", making web content machine-readable by adding structured metadata. It builds on the use of ontologies and the Resource Description Framework (RDF) [63] for representing such metadata. An ontology represents a conceptual descriptions of a specific domain by providing formal and explicit definitions of relevant concepts, properties, and relationships between them. Using strict formalisation, ontologies ease unambiguous data sharing and reuse, and enable automation, reasoning, knowledge discovery as well as inference of implicit information. Representing data using ontologies results in the formation of directed graphs, known as knowledge graphs (KGs), where nodes define concepts, instances, or data, and edges denote their relationships (*i.e.*, properties). KGs provide extensible data structures well suited to represent arbitrarily structured data. Using Internationalised Resource Identifiers (IRIs), KG resources can be uniquely identified, allowing data to be decentralised, *i.e.*, distributed across the web, while maintaining unambiguous links between entities. The approach of Linked Data [12, 14] supports FAIR data principles [110] and amplifies the discoverability of information. Additionally, concepts or relationships can ultimately be traced back to their original definitions. Knowledge graphs can be stored in graph databases, such as RDF4J or Blazegraph [15]. Graph databases are designed to host RDF data, and thus KGs, in the form of subject-predicate-object triples and can be queried and updated using SPARQL [4], a query language designed to interact with semantic information.

Beyond the capabilities of conventional KGs, such as DBpedia or Wikidata, TWA also includes semantically annotated computational capabilities, so-called agents, which operate upon instantiated entities and make the graph inherently dynamic. Computational agents within TWA can be seen as executable knowledge components and perform diverse tasks, such as ingesting real-world data, performing calculations, updating the graph, or transmitting responses to the physical world. By introducing agents as integrated part of the KG, computational capabilities also become discoverable [117]. This means the graph not only provides information about its data, but also reveals potential actions.

Additionally, the derived information framework (DIF) [8] has been proposed as KG-native solution to track data dependencies and manage information flow within TWA. Offering granular data provenance on an instance level, it provides details about the origin of any information and the agent responsible for its acquisition. By representing intrinsic dependencies within the KG, the DIF enables autonomous data handling, allowing information to cascade automatically across the graph.

The combination of ontological descriptions, instantiated data, and autonomous agents makes TWA a powerful, extensible, and FAIR-compliant system for representing and reasoning about complex domains of knowledge. As everything is connected (*i.e.*, data, concepts, and agents), the design creates an interoperable ecosystem of connected digital twins (*i.e.*, tools and services) to describe the behaviour of complex systems of systems. TWA is modular and scalable by design, supporting both decentralisation and interoperability across heterogeneous data sources and software. The knowledge-model based approach aims to provide a technology agnostic and distributed architecture based on open standards and protocols to ensure secure data sharing for both private and public data [28].

## 2.6 Existing ontologies

This section briefly summarises existing ontologies relevant to this work: Rijgersberg et al. [98] have developed the meanwhile widely adopted ontology of units of measure (OM) to represent (measured) quantities, their numerical values, and associated units. For geospatial data, the Open Geospatial Consortium has published several encoding standards, including GeoSPARQL [80] and CityGML [81]. GeoSPARQL forms the de-facto standard for representing and querying geospatial data on the Semantic Web and provides an extension to the SPARQL query language for processing geospatial data. The prevalent approach for handling CityGML data involves storing them in the 3DCityDB database as relational tables without explicit semantics; however, ontology-based data access, as advocated by Botoeva et al. [16] and implemented via tools like Ontop [115], can be used to access such structured data as virtual knowledge graph compliant to underlying ontologies. Ding et al. [30] have proposed a set of declarative mappings to expose 3DCityDB buildings as CityGML ontology concepts, making instantiated buildings available as KG and supporting key GeoSPARQL functions.

**Time series data** Numerous ontologies have been proposed to represent temporal concepts and/or time dependent measurements of (physical) quantities: The time ontology proposed by the World Wide Web Consortium (W3C) [22] provides a vocabulary for describing temporal properties of resources in the world or the web (*e.g.*, instants, intervals, durations). 4D ontologies, like the UK Government's Information Exchange Standard 4 [106], have emerged to honour the fact that certain relationships may only apply to a certain locations and/or phase of an entity and develop over time. While these ontologies offer precise representations in spacetime, simpler alternatives exist for more specific use cases.

The Semantic Sensor Network ontology [59, 112] introduces concepts and relationships to capture the observation time associated with any measurement, reusing concepts from the W3C time ontology. The Smart Applications REFerence ontology (SAREF) [37] adopts a comparable approach by associating timestamps with entities (*e.g.*, a measurement), again reusing temporal entities from the W3C time ontology. However, both approaches focus on a semantically rich representation of measurements and their corresponding observation times rather than an efficient storage of actual time series data. Representing large amounts of time series data as individual triples with full semantic markup can lead to performance issues due to the sheer volume of (partially redundant) statements. This gap has been addressed by a domain extension to SAREF [38], acknowledging that measurements can be either single values or dedicated time series. However, it mandates the arrangement of data in a successive and equally spaced sequence of points in time.

**District heating networks** Becker et al. [9] have proposed a utility network application domain extension for CityGML, introducing concepts for modelling diverse networks in 3D city models, including electricity, freshwater, wastewater, gas, or telecommunication networks. The data model incorporates detailed 3D topography, topology, and functional properties, enabling comprehensive geospatial surveys such as collision analysis and leakage detection. Xu and Cai [116] expanded upon this work, incorporating additional concepts from domain glossaries through natural language processing to create a broad domain ontology for utility infrastructure with high interoperability. The ontology captures both network level (*e.g.*, links between components and their functions in the overall system) and component specific information (*e.g.*, material, elevation, geometry), and focuses on water network related terms for an initial proof-of-concept. Similarly, El-Diraby and Osman [34] proposed an ontology, offering a hierarchical structure of core concepts in the utility infrastructure domain across various media.

The feasibility to automate the optimal coordination of district energy resources using ontologies has been studied by Hippolyte et al. [48]. The work introduces a socio-technical ontology to conceptualise district heating networks, incorporating buildings, sensing and actuation infrastructure, and stakeholders. Web-based services for real-time decision-making have been outlined, positioning the ontology as an intermediary layer between high-level energy management applications and local devices. Li et al. [67] further extended this work to demonstrate its efficacy for real-time optimal control of heat generation, employing a three-layer framework: a sensing and actuation layer, an interoperability layer with semantic data models, and an intelligence layer. Although offering valuable conceptualisations of various district heating network related aspects, reusability is limited as none of these ontologies is publicly available.

**Dispersion data** Eibeck et al. [33] and Zhou et al. [117] presented an initial KG-based attempt to model air pollution dispersion based on power plant emissions. Their work links ontologies from different domains such as chemical plants, weather, and buildings to create dispersion profiles in both Berlin and The Hague. However, no GeoSPARQL is used to represent geospatial features, which makes it difficult to extend their work to different locations in the world. Additionally, the authors do not provide methodologies for storing pollution concentration raster data, an essential aspect we aim to address in this work. Metral et al. [72] attempted to link air quality models and CityGML data, focusing mainly on representing air quality models using ontologies. As we aim to add semantic information to data generated by dispersion simulations to foster cross-domain interoperability between models, their work cannot be directly reused.

Dispersion simulations produce pollutant concentration as a function of coordinates, forming a type of raster data. A number of studies have investigated semantic representation of raster data, but none of them proposed a viable solution [35, 53, 62, 82]. Moreover, the current version of GeoSPARQL [80] only supports vector and not raster data types. It is highly challenging to support raster data in RDF format, as each raster cell is associated with one or more values. Materialising one triple per raster value requires unnecessary high storage and will result in very poor scalability [10, 43]. There is a promising GeoSPARQL extension called GeoLD [2] which was developed to support raster data; however, the system is in the early development stage and, hence, not considered here. Furthermore, it currently does not support aggregation functions (*e.g.*, sum, min, max) which are essential for processing raster data. Lastly, GeoLD requires Rasdaman [93], a specialised database for raster data which is currently not supported by TWA.

**Building energy data and master planning** Reinisch et al. [95] have proposed an ontology to represent home energy consumption; however, multiple essential concepts for urban building energy modelling, such as heating demand or solar potential, are absent. Cuenca et al. [23] have proposed an ontology to represent energy management information in an attempt to unify previous ontologies for energy performance and contextual data, such as building and infrastructure information; however, the ontology lacks sufficient level of detail to describe solar technologies as often required in urban building energy modelling. Moreover, the ontology is not publicly available. Cuenca et al. [24] have proposed another ontology for the energy management domain, including common vocabularies for relevant sub-domains. Despite capturing key energy equipment concepts, the ontology lacks important solar device definitions.

Chadzynski et al. [19] have proposed an ontology based on CityGML 2.0 to represent urban environments, such as transportation networks and buildings. The ontology includes classes to describe buildings in different levels of detail, from simple footprints to detailed architectural models, as defined by CityGML. The ontology captures relevant concepts for city master planning, but lacks a building energy modelling extension. Similarly, both energy ontologies [23, 24] lack interoperability with master planning.

# 3 Implementation

The goal of developing connected digital twins is to operationalise actual real-world data. To harness the advantages of the Semantic Web and Linked Data, it is necessary to develop ontologies as foundational knowledge models. Existing ontologies, as introduced earlier, are reused where applicable, and new ontologies are proposed for identified gaps. This approach honours existing domain expertise and ensures compatibility with established community understanding, while satisfying requirements of the provided data and target use case. The proposed agents exchange data and interact with one another via instantiated ontologies in the dKG, promoting a shared understanding among all actors. In contrast to conventional API-based methods, the explicit semantics of ontologically defined communication establish a uniform, public, and well-controlled framework for expressing and interpreting data. This eliminates ambiguity in data representation, allowing agents to navigate and comprehend information consistently.

## 3.1 Ontologies

Within TWA, ontologies serve as modular components to represent and connect diverse concepts across domains. The used ontologies are supposed to offer a balanced approach between generality and specificity tailored to the target use case. After reviewing the literature, it became clear that prior ontology efforts are either not publicly available (see section 2.6) or do not adequately address the required level of detail in the domain of interest. Consequently, four novel ontologies are proposed to represent time series data as well as relevant aspects of district heating network operations, emission dispersion and the building energy domain. While developing these ontologies, concepts of available ontologies have been reused to the extend possible to comply with the Linked Data paradigm.

We adopt a bottom-up approach in developing our ontologies, with a primary focus on representing real district heating operations data from our industrial partner, Stadtwerke Pirmasens, and capturing outputs from the City Energy Analyst. However, a sufficient level of generality is maintained to foster reusability beyond the target use case. The consistency of all proposed ontologies has been verified using the HermiT reasoner [25]. For a formal representation of the ontologies using description logic [7], please refer to Appendix A.6. The codified [109] versions are publicly accessible on GitHub in OWL format.

### 3.1.1 Time series ontology

This work elaborates an initial approach for a light-weight time series ontology [51], primarily to include a description of forecasts and how they have been derived. The key structure of the ontology is provided in **Fig.** 1.

While SAREF's extension [38] acknowledges that any entity (*e.g.*, measurement) could be represented as either single value or time series, our approach, furthermore, considers cases where entities also have associated time series forecasts. Hence, the domain of `hasTimeSeries` and `hasForecast` remains unconstrained. The `Forecast` concept is the central entity to represent any forecast and is associated with both a `TimeSeries` concept holding the actual predicted values and further meta data about the forecast to ensure proper provenance information. This includes the `ForecastingModel` used to derive the prediction, the length of the historical time series used for training and/or scaling, and the forecast horizon. Concepts from OM and the W3C time ontology are used to represent corresponding units and temporal entities, such as the interval of the forecast horizon. The `ForecastingModel` concept captures key aspects of how forecasts are calculated, including the used training `TimeSeries` to fit the model, potential covariates to be used when creating a forecast, and whether the data shall be scaled when creating predictions (*i.e.*, as required by many neural methods). Previously fitted models can be

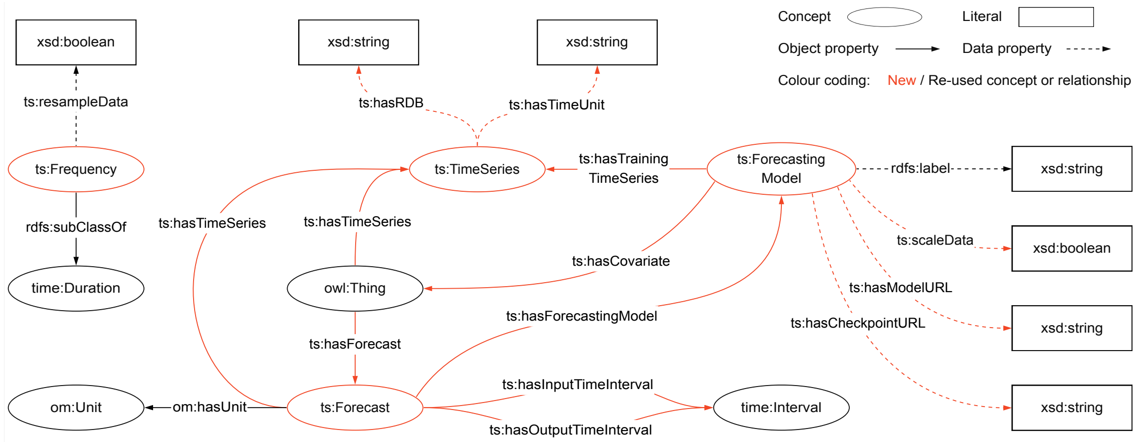



**Figure 1:** *Time series ontology. The OntoTimeSeries ontology provides a light-weight representation for time series data within TWA. It further includes a general description of related forecasting concepts. All referenced namespaces are declared in Appendix A.1.*

incorporated by specifying resolvable URLs for both the saved model and checkpoint files (*e.g.*, pickled pytorch models). Otherwise, default forecast models can be specified using a certain label. Further details on how the *forecasting agent* uses the ontology are provided in section 3.2.1 below. In contrast to the SAREF extension [38], no explicit restriction on the frequency of the represented data is imposed; however, a `Frequency` concept is included, along with a `resampleData` property, indicating whether a time series needs to be resampled when creating a forecast (*e.g.*, to comply with frequency requirements of certain forecasting techniques).

### 3.1.2 District heating network ontology

This ontology aims to conceptualise district heating network operations and has been designed based on actual operations data from Stadtwerke Pirmasens. While previous works often focus on rather static topology and 3D representations [34, 116], the shared information does not contain detailed geo-references to describe the grid structure (*e.g.*, pipes, connectors). Moreover, the operations of the grid are rather dynamic, including customers' heat demand and flow temperature profiles. As the work by Li et al. [67] is not publicly available for direct reuse, this novel ontology is proposed to capture essential aspects of district heating operations, leveraging some design choices from previous works.

The three key concepts in the ontology are `HeatingNetwork`, `HeatProvider`, and `HeatGenerator`. The `HeatingNetwork` connects `HeatProvider` with `Consumer` instances to satisfy their `HeatDemand` (*i.e.*, instantiated as time series). While the location of an individual `Consumer` presently remains undisclosed and is, hence, not explicitly modelled, any `HeatProvider` is connected to the grid via a `GridConnection` with observable properties. These properties include pressure as well as flow and return `om:Temperature`, and provide insights into the temperature spread at these locations and, con-

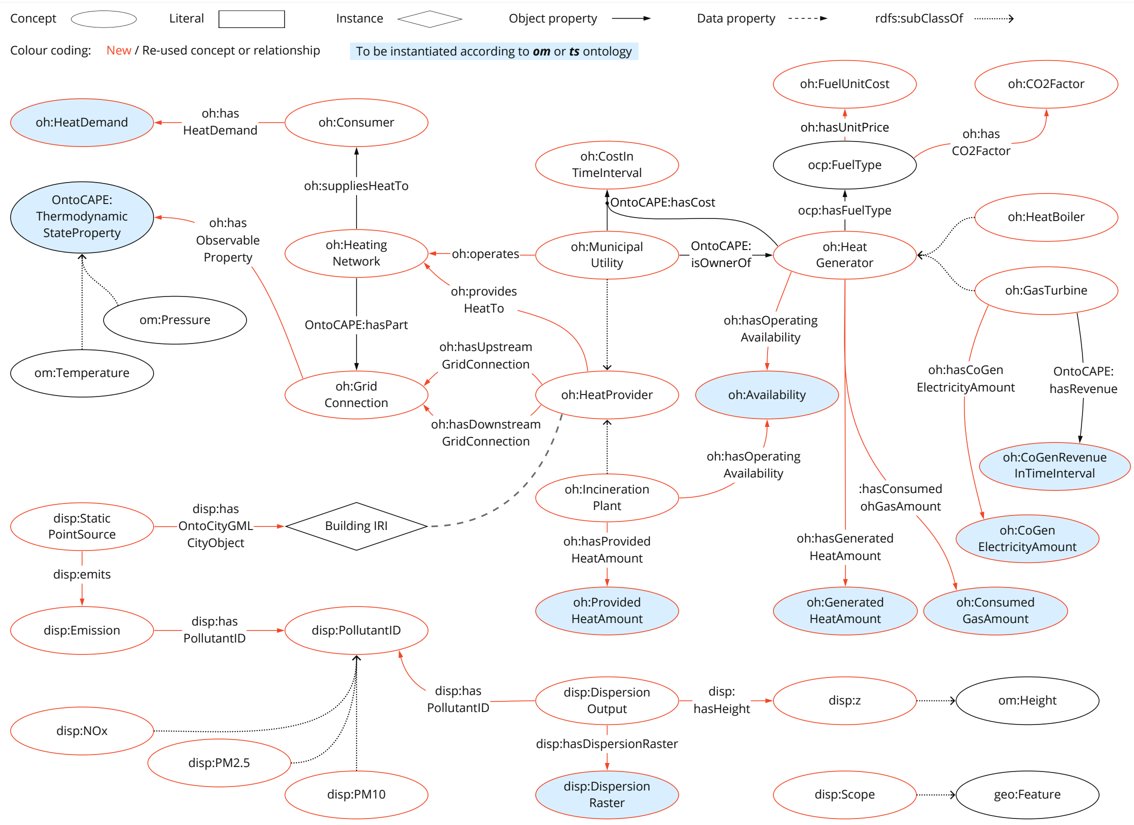



**Figure 2:** *Coupled heating network and dispersion modelling ontology. The OntoHeat-Network ontology conceptualises the operations of a district heating network, while OntoDispersion provides a semantic description of air pollutant emission dispersion. Both ontologies can be linked through shared building instances to support geospatial emission analyses of district heating operations. All referenced namespaces are declared in Appendix A.1.*

sequently, corresponding feed-in heat amounts.

While the `HeatProvider` concept is kept general and represents any entity supplying heat to the grid, two subclasses relevant to the given use case are defined, namely `MunicipalUtility` and `IncinerationPlant`. The geospatial location of a `HeatProvider` is not explicitly modelled as part of this ontology. Instead, geo-references are established via links to corresponding building instances with detailed geometrical and geospatial information as part of the derivation markup as described in section 3.3 and illustrated in **Fig.** 2. This approach keeps the OntoHeatNetwork ontology completely free of geospatial information, by reusing the capabilities of existing ontologies.

An `IncinerationPlant` provides a certain `ProvidedHeatAmount` to the grid based on a supply contract with the grid operator (for details see **Fig.** 22 in the Appendix). A `MunicipalUtility` company can own multiple `HeatGenerators`, including conventional `HeatBoilers` and combined heat and power (CHP) `GasTurbines`. Each `HeatGenerator` is associated with a `GeneratedHeatAmount` and a corresponding `ConsumedGasAmount` concept, according to the generator's efficiency and used `ocp:FuelType`.

Relevant costs are represented on both an individual generator and operator level. As $CO_2$ emissions directly influence operating expenses (OPEX) due to emission certificate cost, they are modelled explicitly as part of OntoHeatNetwork, while other air pollutants are conceptualised as part of OntoDispersion. Similarly, the electricity co-generation of a `GasTurbine` is captured, since respective revenue offsets heat generation OPEX. A detailed overview of the hierarchical cost structure and its components is provided in **Fig.** 21 in the Appendix. An `Availability` concept is introduced to account for periods of plant shut-downs or required idle times of individual generators. Most properties will be instantiated as time series to account for dynamic conditions (*i.e.*, fluctuating prices, time-dependent heat demand) and to align with the hourly-resolved optimisation strategy applied by the municipal utility operator in the target use case.

#### 3.1.3 Dispersion ontology

This light-weight ontology aims to provide semantic markup for dispersion simulation data to create machine-readable inputs and outputs for agents and to foster cross-domain interoperability between various models. The key concepts of OntoDispersion are located in the bottom of Fig. 2, including their intended link to the OntoHeatNetwork ontology.

A geospatial `Scope` concept specifies the simulation domain for the dispersion calculation and is defined as a subclass of `geo:Feature` to enable various geospatial querying and processing capabilities via GeoSPARQL. By using concepts from GeoSPARQL, the ontology is designed to be as robust as possible and easily extendable to different areas. Versatile geospatial capabilities are essential to query which `StaticPointSource`s, emitting one or more pollutant types, are located within a certain scope of interest. `StaticPointSource`s link to corresponding building instances, which describe the actual geometries of the emission outlets. Each instance of `DispersionOutput` holds information on a set of raster data (`DispersionRaster`) for any arbitrary combination of pollutant type (`PollutantID`) and simulated height (`z`). In this work, we do not attempt to materialise raster data as RDF triples. Instead, any `DispersionRaster` instance simply provides the metadata (*e.g.*, name of a GeoTIFF file) of a raster stored in an associated PostGIS database. Thus, an agent querying for dispersion raster data would obtain the metadata via a SPARQL query and perform a subsequent SQL query to obtain the underlying raster values.

#### 3.1.4 Building energy ontology

Energy considerations are important in master planning, and buildings are the main user of urban energy. There are ontologies for urban building energy modelling [23, 24, 95] and master planning [19], but none really links the two fields. This ontology aims to bridge this gap to facilitate information exchange between these closely related domains.

An extract of the proposed OntoUBEMMP ontology is shown in **Fig**. 3. The key concepts are `dabgeo:Building`, `EnergyConsumption` and `EnergySupply`, while the core building concept is shared with the OntoBuiltEnv ontology to facilitate interoperability between this energy-specific perspective and a more comprehensive building description

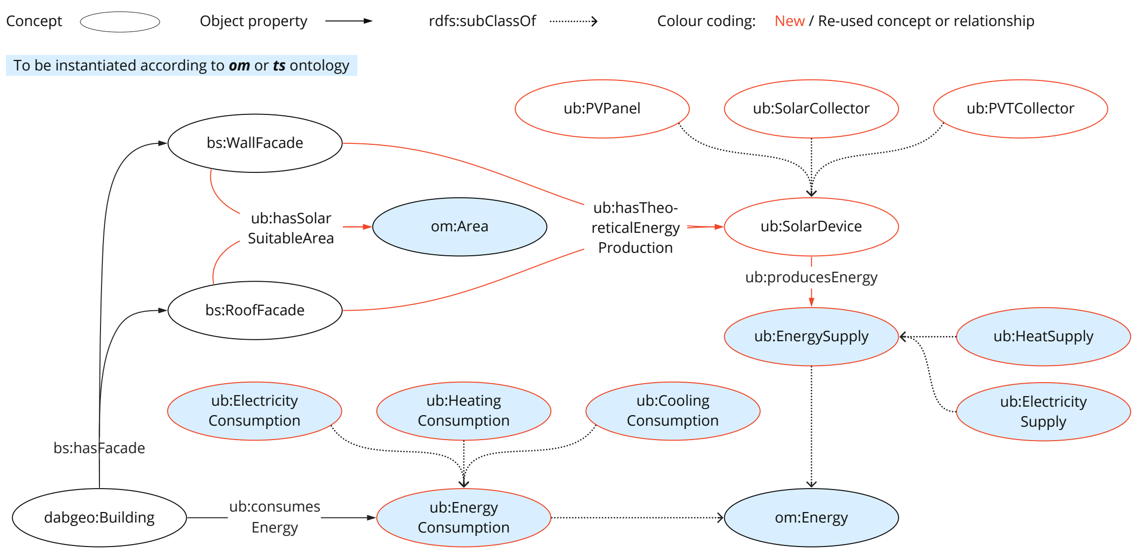



**Figure 3:** *Urban Building Energy Modelling and Master Planning ontology. The OntoUBEMMP ontology represents key concepts in the nexus of urban energy modelling and master planning, including building energy demands and solar potentials. All referenced namespaces are declared in Appendix A.1.*

provided by OntoBuiltEnv. An `EnergyConsumption` concept is linked to its applicable building instance via a `consumesEnergy` relationship to represent a building's energy demand. Renewable energy sources should be taken into consideration during master planning to help offset `EnergyConsumption`. For example, a building can be equipped with `SolarDevices` on its `bs:RoofFacade` and its `bs:WallFacade`. For buildings with suitable areas for solar generation (*i.e.*, `hasSolarSuitableArea` links to a non-zero area), the `hasTheoreticalEnergyProduction` relationship connects relevant areas with their potential `EnergySupply` via the respective `SolarDevice`. There are different subclasses of `EnergySupply`, namely `ElectricitySupply` and `HeatSupply`, depending on the type of `SolarDevice` that could be installed. Installation of `PVPanel` will generate `ElectricitySupply`, whereas `SolarCollector` will generate `HeatSupply` and the hybrid `PVTCollector` will generate both. The ontology uses OntoTimeSeries to instantiate `EnergyConsumption` and `EnergySupply` concepts as well as their subclasses to account for variable demand patterns or changing weather conditions.

## 3.2 Agents

Several agents have been developed to operationalise the proposed ontologies. An overview of all involved agents is provided in **Fig.** 4 and described below. All agents are packaged as individual Docker services to foster distributed and platform-agnostic deployment (*e.g.*, remotely in the cloud, as implemented for this use case).

We were given actual historical operations data for a municipal district heating network of a midsize German town, Pirmasens. Based on this data, the district heating grid is instantiated per OntoHeatNetwork, using 2020 historical time series data. Utilising the instan-

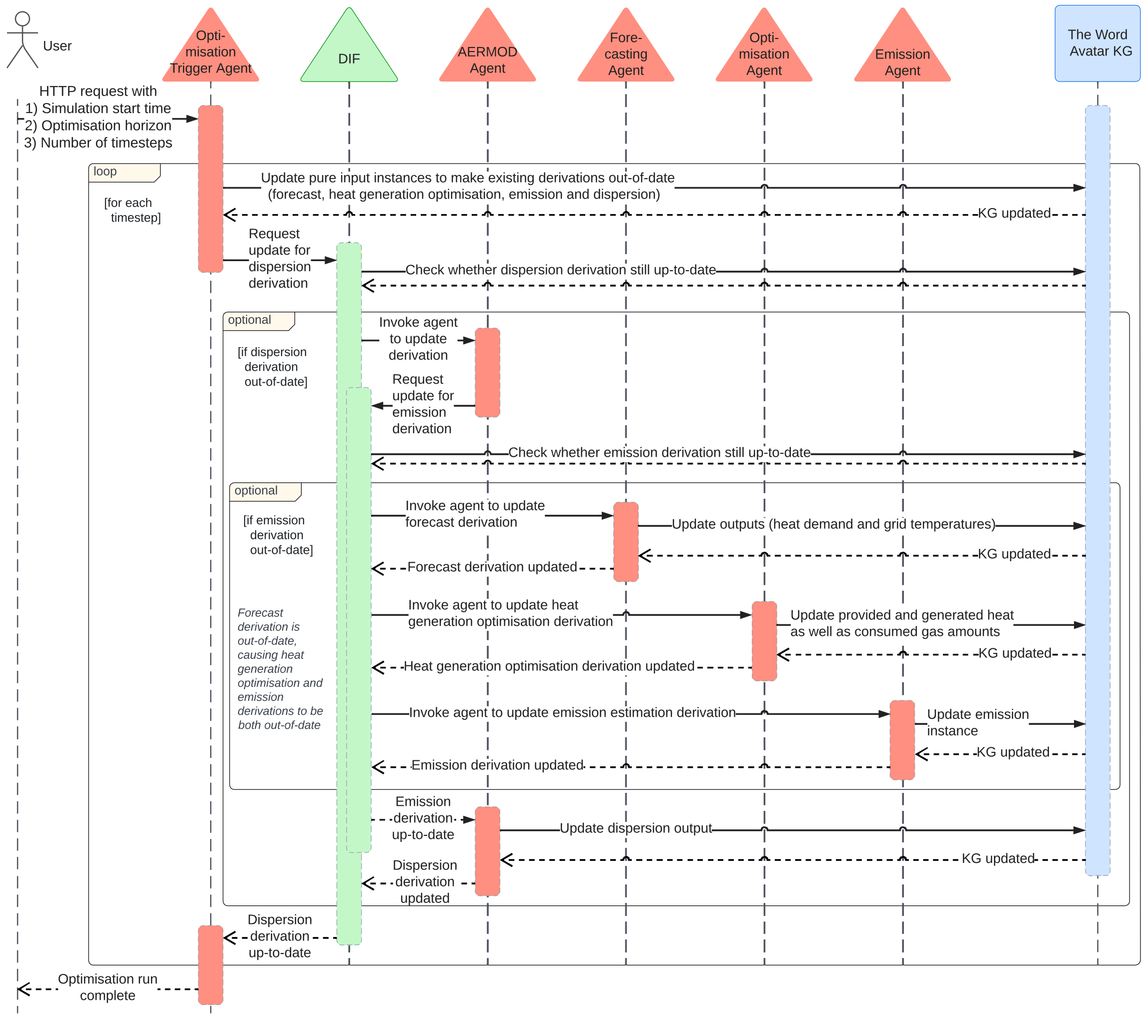



**Figure 4:** *Agent interplay. Sequence diagram of all agents involved in the heat generation optimisation with integrated emission dispersion modelling (depicted for case of already instantiated derivation markups).*

tiated time series data, a *forecasting agent* can be used to predict any quantity, including the community's `HeatDemand`. A *district heating optimisation agent* then generates a cost-optimised generator dispatch strategy to satisfy the forecast `HeatDemand`, considering both internal heat generators and sourcing from an external waste incineration plant. The respective amounts of burned natural gas as well as heat from waste incineration are then converted by an *emission estimation agent* into corresponding $NO_x$, $PM_{2.5}$, and $PM_{10}$ emission streams. Together with the associated location information, these emission streams form inputs to an *dispersion modelling agent* to create a steady-state emission dispersion map using actual historical wind data. All agents are implemented as derivation agents based on the DIF [8] and communicate directly via the dKG to ensure unambiguous provenance tracking of how a certain output has been derived and which inputs it depends on. We introduce an *optimisation trigger agent* as coordinating link between a user and the automated forecasting, optimisation, and subsequent emission dispersion simulation.

While the dynamic load forecasting and supply-side optimisation use actual historical data, the *City Energy Analyst agent* can provide general insights into the energy performance of buildings in case historical data is not available: utilising building-specific construction characteristics and weather data, various energy demand and generation profiles can be estimated. This complementary perspective provides valuable insights into building-resolved heat demands, *e.g.*, relevant to analyse any potential extension of the district heating grid.

### 3.2.1 Forecasting agent

This agent provides generic forecasting capabilities as part of TWA: it can retrieve instantiated time series, predict future values, and instantiate respective forecasts back into the dKG using the OntoTimeSeries ontology. Based on the Python library Darts [47], the agent supports forecasting via a wide range of methods, ranging from classical white box models (*e.g.*, established statistical methods such as autoregressive integrated moving-average (ARIMA) models and its derivatives) to black box machine learning techniques (*e.g.*, state-of-the-art transformer models), as well as grey box approaches such as Facebook's Prophet [104].

The required input instances to derive any forecast comprise the instance associated with the time series to predict, a `ts:ForecastingModel` describing the prediction model to use, the target `ts:Frequency` of the forecast to be created, a `time:Interval` denoting the target forecast horizon, and a `time:Duration` denoting the historical data length to use for fitting and/or scaling of the historical time series data prior to creating the forecast. New forecasts are instantiated with relevant metadata, such as input and output time intervals as well as potentially applicable unit, as depicted in Fig. 1.

The agent supports most forecasting models from Darts by instantiating a `ts:ForecastingModel` instance with the corresponding properties. If an instance is labelled "Prophet", it utilises Prophet [104] for forecasting, which represents the default model for arbitrary time series. Predictions with and without covariates are supported, depending on whether `ts:hasCovariate` relationships are present for the target model instance. Additionally, custom models can be trained and stored within TWA for future forecasting. This involves creating custom `ts:ForecastingModel` instances with specific properties, such as resolvable URLs for saved model files, relevant covariates, and scaling parameters. Thus, the agent offers both out-of-the-box forecasting capabilities and the flexibility to leverage custom fine-tuned models as needed.

The agent can be used to predict any arbitrary time series and is deployed to forecast heat demand and grid temperatures of a district heating network, using fine-tuned temporal fusion transformer (TFT) [68] models in the context of the current work. Further implementation details as well as a comparison of the trained TFTs and previously fitted SARIMAX models [50] are provided in Appendix A.4.1. Compared to many other deep learning methods, attention-based TFTs provide better explainability and interpretability thanks to insights into underlying attention weights, which outline what a model focuses on. For instance, it can be observed that the heat demand model shows a clear daily seasonal pattern, with increased attention to the last few hours, which aligns well with prior expectations (see Fig. 23 in the Appendix).

### 3.2.2 District heating optimisation agent

This agent leverages a previously developed optimisation routine to minimise total heat generation cost for a district heating provider [50]. The optimisation follows a hierarchical approach based on a merit-order principle to determine the OPEX-optimised short-term heat generation mix for a system comprised of multiple gas boilers, a CHP gas turbine as well as external heat sourcing from a waste incineration plant. While the effectiveness of the solution has been demonstrated elsewhere based on real-world operations data [50], this agent integrates the existing model semantically into TWA.

While the initial optimisation relied on internally created SARIMAX predictions for key inputs, this agent increases modularity and fosters a micro-service architecture enabled through task-oriented connected digital twins by using externally instantiated forecasts. The agent requires five `ts:Forecast` and one `time:Interval` instance to perform an optimisation. The interval specifies the optimisation horizon, describing the period for which to derive the optimal dispatch strategy, while the five forecasts denote the forecasted `oh:HeatDemand` and four `om:Temperatures` (*i.e.*, representing flow and return temperatures at the waste incineration and municipal heating plant) over this period, respectively. Besides these key inputs, further information are queried from the dKG during agent operation. Upon successful optimisation, the following results are instantiated back into the dKG according to the OntoHeatNet ontology: a `oh:ProvidedHeatAmount` instance describing the heat amount to be sourced from the waste incineration plant; an `oh:GeneratedHeatAmount` and `oh:ConsumedGasAmount` instance for each gas boiler and CHP gas turbine denoting the heat amount to be provided and corresponding gas amount to be consumed by each heat generator, respectively; an `oh:CoGenElectricityAmount` instance describing the amount of co-generated electricity by the gas turbine while providing the required amount of heat; an `oh:Availability` instance for each heat provider indicating its anticipated availability in the coming time steps. All optimisation outputs are instantiated as `ts:Forecast` instances for the respective concepts to not interfere with instantiated actual historical data. Newly created optimisation outputs automatically overwrite previously instantiated ones.

Upon first invocation of the agent, historical gas consumption, heat generation, and (if applicable) electricity generation data is queried to fit data-driven generator specific gas consumption and co-generation models to be used during the optimisation. These models will be reused for all subsequent optimisation requests. The results of the agent implementation have been verified against previous optimisation results [50].

### 3.2.3 Emission estimation agent

This agent estimates the emission rates associated with heat production from burning natural gas (*i.e.*, in gas boilers or the CHP gas turbine) or waste (*i.e.*, in the waste incineration plant). For the time being, the assessment is limited to $NO_x$, $PM_{2.5}$, and $PM_{10}$ as the major airborne emissions [39, 40, 111] and relies on literature-based emission factors instead of detailed combustion models for this proof-of-concept and due to the absence of detailed information on the waste incineration plant internals.

Implemented as a derivation agent, all required inputs need to be available in the KG. This includes one `dh:ProvidedHeatAmount` or, alternatively, one or more `dh:ConsumedGasAmounts` representing the (optimised) time series for externally sourced heat or consumed gas amounts, respectively. A collection of consumed gas amounts resembles multiple gas boilers and gas turbines housed within the same building, emitting exhausts through a shared chimney. A `disp:SimulationTime` marks the timestamp for which to estimate the emissions, *i.e.*, for which later to simulate the emission dispersion. Lastly, a `disp:StaticPointSource` instance specifies the location at which the estimated emissions will be emitted.

During assessment, the time series values for provided heat or consumed gas corresponding to the target `disp:SimulationTime` are extracted. If multiple `dh:ConsumedGasAmounts` are given, their individual values are added together and processed collectively. Subsequently, emission factors are applied to convert the energy amounts into corresponding mass flow rates for $NO_x$, $PM_{2.5}$, and $PM_{10}$, as required for the air pollutant dispersion simulation. The flue gas stream is treated as hot air, using typical values from the literature. Refer to Appendix A.4.2 for detailed information on the estimation methods. All outputs are instantiated according to the OntoDispersion ontology as `disp:Emission` instances with associated quantities for mass flow rate of pollutant as well as temperature and density of the exhaust stream. One emission instance per pollutant type is created.

### 3.2.4 Dispersion modelling agent

This agent utilises AERMOD [21, 85] to simulate the dispersion of various air pollutants in a specific area of interest. It considers instantiated wind and emission stream data (*i.e.*, mass flow rate, temperature) from multiple point sources to generate emission concentration maps. Upon invocation, the agent performs three key steps: querying relevant inputs from the KG, executing AERMOD using this information, and finally instantiating the results back into the KG. As shown in **Fig.** 5, the key inputs are `disp:Scope` and `disp:SimulationTime`. The `disp:Scope` defines the polygon of the simulation domain (*i.e.*, a rectangle) and `disp:SimulationTime` determines the time step of interest for which to run the dispersion calculation. Note that this input is shared with the *emission estimation agent*.

The agent requires at least one instance of `disp:StaticPointSource` (*e.g.*, a chimney emitting pollutants) that is located within `disp:Scope` to simulate a plume. Instances of `disp:StaticPointSource` are not linked directly to dispersion derivations in order to facilitate future use cases involving mobile point sources (*e.g.*, ships), which may move in and out of `disp:Scope`, making explicit markups very difficult to maintain. Instead, the agent uses `disp:Scope` to obtain relevant buildings and emission sources within the simulation area for the relevant timestamp. The `disp:SimulationTime` is also used to query the actual (historical) weather data for that given time. Having retrieved all necessary information, the agent composes relevant input files and executes AERMOD. Subsequently, the agent processes the dispersion results into raster form and updates the `disp:DispersionOutput` time series instance in the dKG.

Although we use AERMOD in this work, it could be swapped with any other dispersion model (*e.g.*, EPISODE [44]) with minimal changes to the overall workflow outlined in Fig. 5. It is inevitable that a new agent would need to be developed; however, the proposed ontology would still suffice to represent relevant concepts (*e.g.*, `disp:DispersionRaster`).

### 3.2.5 City Energy Analyst agent

This agent calculates various aspects of a building's energy performance using CEA as its simulation engine. To overcome limitations with built-in CEA assumptions, actual building stock data from the dKG are incorporated to allow for building-specific analyses, namely a building's geometry and usage, the geometry of surrounding buildings, weather, and terrain data. The implementation maintains CEA's broad applicability while adopting a building-resolved bottom-up approach.

Upon invocation, the CEA agent attempts to retrieve the geometry of the target building(s) (*i.e.*, specified by the building IRI(s) in the received HTTP request) as well as the geometries of the surrounding buildings. Retrieved geometries from TWA replace CEA's default OSM footprints. Subsequently, the agent will attempt to retrieve building specific usage data from TWA to produce most meaningful energy consumption profiles. After retrieving building level input data, the agent attempts to retrieve actual weather information at the target location from the dKG. Available local weather data supersedes CEA's default behaviour of interpolating weather information based on a few selected locations within its own database. Lastly, the agent attempts to retrieve terrain data, specifically, the elevation of the land surrounding the target building(s) from TWA, to replace CEA's default terrain input of a fixed elevation. Only a building's geometry is strictly necessary for the agent to run successfully. In cases where other inputs cannot be retrieved from TWA (*i.e.*, surroundings, usage, weather, terrain), it proceeds to run CEA with its corresponding default assumptions. For details please refer to Appendix A.4.3.

After running the simulations, relevant results are instantiated according to OntoUBEMMP in TWA. The various building energy demands (*i.e.*, heating, cooling, electricity, grid) are instantiated as `ub:EnergyConsumption` instances. The agent also provides solar potential estimates for various types of solar generators: PV panels, flat plate and evacuated tube solar collectors, and combined flat plate and evacuated tube PV-thermal collectors. These generators are instantiated as the corresponding subclasses of `ub:SolarDevice`, with their associated energy potentials instantiated as `ub:EnergySupply` entities. The suitable area for installing solar devices is instantiated via the `ub:hasSolarSuitableArea` property.

## 3.3 Connecting agents to create cross-domain interoperability

Individual agents are chained together via their input and output instances using TWA's native derived information framework. This ensures that whenever a specific piece of information is requested from the dKG, all dependent upstream inputs are scrutinised first to determine if they are still up-to-date or require updating before retrieval. We leverage this infrastructure to automatically simulate associated air pollution dispersion whenever a new heat generation optimisation is computed and corresponding emission streams get instantiated.

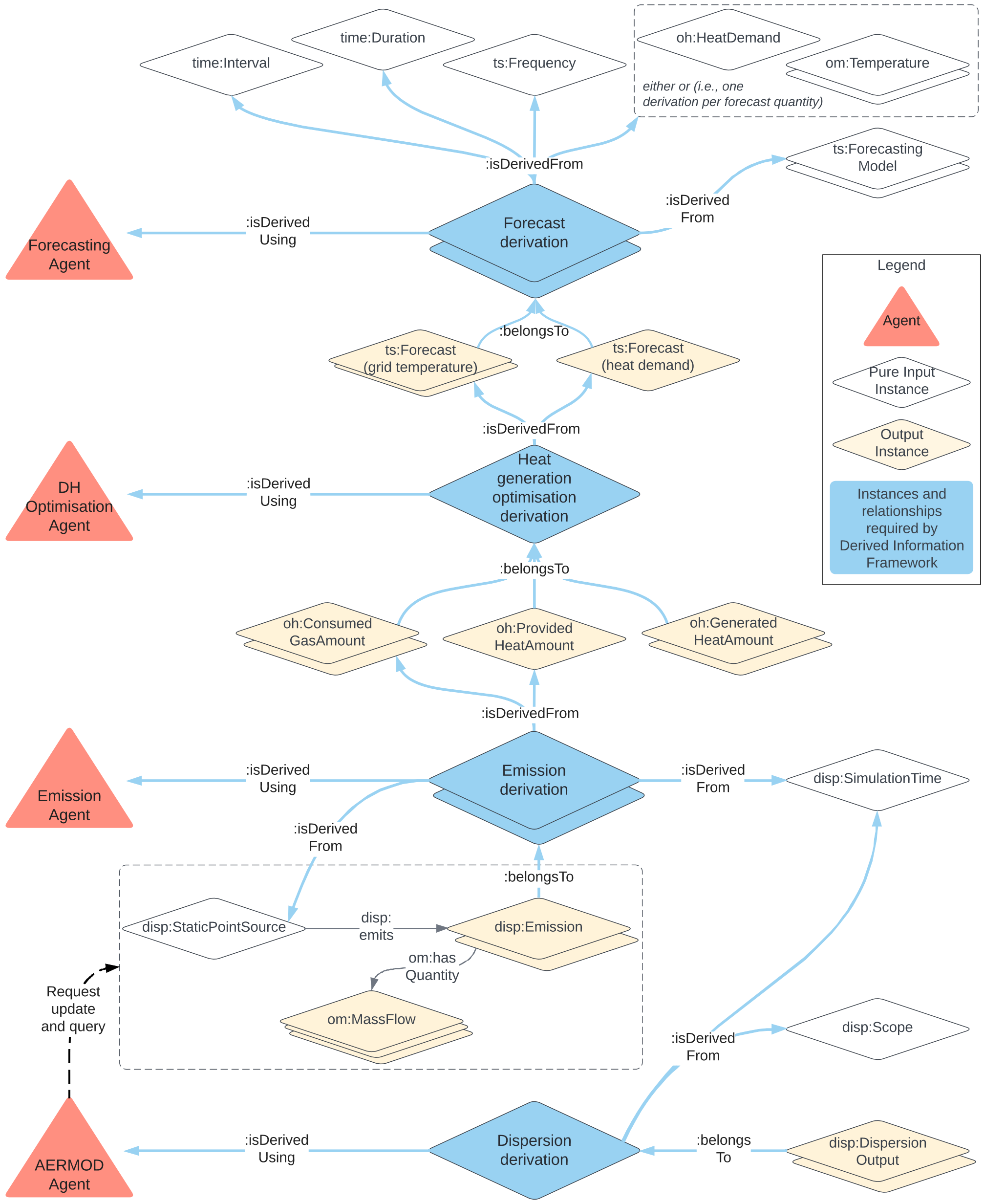



**Figure 5:** *Derivation chain (simplified). Schematic depiction of knowledge graph native instance markup to resemble a model predictive control loop, coupled with automated air pollutant dispersion modelling. All referenced namespaces are declared in Appendix A.1, with not explicitly stated prefixes referring to OntoDerivation.*

As illustrated in Fig. 4, the *optimisation trigger agent* acts as external input and coordination agent. To initiate an optimisation run, an HTTP POST request is expected, specifying 1) the optimisation start time, 2) the optimisation horizon (*i.e.*, the number of time steps to be considered within each optimisation), and 3) the number of subsequent time steps to optimise in total. Upon receiving and verifying a request, corresponding instances are created/updated within the dKG, and an update is requested from the *dispersion modelling* *agent* (also referred to as *AERMOD agent* due to the implemented model) via the DIF. The DIF then assesses whether an up-to-date dispersion instance already exists by comparing the instantiation timestamp of the derivation instance against the ones of corresponding inputs. If necessary, an update is requested, in which case the DIF works backwards through the dependencies: dispersion simulations depend on emission estimation outputs, which depend on the heat generation optimisation, itself dependent on heat demand and grid temperature forecasts. The DIF initiates updates by invoking the associated agents, starting with the *forecasting agent*, responsible for the most upstream derivation, to ensure a proper cascade of all dependent information. Once all information is up-to-date, the initially requested dispersion outputs are simulated, which marks the end of the current optimisation run. This loop is repeated until the number of time steps to optimise is reached.

To ensure automated information cascading, derivation markups need to be instantiated at the instance level, as illustrated in **Fig**. 5. It is important to note that this figure is a simplified representation for readability, with a more detailed diagram provided in **Fig**. 25 in the Appendix. The *optimisation trigger agent* instantiates initial inputs for `time:Interval`, `time:Duration`, and `ts:Frequency` for each requested optimisation run (and updates them accordingly for subsequent time steps). Additionally, the agent programmatically creates the depicted derivation markup if not already present and requests an initial assessment from responsible agents to generate corresponding outputs for further markup.

One derivation instance is created for each required forecast, *i.e.*, one `oh:HeatDemand` and four `om:Temperature` instances denoting the flow and return temperatures at the municipal heat and waste incineration plant, respectively. The derivation outputs (*i.e.*, updated `ts:Forecast` instances) are then collectively marked up as inputs to the heat generation optimisation derivation. After optimising the generation dispatch based on the provided forecasts, multiple outputs are instantiated by the *district heating optimisation agent*, including one `oh:ProvidedHeatAmount` and several `oh:ConsumedGasAmount` instances, representing the time series of external heat provision from the waste incineration plant and gas consumption of several internal heat generators, respectively.

Subsequently, two individual emission derivations are marked up to account for different emission factors used for waste and natural gas burning when estimating associated emission rates. Different derivation instances also account for different locations of the respective pollutant streams, as each derivation is derived from a `disp:StaticPoint-Source`, which introduces a geospatial reference to the dynamic optimisation. The estimated `disp:Emission` outputs are used as source terms by AERMOD to simulate pollutant dispersion maps. Although not explicitly marked up as inputs, the agent requires at least one `StaticPointSource` (*e.g.*, a chimney) within the `disp:Scope` of interest. The `disp:SimulationTime` instance represents the time for which to simulate the emission dispersion, and matches the first time step of the forecast and optimised heat generation.

The *district heating optimisation agent* is designed to handle each optimisation request independently, *i.e.*, without consideration of any preceding requests. The sole exception to this behaviour occurs when two consecutive requests are precisely +1 hour apart. In this scenario, the second request is treated as dependent on the previous one, facilitating the tracking of relevant system state variables, such as cumulative profit from ongoing gas turbine activity. This enables a dKG-native receding horizon optimisation implementation, representing the first model predictive control-style application within TWA. While demonstrated for energy dispatch with integrated emission modelling, similar derivation chains can automate various other (cross-domain) smart city workflows. Although the current implementation relies on an *optimisation trigger agent* as external input agent, this can easily be replaced with autonomous input agents in the future.

# 4 Comprehensive energy perspective

This section unveils the results and insights from our connected digital twin use case. Utilising the developed ontologies and semantic agents, connected via the derived information framework, allows for novel insights and capabilities, such as 1) the knowledge graph-native control of a district heating system, 2) the refinement of building energy analyses with latest instantiated building stock data, and 3) cross-domain insights into heat generation induced air pollutants dispersion.

The World Avatar offers a versatile visualisation interface to explore and interact with the underlying data, and supports both Mapbox (*i.e.*, mainly for geographic information) and Cesium (*i.e.*, mainly for detailed geometrical representations) as well-established visualisation frameworks; however, it is crucial to understand that the presented visualisations are not the digital twin itself. Instead, the digital twins are a dynamic collection of knowledge, data, and models embedded in the dynamic knowledge graph running in the background, with the visualisation being only one way to access it. Further options include a mobile app [97], virtual reality goggles, and a question answering system [118] besides a unified SPARQL endpoint.

The integrated visualisation interface provides both map-based and (real-time) dashboard features. While map-based visualisations can help to understand the geospatial distribution of energy demand or the implications of certain heat sourcing strategies on air pollution, dashboards focus on time series data and offer more details about the current operational state of assets, such as the latest historical and forecast heat demand or the optimised generation strategy to satisfy it.

## 4.1 Resource-efficient heat provision

The municipal district heating network has been instantiated based on actual data, including historical operation, weather, and market conditions. Operations data include details about the grid itself as well as attached heat providers, while market conditions cover electricity spot, gas, or $CO_2$ certificate price time series. Instantiated heating network data comprises the total heat demand profile of all attached customers, operational boundaries of the grid (*e.g.*, minimum volumetric flow rate to ensure hydrodynamic stability), and connection properties for the municipal heating and waste incineration plant, such as observed flow and return temperatures. Plant data includes information about the buildings hosting individual heat generators, along with their design characteristics (*e.g.*, rated thermal power) and dynamic properties, such as time series of generated heat and electricity as well as consumed gas. The integration across scales (*i.e.*, from city level to detailed boiler specifications) as well as the inherent dynamism due to the dKG-native control implementation combines and exceeds the capabilities of isolated energy system modelling and geographic information system-based approaches.

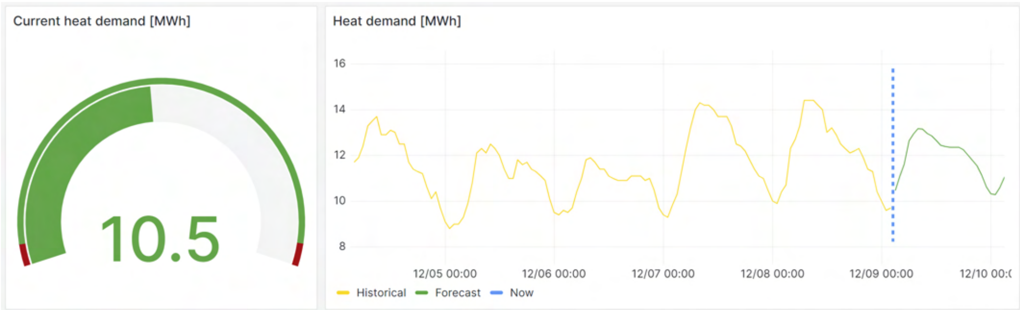



**Figure 6:** *Heat demand forecast. Dashboard view of the latest historical and forecast heat demand at any given time step. The historical load profile is shown left of the dashed line, with predicted values to its right.*


**Figure 7:** *Optimised heat generation. Dashboard view of the cost-optimised heat distribution across generators and external sources, considering a waste incineration plant, three conventional gas boilers, and a gas turbine (based on forecast heat demand).*

**Figure** 6 presents a snapshot of the dynamic heat demand forecast dashboard. It shows the recent municipal heat demand history of all district heating consumers as well as the latest 24-hour demand forecast. The dashboard updates automatically with each new forecast, offering real-time insights into the latest operational state. In addition to time series visualisation (right), a gauge indicates the current state relative to operational/observed minimum and maximum values (left).

**Figure** 7 illustrates the optimal dispatch of three conventional heat boilers, one CHP gas turbine, and external heat sourcing from the nearby waste incineration plant to satisfy the predicted demand (refer to Fig. 6). Currently, the demand of approximately 10.5 MWh is met through external sourcing and one heat boiler, while the remaining heat generators remain idle. Based on the projected demand as well as anticipated electricity spot prices, the CHP gas turbine is expected to be the main contributor to heat production as of in 4 hours, with minor support from the waste incineration plant and one additional gas boiler.

## 4.2 City energy analyses and scenario planning

As actual (historical) energy data are not always available, an alternative approach is needed to estimate relevant quantities and gain insights on a broader scale, such as city level. The CEA agent provides estimates for various aspects of buildings' energy performance, such as demands for different types of energy and on-site solar generation potentials. Compared to the official CEA toolkit, actual building stock (*i.e.*, building geometry, geometry of surrounding buildings, property usage) as well as weather and terrain data are used for the underlying simulations (where available) to derive building-specific

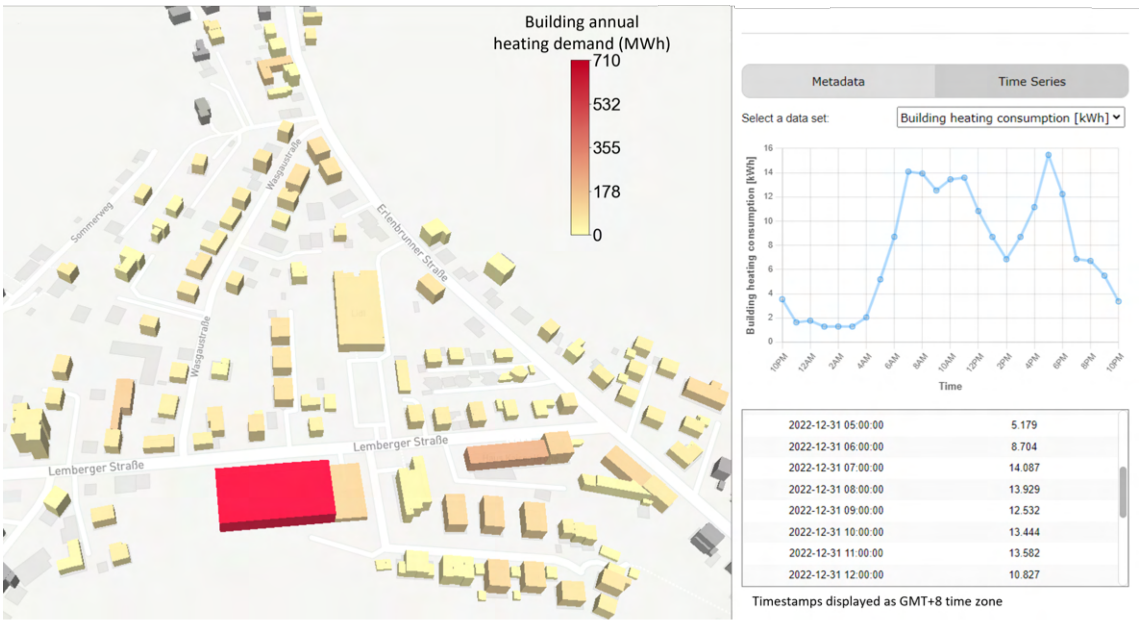



**Figure 8:** *Visualisation of annual heating demand of each building simulated by the CEA agent. While map-based visualisation allows for quick identification of buildings with high/low heating demand, time series support the inspection of load profiles for individual buildings.*

estimates. The outputs of the agent are instantiated and attached to the corresponding building in TWA and can be inspected via its unified visualisation interface. **Figure** 8, for example, shows the annual heating demand for a selected neighbourhood in Pirmasens, allowing for a quick identification of buildings (and areas) with high/low heating demand. Further simulation results, such as photovoltaic potential or gross-floor area specific values are provided in Appendix A.2.3.

While Fig. 8 offers rather qualitative insights, the credibility of the results has been evaluated for both electricity consumption and on-site solar PV potential. The assessment compares instantiated agent results with actual historical consumption data or the official PV potential estimates provided by the state of Rhineland-Palatinate [73], respectively. By leveraging more granular building an weather information from TWA, significant accuracy improvements compared to native CEA (*i.e.*, the unaltered CEA toolkit) can be achieved. The mean absolute percentage error (MAPE) relative to the above benchmarks could be reduced from 57.6% to 13.7% and from 28.1% to 12.9% for annual electricity consumption and solar PV potential, respectively. This improvement outlines the value of our bottom-up approach to remove default assumptions in the underlying CEA toolkit where actual data is available from the dKG.

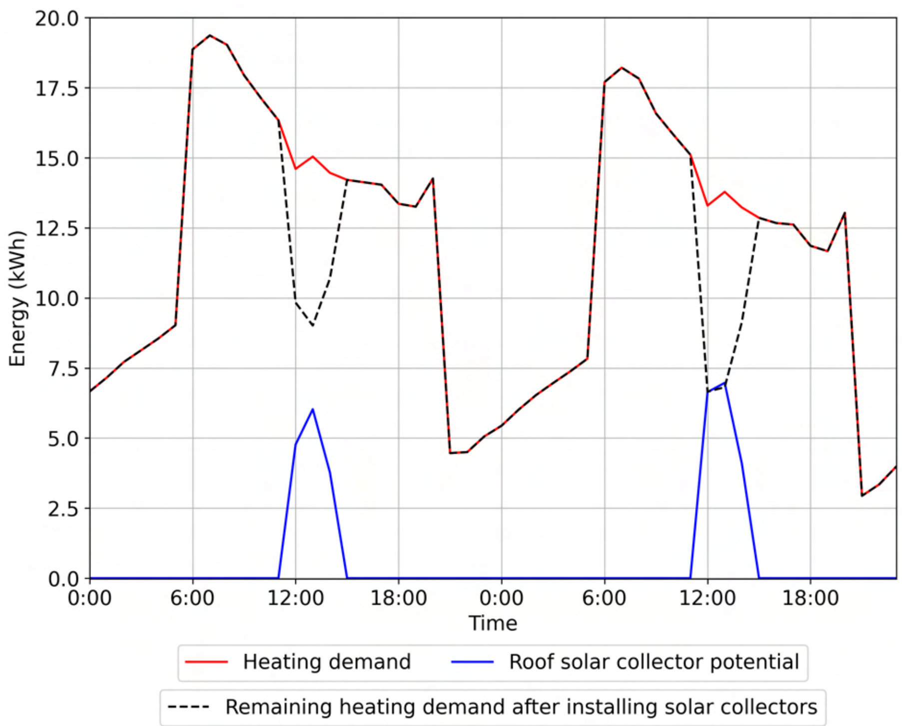



**Figure 9:** *Heating load profiles simulated by the CEA agent. Heating demand and solar generation time series can be used to evaluate potential energy savings of on-site solar collector installations (depicted for a typical day in March).*

Beyond cumulative annual figures, the CEA agent also provides both overall heat demand and solar potential time series, which facilitate the assessment of possible energy savings achievable with the installation of solar collectors. A basic analysis could explore utilising heat from rooftop solar panels to directly offset a building's heating demand, without factoring in any thermal storage. This simplistic assessment provides a preliminary estimate for remaining heating demand from alternative sources such as gas or district heating, together with the potential energy savings conferred by on-site generation (see **Fig**. 9). This capability can help to develop highly-granular heat maps of a city's heating demand (with or without considering on-site generation of solar energy), *e.g.*, as currently required for the municipal heat planning initiative in Germany. The building-resolved insights exceed the accuracy of most publicly available dataset, which are usually restricted to simple raster maps with $50 \times 50$ m or $100 \times 100$ m resolution [54]. Combined with actual district heating grid location data, this information can be used to evaluate potential grid extension scenarios, both with regards to the total geospatially distributed heat demand as well as prevalent heat demand profiles considering actual building usage patterns.

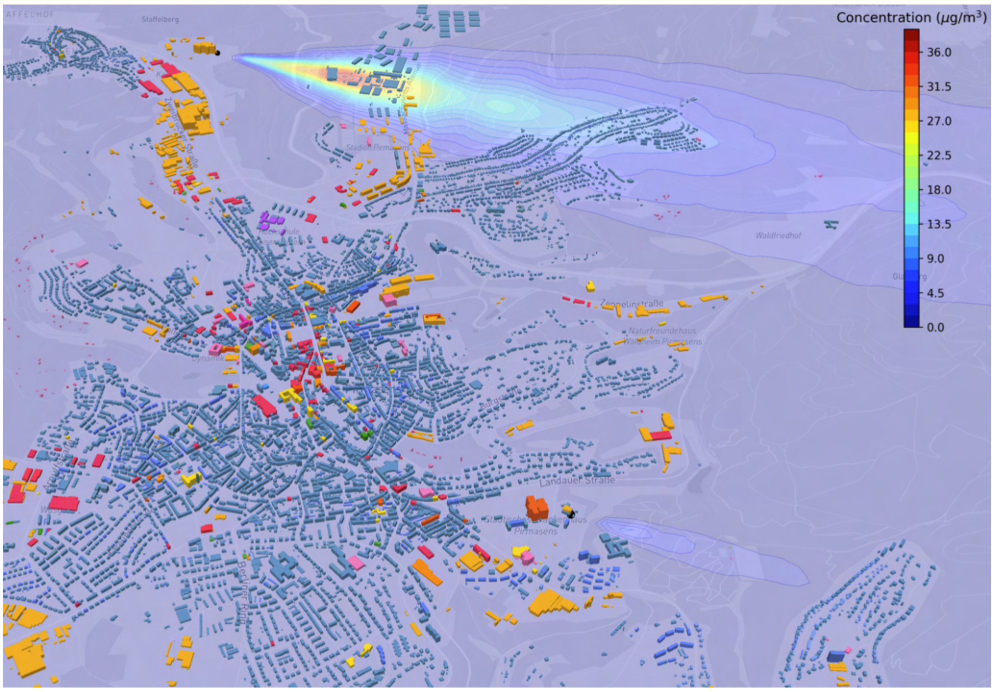
## 4.3 Impact on air quality

Beyond insights into the energetic behaviour of buildings and their optimised heat provision, a key strength of the World Avatar lies in generating cross-domain insights: Emission dispersion simulations are triggered automatically by each heat generation optimisation to immediately understand potential impacts of the projected heat sourcing strategy, comprising multiple locations, on the exposure of various parts of the surrounding population to associated airborne emissions. While this proof-of-concept predominantly focuses on connecting the dynamic cost optimisation with geospatial emission implications, a detailed investigation of potential health consequences remains unexplored. Nevertheless, this avenue holds promise for future research, potentially incorporating adverse health impacts within a multi-objective optimisation framework.

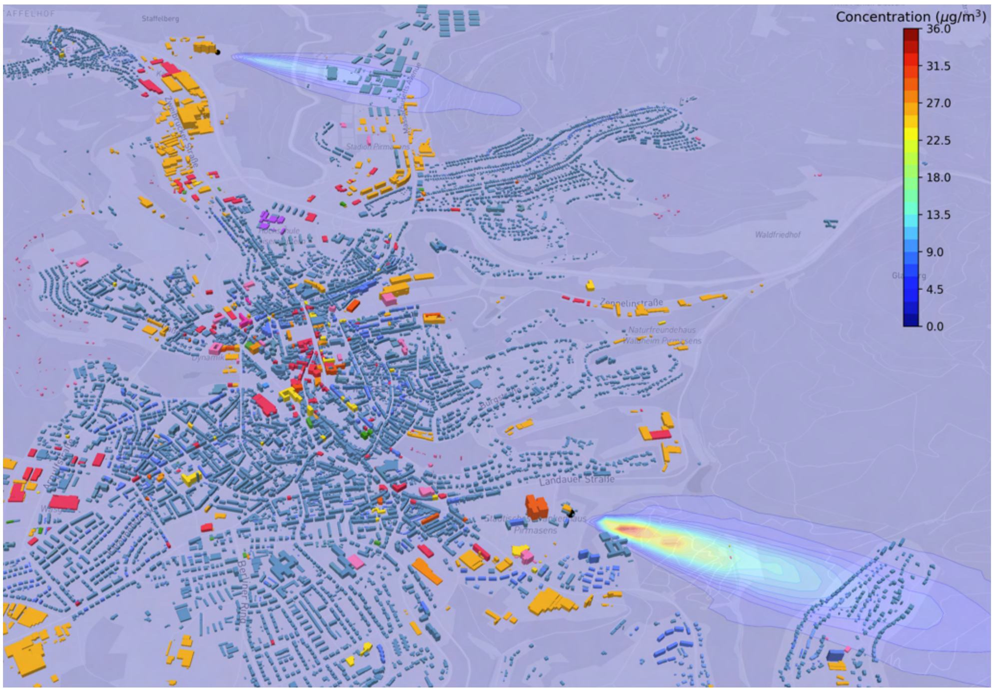
To mirror actual operating conditions, the dynamic optimisation is deployed with an hourly resolution. Hence, also the emission dispersion is simulated for each optimised hour, producing one instantiated dispersion raster per air pollutant and elevation of interest. As generic Gaussian plume model, AERMOD supports various emission types; however, this work focuses on $NO_x$, $PM_{2.5}$, and $PM_{10}$ as major pollutants (see Appendix A.4.2 for details), with $NO_x$ typically exhibiting the highest proportional concentrations. While the dispersion at arbitrary elevations relative to the underlying terrain can be studied, our focus centres on ground level (*i.e.*, 0 m of elevation), given its importance for pedestrians and the general public. A summary of relevant parameters for the AERMOD simulations is provided in Appendix A.2.2.

Instantiated dispersion maps can be overlaid with buildings or population density data to inspect various aspects and potential implications of heat generation induced emissions. In **Fig**. 10, this capability is showcased, displaying $NO_x$ emission values in conjunction with instantiated building stock, where the colours indicate the usage of properties, with blue representing predominantly residential buildings. The figure illustrates a historical scenario across two consecutive hours, during which the start-up of the CHP gas turbine has been deemed profitable by the optimisation routine.

Figure 10(a) illustrates a heat provision situation where the majority of heat is sourced from the waste incineration plant situated in the North of the town. Figure 10(b) depicts the heat generation one hour later, including the active gas turbine located at the munic-




(a) Heat generation related $NO_x$ emission dispersion as of 09 Dec 2020 06:00 UTC.




(b) Heat generation related $NO_x$ emission dispersion as of 09 Dec 2020 07:00 UTC.

**Figure 10:** *Integrated emission dispersion simulation. The integrated simulation of heat generation induced air pollutants provides insights into air pollution implications of various heat generation/sourcing strategies, considering actual (historical) weather data.*

ipal heating plant in the Southern part of the town. Given the different geo-locations of various heat sources, distinct exposure scenarios emerge based on the chosen heat provision strategy due to the incorporation of wind data. Despite similar wind conditions and comparable maximum concentrations, both situations exhibit significantly different exposure potentials. In the first scenario, multiple residential buildings face relatively high $NO_x$ concentrations, whereas these areas are shifted to regions without residential buildings in the second scenario. For additional details on the emission exposure of various population segments, refer to Appendix A.2.1, where emission maps are overlaid with population density data.

Beyond the sole map view, virtual sensors can be placed at arbitrary locations to study air pollution exposure over time. These sensors extract data from underlying raster files and display corresponding values as time series for the respective pollutant types in the visualisation side panel, as illustrated in **Fig**. 11. The depicted emission profiles for different pollutants look similar due to deploying a Gaussian dispersion model. Results could not explicitly be validated against actual local air quality readings due to a lack of available historical data (*i.e.*, sensor readings) within the area of interest. However, given the numerous previous calibration studies of AERMOD, we believe that the derived values possess at least indicative meaning. Moreover, simulated values align well with applicable emission thresholds (see Table 6 in the Appendix) as well as published hourly and daily mean readings reported by the waste incineration plant operator [32]. While the current dispersion model is intentionally simplistic for this proof-of-concept, the workflow

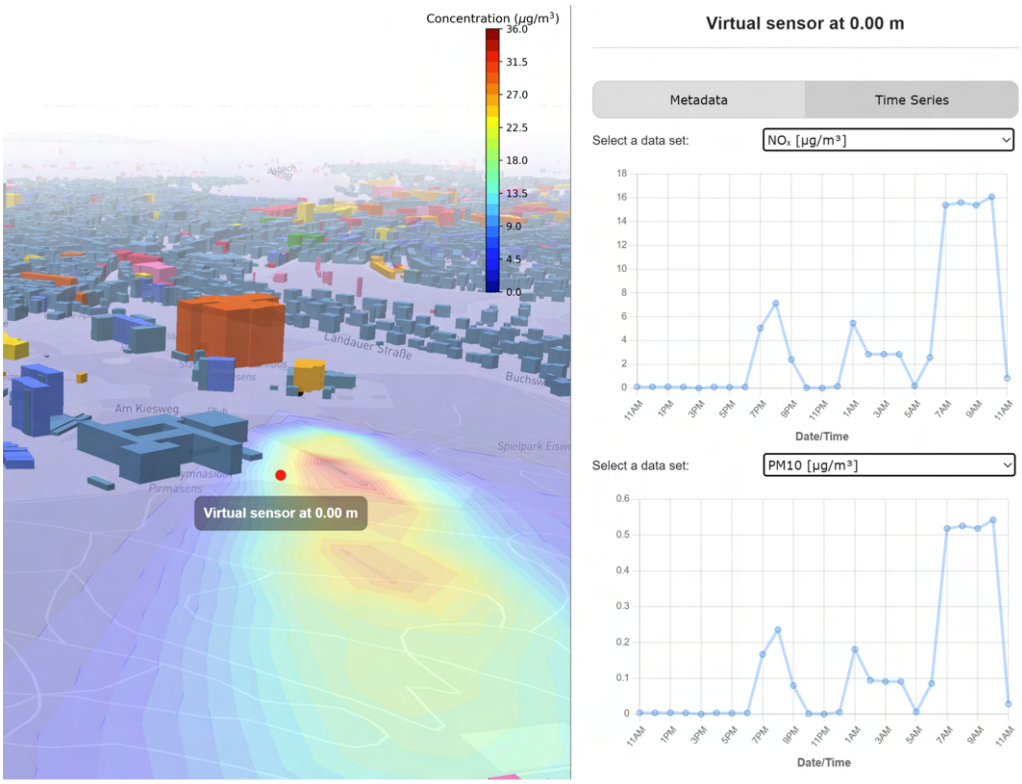



**Figure 11:** *Emission time series. Virtual sensors allow to inspect simulated emission concentrations for arbitrary locations of interest as well as time scales.*

can easily accommodate more sophisticated models in future iterations without significant modifications.

## 4.4 Further directions

Having established the foundation for integrating district heating network operations with an automated assessment of corresponding air pollutant dispersion, a logical extension of this work involves evaluating the potential health implications of these emissions for the community. Expanding on this, a health impact assessment could be integrated as an additional target within the optimisation framework, adopting a multi-objective approach that directly minimises exposure and potential health risks for the affected population.
Moreover, the current dynamic supply-side optimisation utilises historical district heating data, while the building energy analyses based on CEA take a rather generic perspective - also due to the current lack of information about which buildings are connected to the heating grid. Once this data gap is addressed, the use of CEA to gradually replace the reliance on historical data shall be explored. This advancement would allows for a more robust link between demand and supply side in the district heating system.
To further improve the accuracy of the CEA agent, additional building data shall be incorporated. Currently, the agent still relies on CEA's assumptions for parameters like setpoint temperature or the fraction of air-conditioned gross floor area, which significantly influence energy demand calculations. While the agent already uses actual building usage, it lacks specific data on occupancy schedules and electrical appliance usage patterns. With the ongoing integration of detailed building information models into TWA [90], the agent shall be updated accordingly, thereby further improving the accuracy of deployed energy simulation.

# 5 Conclusions

In this work we demonstrate the capabilities of the World Avatar dynamic knowledge graphs to create comprehensive energy perspectives for smart cities by connecting detailed energy analyses for individual buildings with the dynamic control of a municipal district heating system and the simulation of associated emissions - all integrated within the very same system, thereby bridging domains in the nexus of energy, the built environment, and atmospheric emission dispersion.

We have extended the ontological coverage of the World Avatar with knowledge models to describe time series and forecasts, district heating network operations, building energy characteristics, and air pollutant dispersion and leverage these ontologies to instantiate real-world data. We have developed multiple semantic agents to act upon the instantiated data and deploy them in a connected fashion to provide the proof-of-concept for the first model predictive control application within TWA. We have implemented a generic forecasting agent and use it as part of this knowledge graph-native receding horizon optimisation to minimise the total heat generation cost of a district heating network. The outputs of the optimisation are directly linked with an integrated emission dispersion model to understand the impact of various heat generation and sourcing strategies on air pollution.

The showcased degree of interoperability and automation, spanning from energy forecasting to cost-optimal generator dispatch to airborne emission dispersion, is enabled by the automated tracking of dependencies between various agent and data instances within the dynamic knowledge graph. This approach ensures convenient automated information flow between interconnected entities, while upholding an extraordinary level of explainability about how a certain piece of information has been derived.

Furthermore, the City Energy Analyst is made available as part of TWA to provide valuable information about buildings' energy demands and on-site generation potentials with regards to solar energy. The developed agent offers a flexible enhancement to the original CEA toolkit by utilising latest instantiated building and weather data from the knowledge graph to reduce the dependency on default assumptions; thus, promoting a data-driven bottom-up approach for comprehensive energy assessments of building stock at various levels (*e.g.*, building level, district, and city). This perspective complements the dynamic generation optimisation with a more strategic angle, relevant to analyse any potential expansion of the district heating grid to drive low-carbon heating solutions.

Given its extensibility, scalability, and suitability to represent arbitrary data, TWA provides a promising solution to merge and advance individual digital twins collectively, with a semantic layer at its core ensuring that individual agents act upon an aligned data basis. This work shows that semantically chained agents offer great potential to resemble the behaviour of complex systems, address interoperability challenges holistically, and implement automatable cross-domain smart city workflows, not only in the energy space. By integrating real-world sensor data, the accuracy of deployed models, such as AERMOD or the City Energy Analyst, can continuously be refined and, reversely, virtual sensors can easily be deployed to fill gaps in the actual sensor landscape with simulated readings, thereby creating a truly interoperable and dynamic cyber-physical system of connected digital twins.

# 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.

The authors express gratitude to the Stadt Pirmasens, especially mayor Michael Maas and his team, as well as the Stadtwerke Pirmasens, with Christoph Dörr and his team, for their invaluable collaboration and generous support in sharing relevant data, enhancing the depth and quality of this research. This work also leverages data from ©GeoBasis-DE/LVermGeoRP 2023. Furthermore, the authors express gratitude to L.F. Ding and G.H. Xiao for their valuable contributions, particularly in sharing the Ontop mapping and engaging in helpful discussions.

The graphical abstract leverages material designed by macrovector/Freepik.

# Nomenclature

**AERMOD** AMS/EPA regulatory model (air dispersion model)
**API** Application programming interface
**ARIMA** Autoregressive integrated moving average
**CEA** City Energy Analyst
**CHP** Combined heat and power
**DIF** Derived information framework
**dKG** Dynamic knowledge graph
**EEA** European environment agency
**GeoSPARQL** Geographic query language for RDF Data
**IRI** Internationalized resource identifier
**KG** Knowledge graph
**LSTM** Long short-term memory
**MAPE** Mean absolute percentage error
**ME** Maximum error
**$NO_2$** Nitrogen dioxide
**$NO_x$** Nitrogen oxides
**OM** Ontology of units of measure
**OPEX** Operating expense
**OSM** OpenStreetMap
**$PM_{10}$** Particulate matter less than 10 $\mu$m in diameter
**$PM_{2.5}$** Particulate matter less than 2.5$\mu$m in diameter
**PM** Particulate matter
**PV** Photovoltaics
**RDF** Resource description framework
**RMSE** Root mean square error
**SAREF** Smart Applications REFerence ontology
**SARIMAX** Seasonal autoregressive integrated moving average with exogenous regressors
**SMAPE** Symmetric mean absolute percentage error
**SPARQL** SPARQL protocol and RDF query language
**SQL** Structured query language
**TFT** Temporal fusion transformer
**TWA** The World Avatar (dynamic knowledge graph)
**W3C** World Wide Web Consortium
**WHO** World Health Organization

### Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to enhance the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Data and code availability

All the codes developed are available on The World Avatar GitHub repository:
https://github.com/cambridge-cares/TheWorldAvatar.
Developed ontologies can be found in the ontology subdirectory and instructions to reproduce the use case are detailed in the Pirmasens repository.

# A Appendix

## A.1 Namespaces

deriv: <https://www.theworldavatar.com/kg/ontoderivation/>
disp: <https://www.theworldavatar.com/kg/ontodispersion/>
oh: <https://www.theworldavatar.com/kg/ontoheatnetwork/>
ts: <https://www.theworldavatar.com/kg/ontotimeseries/>
ocp: <http://theworldavatar.com/ontology/ontochemplant/OntoChemPlant.owl#>
OntoCAPE: <http://www.theworldavatar.com/ontology/ontocape/chemical_process_system/CPS_performance/economic_performance.owl#>
OntoPowSys: <http://www.theworldavatar.com/ontology/ontopowsys/PowSysRealization.owl#>
ub: <https://www.theworldavatar.com/kg/ontoubemmp/>
bs: <https://www.theworldavatar.com/kg/ontobuildingstructure/>
contract: <https://spec.edmcouncil.org/fibo/ontology/FND/Agreements/Contracts/>
dabgeo: <http://www.purl.org/oema/infrastructure/>
geo: <http://www.opengis.net/ont/geosparql#>
om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>
owl: <http://www.w3.org/2002/07/owl#>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>
time: <http://www.w3.org/2006/time#>
xsd: <http://www.w3.org/2001/XMLSchema#>

## A.2 Results continued

### A.2.1 Emission dispersion

**Figure** 12 depicts the exposure of the town's population to additional air pollution for two different heat sourcing strategies by overlaying the dispersion visualisation over the population density raster: Despite similar weather conditions, certain population segments experience significantly different exposure. The left strategy sources most of the heat from the waste incineration plant located at the outskirts, resulting in relatively higher but more remote emissions (*i.e.*, further away from the town). The right strategy distributes heat sourcing more evenly between the two available sites, resulting in lower overall concentrations; however, the municipal heating plant's plume affects central areas with higher population density. This trade-off shall be addressed in a future version of the optimisation by coupling heat generation with potential health implications for the surrounding population.

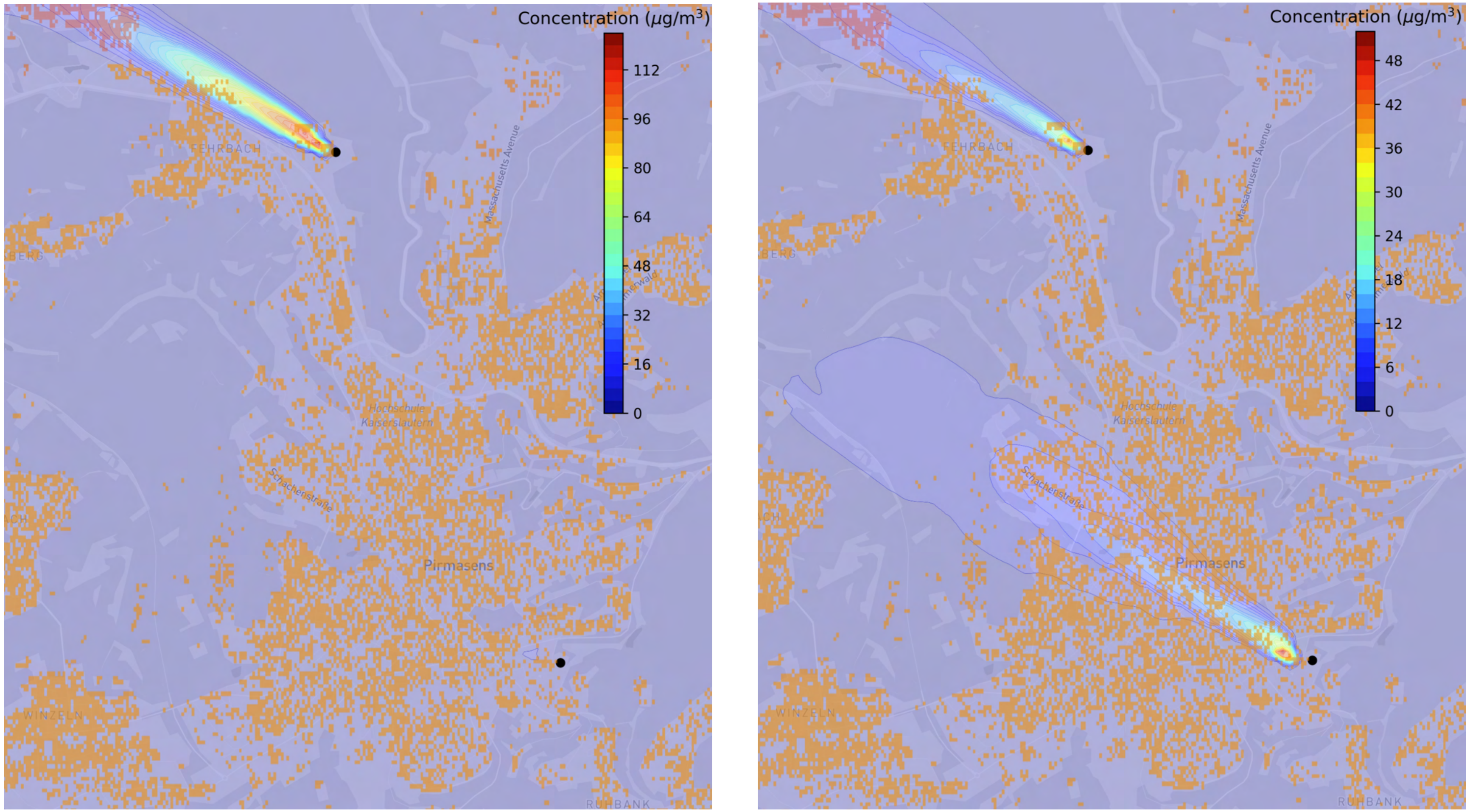



**Figure 12:** *Emission exposure. The integrated dispersion simulation provides insights into the exposure of certain parts of the population to additional air pollutants as a result of heat generation. Illustrated for $NO_x$ based on simulated optimisation results and actual historical weather data as of 10 Dec 2020 04:00 UTC (left) and 17:00 UTC (right), overlayed with the population density raster.*

### A.2.2 AERMOD simulation inputs

Key emission outlet (*i.e.*, stack) parameters used in the AERMOD simulations are summarised in Table 1, with corresponding weather parameters provided in Table 2.

**Table 1:** *Stack parameters used in the AERMOD simulations.*

| Simulation | Figure 10(a) | | Figure 10(b) and 11 | |
|---|---|---|---|---|
| Emission source | Waste incineration plant | Municipal heating plant | Waste incineration plant | Municipal heating plant |
| Height (m) | 50 | 35 | 50 | 35 |
| Diameter (m) | 1.69 | 2.66 | 1.69 | 2.66 |
| Exit velocity (m/s) | 3.14 | 0.02 | 0.42 | 0.15 |
| Temperature (K) | 493.15 | 473.15 | 493.15 | 473.15 |
| $NO_x$ emission rate (g/s) | 11.23 | 0.18 | 1.52 | 1.66 |

**Table 2:** *Historical weather parameters used in AERMOD simulations.*

| Simulation | Figure 10(a) | Figures 10(b) and 11 |
|---|---|---|
| Wind direction (°) | 290 | 290 |
| Wind speed (knot) | 6 | 7 |
| Temperature (°F) | 33 | 33 |
| Humidity (%) | 98 | 98 |
| Cloud cover (%) | 90 | 90 |

### A.2.3 City energy analyses

The CEA agent instantiates various energy outputs beyond heating demand which can also be visualised via TWA, including a building's annual electricity demand (**Fig**. 13) as well as annual generation potential for heat from solar collectors (**Fig**. 14) or electricity from PV installations on the different walls and roof. Furthermore, the gross floor area specific values for electricity and heat demand (**Fig**. 15 and **Fig**. 16) as well as area normalised values for solar potential (**Fig**. 17) can be assessed. The visualisations allow for quick inspection and identification of different building demands and potentials, which can be a good starting point for more detailed analyses or scenario planning.

The TWA visualisation uses OpenStreetMap [83] as base layer, which may result in the perception that the instantiated building data is incomplete. However, this solely stems

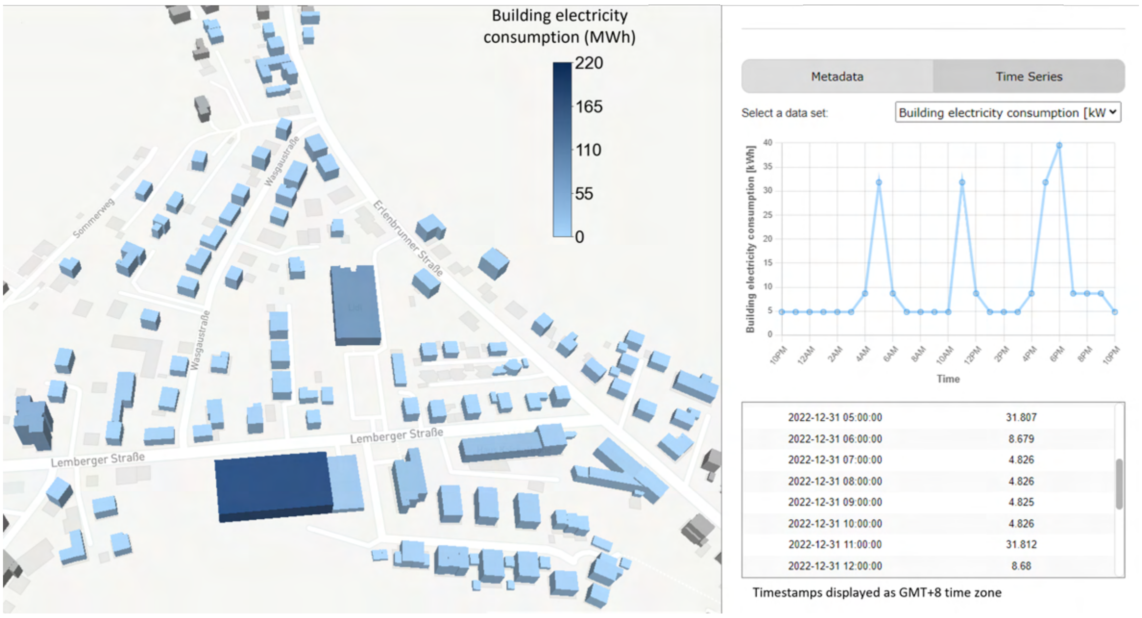



**Figure 13:** *Visualisation of building annual electricity consumption as simulated by the CEA agent.*

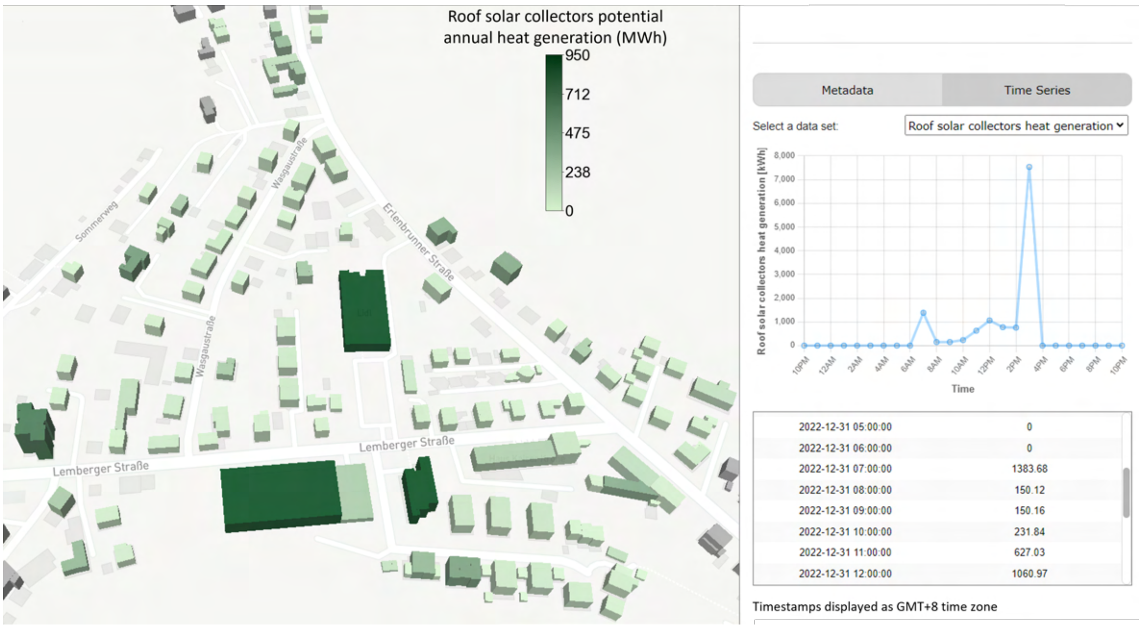



**Figure 14:** *Visualisation of potential heat generation by solar collectors on building roofs as simulated by the CEA agent.*

from the fact that the instantiated buildings for all analyses are derived from an official 3D building stock data set of the state of Rhineland-Palatinate [65], with slight discrepancies compared to OSM.

Beyond assessing a building's annual consumption or generation potential, the CEA agent also evaluates respective load profiles. The outputs can provide an initial assessment of

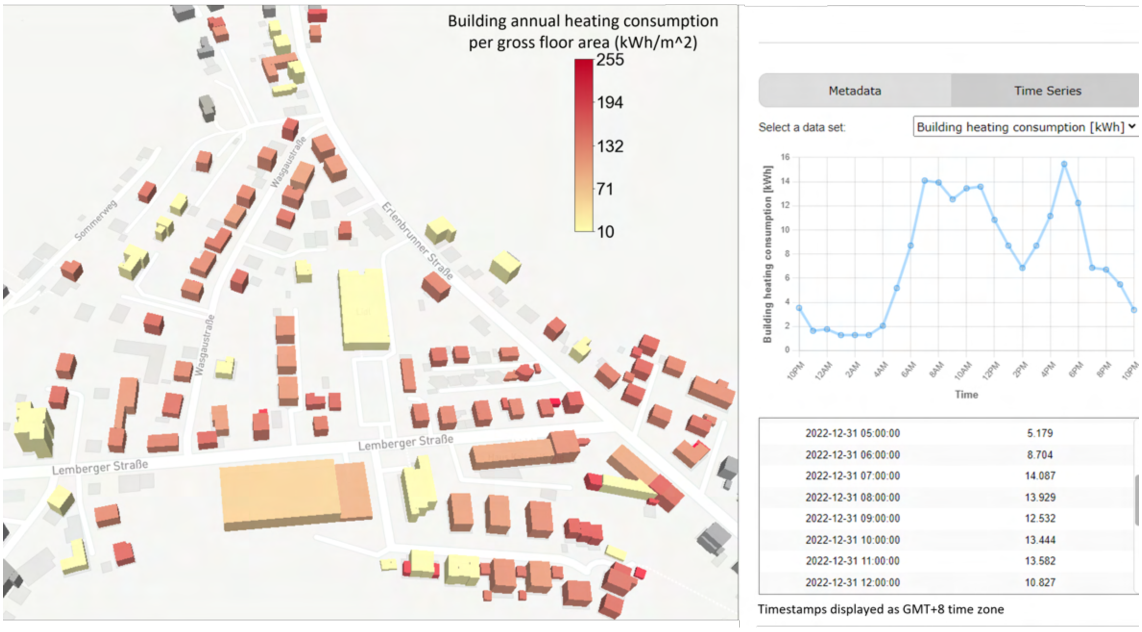



**Figure 15:** *Visualisation of building annual heating consumption per gross floor area as simulated by the CEA agent.*

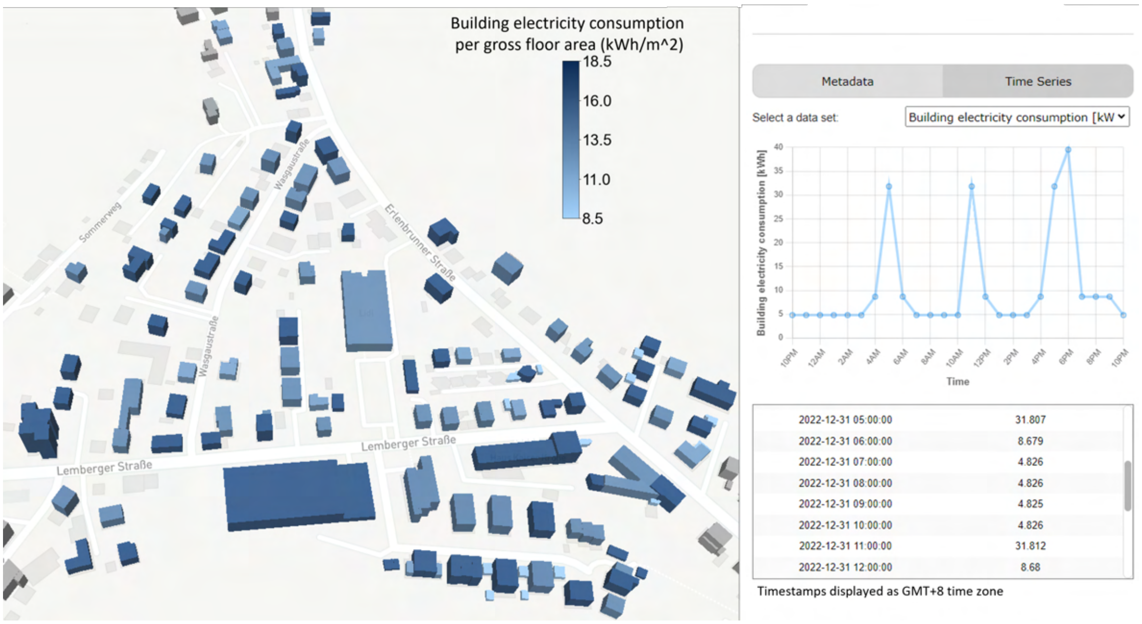



**Figure 16:** *Visualisation of building annual electricity consumption per gross floor area as simulated by the CEA agent.*

directly utilising own generated electricity from rooftop PV to offset demand, without factoring in potential storage solutions, as shown in **Fig**. 18. Negative values indicate an excess of electricity generation, which could be supplied to the grid in active prosumer scenarios.

To benchmark the performance of the CEA agent with regards to electricity demand, sim-

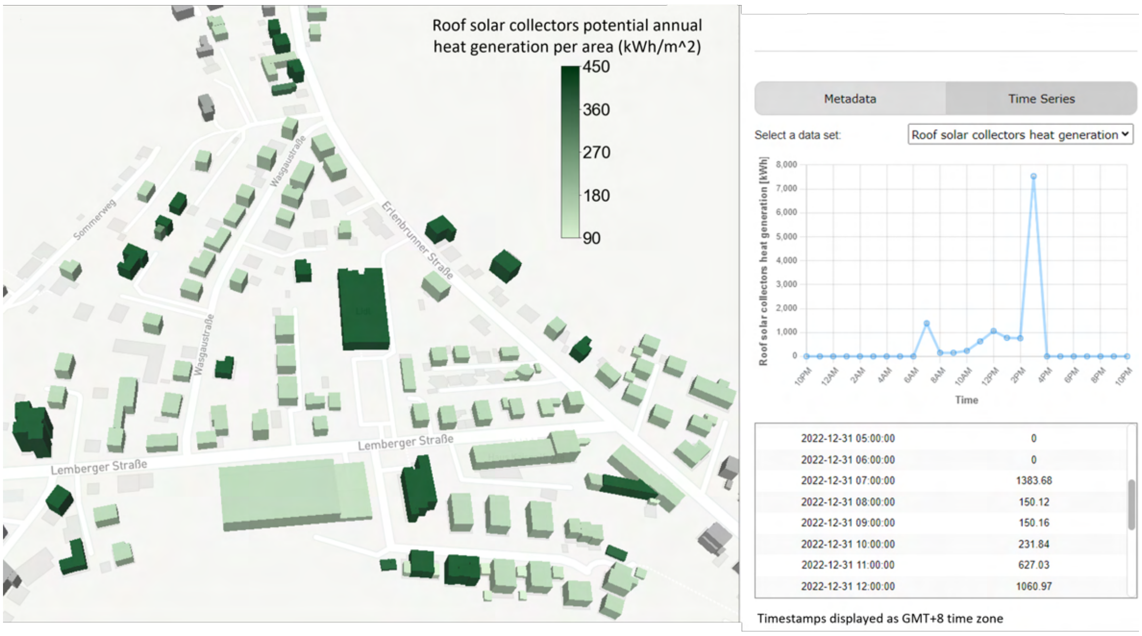



**Figure 17:** *Visualisation of potential heat generation by solar collectors on building roofs per suitable roof area as simulated by the CEA agent.*

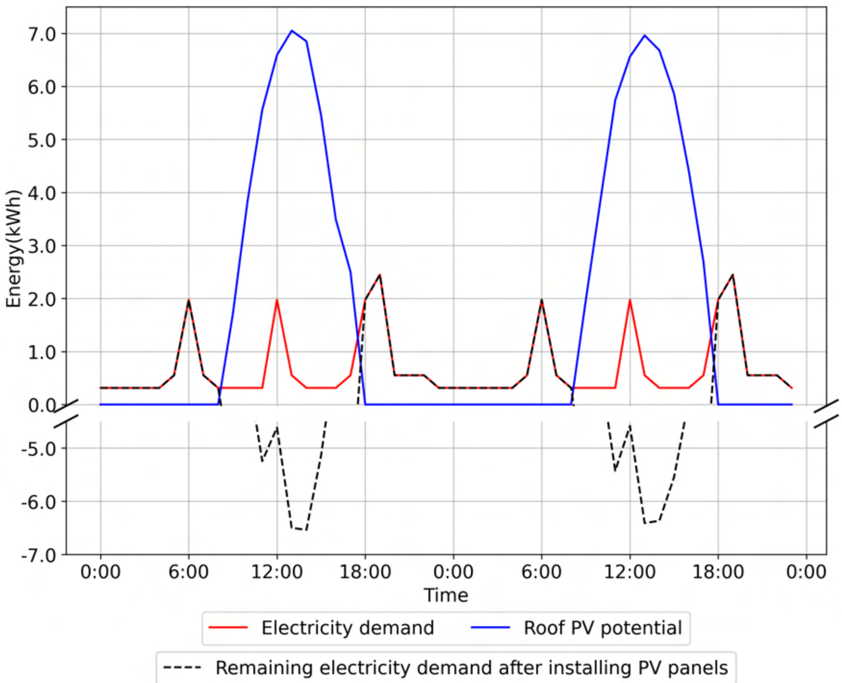



**Figure 18:** *Electricity load profiles simulated by CEA agent. Electricity demand and PV generation time series can be used to evaluate potential energy savings of on-site PV collector installations (depicted for a typical day in March).*

ulated values have been compared against actual historical readings. Annual electricity consumption data for four buildings has been obtained and compared in **Fig**. 19, juxtaposing the CEA agent results with native CEA estimates (*i.e.*, the unaltered CEA software) and actual historical values. While using native CEA results in a MAPE of 57.6% compared to historical consumption, the CEA agent yields a reduced MAPE of 13.7%, by using building specific input data as well as actual weather and terrain inputs representative for the actual location. This improvement outlines the value of removing assumptions in the underlying CEA toolkit by replacing them with actual data from the dKG; however, the small sample size due to limited availability of historical electricity readings needs to be noted. The improved accuracy can mainly be attributed to the use of actual location-specific weather data, including dry bulb temperature, dew point temperature, cloud cover, and direct normal as well as diffuse horizontal irradiance.

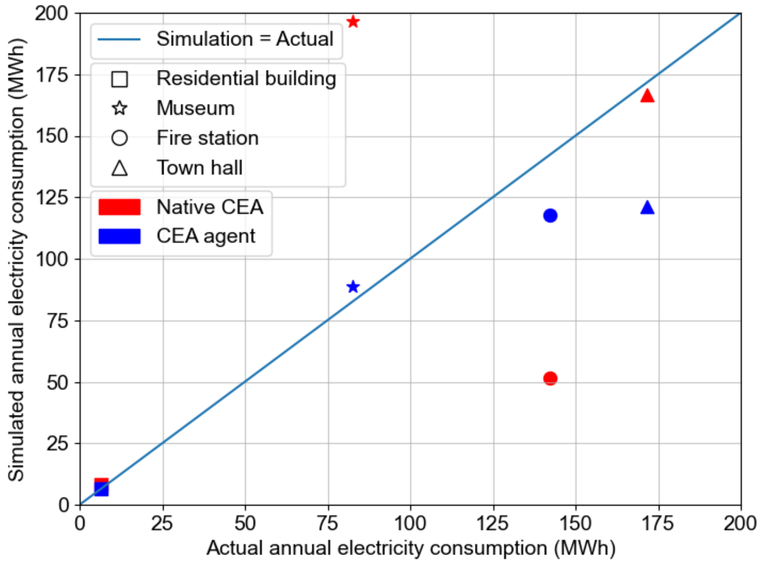



**Figure 19:** *Comparison of annual electricity consumption estimates. Incorporating actual building stock data, the CEA agent aligns more closely with observed historical consumption figures compared to the native CEA toolkit.*

To scrutinise the reliability of CEA agent's solar potential assessment, rooftop PV electricity estimates have been compared against figures published by the Ministry for Climate Protection, Environment, Energy, and Mobility in Rhineland-Palatinate (*i.e.*, Solarkataster [73]) for 25 buildings, as depicted in **Fig**. 20. Comparing generation potential per area (*i.e.*, to mitigate discrepancies stemming from differences in area estimations), the native CEA toolkit yields a 28.1% MAPE, whereas the agent reduces this error to 12.9%. Detailed results can be found in **Table** 3. As shown in Fig. 20, the native CEA consistently underestimates PV potential compared to Solarkataster. The agent, while showing a slight overestimation, demonstrates a narrower distribution closer to a zero relative error. The significantly larger discrepancies in the native CEA are likely a result of biased or incorrect built-in assumptions.

Despite significant improvement, certain discrepancies remain between the agent results and the used benchmarks. This could be attributed to potential differences in comparison

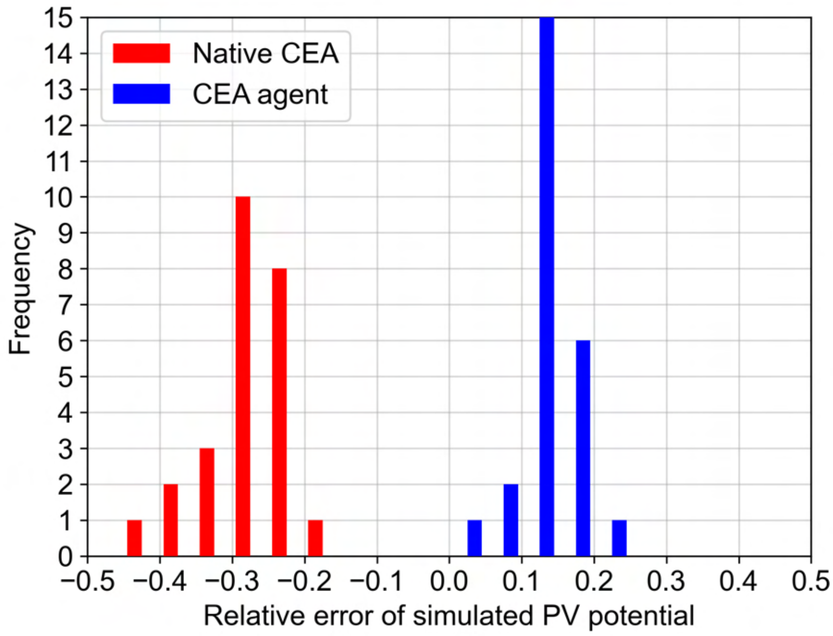



**Figure 20:** *Comparison of simulated roof solar potentials per area. The CEA agent's photovoltaic estimations exhibit less bias and higher precision than native CEA when compared to the official Solarkataster [73] values (analysed for 25 buildings).*

bases due to a lack of detailed context information about the used benchmarks. Historical electricity consumption figures, for instance, may represent aggregated data for complex building structures (*e.g.*, a school with an integrated laboratory and gym), while the agent currently assumes a single overarching usage classification per building. This hypothesis is supported by the residential building, being a single structure with one usage type, exhibiting best alignment with historical data among all studied buildings (see Fig. 19). Discrepancies in the solar potential estimation likely stem from unaligned assumptions in the underlying simulations, including the use of slightly different roof geometries and angles, as well as variations in PV panel materials and efficiency. The consistent overestimation supports this hypothesis, though it cannot be verified as the Solarkataster methodology is not published.

It has been highlighted that the use of actual weather data significantly contributes to the agent's improvement over native CEA. Three crucial weather parameters for PV potential calculations are direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and total sky cover. Hence, the difference between the actual historical values for the location of interest has been compared against the CEA assumptions for these parameters. As CEA simulations consider hourly weather data, the comparison has been conducted for weather time series data in hourly format over a year. Comparing the weather data at the actual building location used by CEA agent against the CEA assumptions, it has been found that the DNI increased by $409.3\,\mathrm{kWh/m^2}$ and the DHI decreased by $118.8\,\mathrm{kWh/m^2}$ when DNI and DHI are summed over the year. The total sky cover has been averaged over the year and the average hourly value of the weather data used by the agent is 0.74 less than

**Table 3:** *Comparison of simulated roof PV potential per area against official Solarkataster [73] estimates for 25 buildings.*

| | Roof PV potential per area (in $\mathrm{kW\,h\,m^{-2}}$) | | | Relative error compared to Solarkataster estimate | |
|---|---|---|---|---|---|
| # | Solarkataster | Native CEA | CEA agent | Native CEA | CEA agent |
| 1 | 155.7 | 121.0 | 185.7 | −0.22 | 0.19 |
| 2 | 156.8 | 125.8 | 190.9 | −0.20 | 0.22 |
| 3 | 161.2 | 101.5 | 191.0 | −0.37 | 0.18 |
| 4 | 161.5 | 125.0 | 187.7 | −0.23 | 0.16 |
| 5 | 164.4 | 122.3 | 190.9 | −0.26 | 0.16 |
| 6 | 164.6 | 119.1 | 181.4 | −0.28 | 0.10 |
| 7 | 166.1 | 126.3 | 189.9 | −0.24 | 0.14 |
| 8 | 166.2 | 122.3 | 190.2 | −0.26 | 0.14 |
| 9 | 166.2 | 125.0 | 191.0 | −0.25 | 0.15 |
| 10 | 166.3 | 120.2 | 190.9 | −0.28 | 0.15 |
| 11 | 166.9 | 92.6 | 190.1 | −0.45 | 0.14 |
| 12 | 167.0 | 117.1 | 189.4 | −0.30 | 0.13 |
| 13 | 167.2 | 126.1 | 186.2 | −0.25 | 0.11 |
| 14 | 167.8 | 126.6 | 190.9 | −0.25 | 0.14 |
| 15 | 167.8 | 125.2 | 191.0 | −0.25 | 0.14 |
| 16 | 168.4 | 125.0 | 168.5 | −0.26 | 0.00 |
| 17 | 168.9 | 112.5 | 190.4 | −0.33 | 0.13 |
| 18 | 169.0 | 125.3 | 190.9 | −0.26 | 0.13 |
| 19 | 169.0 | 126.4 | 190.9 | −0.25 | 0.13 |
| 20 | 169.3 | 126.1 | 189.4 | −0.26 | 0.12 |
| 21 | 169.8 | 125.9 | 191.1 | −0.26 | 0.13 |
| 22 | 171.1 | 104.0 | 191.0 | −0.39 | 0.12 |
| 23 | 171.8 | 115.5 | 191.0 | −0.33 | 0.11 |
| 24 | 179.9 | 126.5 | 189.7 | −0.30 | 0.05 |
| 25 | 181.5 | 124.6 | 191.0 | −0.31 | 0.05 |
| | | | Average | 0.281 | 0.129 |

the CEA assumption. This disparity in weather parameters indicates that native CEA calculations assume lower solar irradiance and higher sky cover than the Pirmasens weather, leading to an underestimation of solar radiation received by PV panels and, consequently, a decrease in calculated PV potential. The CEA agent rectifies these inaccuracies by utilising weather data representative for the actual building location.

## A.3 District heating network ontology

**Figure** 22 provides a more detailed illustration of the proposed conceptualisation for (municipal) district heating networks, encompassing the overall network structure down to generator-specific operational data. **Figure** 21 outlines the proposed hierarchical structure for organising key operational and cost concepts.

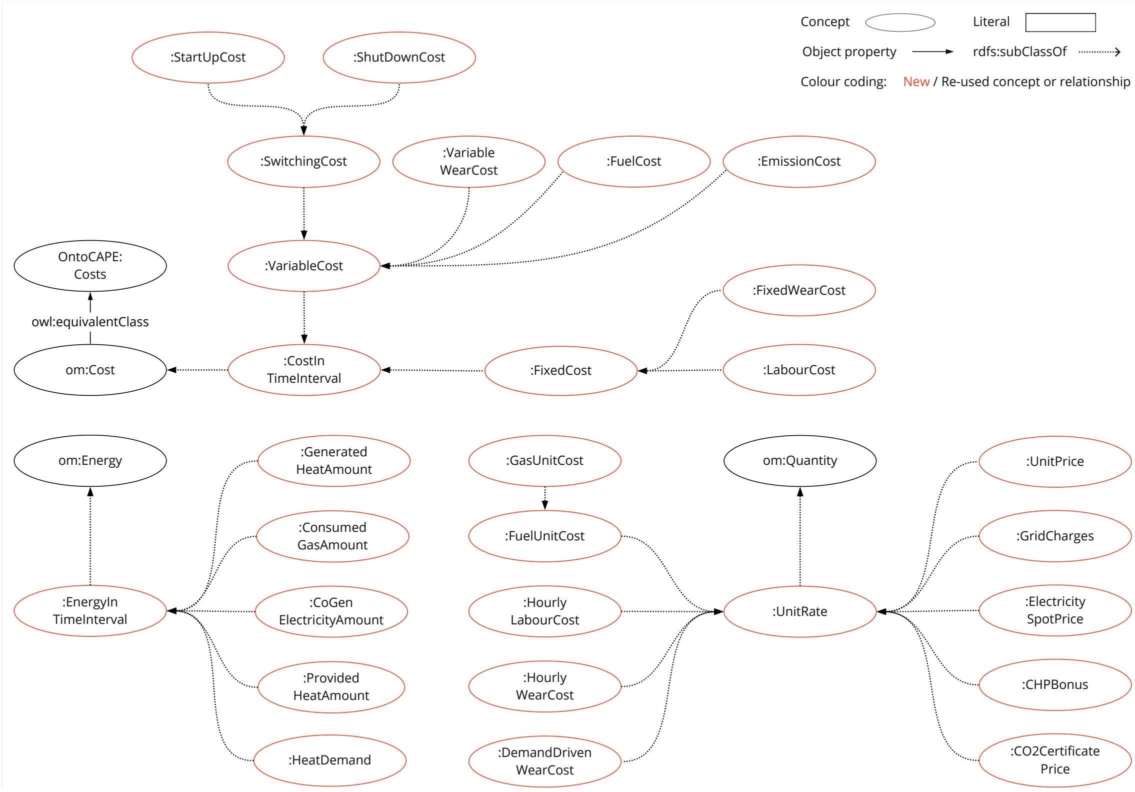



**Figure 21:** *Hierarchical structure of relevant cost components and heat generation outputs as defined by the OntoHeatNetwork ontology. All concepts are instantiated using the* `om` *or* `ts` *ontology, as single values or time series, respectively. All referenced namespaces are declared in Appendix A.1, with not explicitly stated prefixes referring to OntoHeatNetwork.*

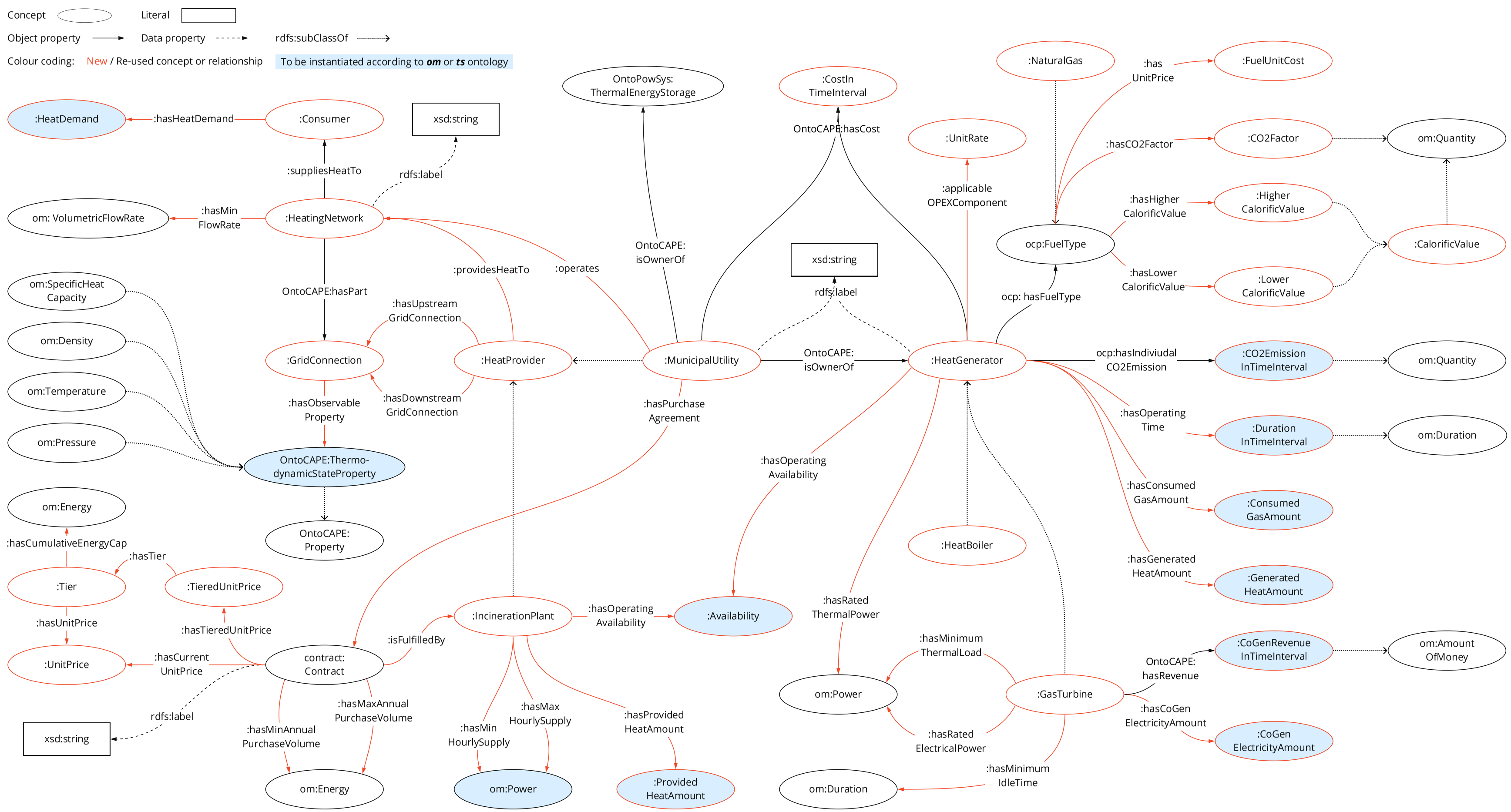



**Figure 22:** *OntoHeatNetwork. The OntoHeatNetwork domain ontology captures key concepts and relationships of a municipal district heating network, and links to* `om` *and* `ts` *to instantiate and operationalise actual numerical data. All referenced namespaces are declared in Appendix A.1, with not explicitly stated prefixes referring to OntoHeatNetwork.*

## A.4 Agent details

### A.4.1 Forecasting agent

**Forecasting logic** Upon invocation, the agent first verifies the suitability of received inputs to derive a forecast, *i.e.*, checks that exactly one IRI to forecast, one `ts:ForecastingModel`, one `ts:Frequency`, one `time:Interval`, and one `time:Duration` concept are marked up as derivation inputs. The instance to be predicted must pertain to a type that is either a direct or nested subclass of `om:Quantity` or `owl:Thing`. Although most instances to forecast will likely be of (sub-)class `om:Quantity`, the support for the more general `owl:Thing` ensures that anything with an associated time series can be predicted. After successful verification, the agent queries further relevant inputs from the KG and creates an overarching configuration dictionary describing the forecast to create. If the instantiated `rdfs:label` of the target forecasting model is "prophet" (irrespective of capitalisation), the default Prophet model [104] will be used. Otherwise, the agent will try to load the corresponding custom pre-trained model as specified in a model mapping file.

Facebook Open Source has published Prophet [104] as a modular regression model designed for univariate automatic forecasting. Accepting a univariate time series as input, Prophet autonomously identifies data frequency and conducts seasonality analysis. Moreover, it can effectively model special events, holidays, and series with multiple seasonalities, overcoming common challenges like stationarity requirements or manual parameterisation as posed by the ARIMA family. In contrast to many machine learning-based forecasting methods, Prophet is less complex, easier to implement and computationally less intensive; hence, perfectly suited as versatile default forecasting model.

To use pre-trained custom models, both a model and covariate loading function need to be specified in a model mapping file. The mapping file is added as bind mount into the agent Docker container to allow for seamless addition of custom models even after agent deployment. Custom loading functions are required to 1) ensure consistent handling of time series data between training and forecasting (*i.e.*, identification of covariate order, data scaling, pre-processing, *etc.*) and 2) capture model specific requirements of different forecasting techniques (*e.g.*, custom handling of GPU trained models on a CPU machine). Furthermore, custom covariate loading functions allow for the incorporation of additional time-dependent covariates, such as representing (long-term) seasonalities through (multiple) Fourier series.

After creating the forecast configuration, the agent loads the time series data, including covariates if specified. Potentially required resampling of the retrieved time series (*i.e.*, as most forecasting models require equally spaced time series data) does not alter the instantiated data, but creates an internal copy for forecasting. Subsequently, the pre-trained model is loaded or a new Prophet model is created and fitted to the retrieved time series to predict the data. Afterwards, the forecast is created for the specified `time:Interval` (with both bounds inclusive). If the forecasting model requires scaled data, both the time series and covariates (if applicable) are scaled based on their historical values during `time:Duration` prior to the forecast start date. Lastly, the predicted time series is transformed back to its original scale and instantiated/updated in the dKG. For new forecasts, a new `ts:Forecast` instance is created and attached to the `om:Quantity` or `owl:Thing`

IRI which has been forecasted. Further metadata, *e.g.*, which data and model have been used, are instantiated according to the OntoTimeSeries ontology. For existing forecasts, only the time series value in the relational database are replaced and respective meta data gets updated.

**Forecasting performance** Temporal fusion transformers have been fitted to the historical hourly spaced data and instantiated in TWA to forecast heat demand and grid temperatures for the target use case. TFTs have initially been developed to overcome typical forecasting problems, such as considering the interaction between time-invariant covariates, known future inputs (*e.g.*, tomorrow is Friday), and (un)known exogenous future time series (*e.g.*, tomorrow's outside temperature). TFTs have shown significant performance improvements and interpretability gains compared to complete black box neural networks [68].

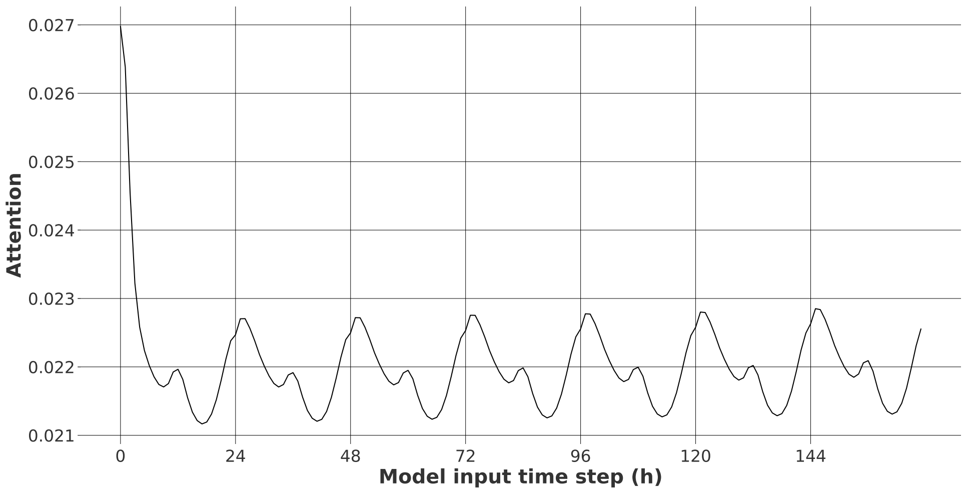



**Figure 23:** *TFT attention weights for heat demand. Attention values for a multi-step forecast with 24 steps and 168 given past values, averaged over 365 forecast days. The plot shows a clear increased attention to the recent history, as well as related times in the more distant past (i.e., 24 steps seasonal cycle).*

Both district heating demand and grid flow and return temperatures show typical daily and weekly patterns, similar to many other energy consumption time series. The attention weights of the TFT model used for heat demand forecasting is shown in **Fig.** 23, highlighting the relative importance of the recent history as well as the expected daily pattern. The performance of the fine-tuned TFT models for both heat demand and grid flow temperature has been compared to a previously suggested fine-tuned SARIMAX model [50]. Considered error metrics are the root mean squared error (RMSE) (in MWh), the symmetric mean absolute percentage error (SMAPE) (in %) and the maximum error (ME) (in MWh) for 1 day, 1 week and 1 month ahead forecasts. The results are provided in **Table** 4 and **Table** 5, respectively, outlining the improved accuracy of the fine-tuned TFT model

**Table 4:** *Forecast error comparison for municipal heat demand. The stated SARIMAX model refers to the fine-tuned model from a previous study [50] and the prefix "_X" refers to forecasting models with covariates.*

| | 24 (1 day) | | | 168 (1 week) | | | 720 (1 month) | | |
|---|---|---|---|---|---|---|---|---|---|
| | RMSE | SMAPE | ME | RMSE | SMAPE | ME | RMSE | SMAPE | ME |
| Prophet | 1.53 | 30.61 | 2.58 | 1.43 | 28.99 | 3.13 | 1.55 | 32.47 | 4.54 |
| SARIMA | 0.90 | 14.65 | 1.79 | 1.47 | 23.63 | 3.43 | 2.16 | 31.23 | 5.41 |
| LSTM | 1.12 | 18.72 | 2.24 | 1.68 | 36.27 | 4.23 | 2.87 | 66.80 | 6.90 |
| TFT | 1.25 | 19.89 | 2.69 | 1.76 | 25.27 | 4.49 | 2.40 | 36.91 | 6.31 |
| SARIMAX | 0.74 | 15.62 | 1.38 | 0.94 | 17.94 | 2.58 | 1.29 | 26.52 | 3.97 |
| LSTM_X | 0.84 | 13.75 | 1.77 | 0.99 | 20.04 | 2.57 | 1.27 | 23.02 | 3.91 |
| TFT_X | 0.73 | 12.28 | 1.43 | 0.76 | 12.41 | 2.12 | 0.89 | 14.76 | 2.91 |
| TFT_X_tuned | **0.53** | **9.48** | **1.06** | **0.61** | **10.43** | **1.80** | **0.72** | **11.81** | **2.27** |

**Table 5:** *Forecast error comparison for grid flow temperature. The stated SARIMAX model refers to the fine-tuned model from a previous study [50] and the prefix "_X" refers to forecasting models with covariates.*

| | 24 (1 day) | | | 168 (1 week) | | | 720 (1 month) | | |
|---|---|---|---|---|---|---|---|---|---|
| | RMSE | SMAPE | ME | RMSE | SMAPE | ME | RMSE | SMAPE | ME |
| Prophet | 2.87 | 2.61 | 5.26 | 3.10 | 2.68 | 8.39 | 3.50 | 3.01 | 10.70 |
| SARIMA | 2.70 | 2.30 | 5.42 | 4.30 | 3.76 | 10.28 | 5.84 | 5.44 | 14.47 |
| LSTM | 2.97 | 2.62 | 5.59 | 5.39 | 4.71 | 11.86 | 7.57 | 7.15 | 17.11 |
| TFT | 2.56 | 2.15 | 5.12 | 3.95 | 3.35 | 10.14 | 4.85 | 4.11 | 13.81 |
| SARIMAX | 2.00 | 1.71 | 4.09 | 3.03 | 2.59 | 7.82 | 3.97 | 3.57 | 10.57 |
| LSTM_X | 2.28 | 1.94 | 4.63 | 2.63 | 2.15 | 7.70 | 2.69 | 2.22 | 8.89 |
| TFT_X | **0.90** | **0.79** | **2.02** | **1.09** | **0.83** | **4.22** | **1.20** | **0.93** | **6.20** |

over the previously deployed SARIMAX one, especially for longer forecast horizons.

### A.4.2 Emission estimation agent

Both the European Environment Agency (EEA) [39] and the World Health Organization (WHO) [111] specify threshold concentrations for various airborne pollutants. Exposure to air containing pollutant concentrations above these values may increase the risk of various respiratory diseases [111]. The annual, daily, and hourly average threshold concentrations for some pollutants are summarised in **Table** 6. For other pollutants, different averaging periods are used to formulate air quality standards.

Emissions rates of various pollutants such as $SO_x$, $NO_x$ and $PM_{2.5}$ were decreasing in Europe between 2000 and 2017 due to the implementation of pollution control measures [102]. Nevertheless, the concentrations of $PM_{2.5}$, $PM_{10}$, $O_3$, $NO_2$ and Bezo($\alpha$)pyrene remain unacceptably high.

**Table 6:** *Air pollutant concentration thresholds in* $\mu g\, m^{-3}$ *for different reference (i.e., averaging) periods stipulated by EEA [39] and WHO [111].*

| | **Annual limits** | | **Daily limits** | | **Hourly limits** | |
|---|---|---|---|---|---|---|
| Pollutant | EEA | WHO | EEA | WHO | EEA | WHO |
| $NO_2$ | 40 | 10 | | | 200[1] | 200 |
| $PM_{10}$ | 40 | 15 | 50[2] | 45 | | |
| $PM_{2.5}$ | 25 | 5 | | | | |

The objective of the *emission estimation agent* is to determine the extent to which heat sourced from the waste incineration plant as well as heat generated by natural gas burning in the district heating plant in Pirmasens affect the concentrations of major airborne pollutants in the surroundings. EEA data for 2022 for the nearby city of Kaiserslautern [40] indicates that the concentrations of $PM_{2.5}$, $PM_{10}$ as well as $NO_2$ were above the WHO limits given in Table 6 with the concentrations of all other pollutants meeting the standards. Hence, this work focuses on evaluating concentrations of the three pollutants. The following paragraphs summarise the literature review on relevant emission factors, while the medians stated in **Table** 7 and **Table** 8 are implemented by the *emission estimation agent*.

**Assumptions and limitations** Based on limited data availability and ambiguous referencing in the literature, emission values for $NO_2$ and $NO_x$ have been used interchangeably throughout this work. Furthermore, particulate matter emissions are not consistently reported as separate values for $PM_{2.5}$ and $PM_{10}$ in the literature: if separate values for $PM_{2.5}$ and $PM_{10}$ are provided, those have been used; if only $PM_{2.5}$ values are reported, corresponding $PM_{10}$ values are assumed equal, since $PM_{2.5}$ is a subset of $PM_{10}$; if only $PM_{10}$ values are reported (*e.g.*, reported collectively as "dust" or "fine dust"), corresponding $PM_{2.5}$ values have been derived using $PM_{2.5}$ to $PM_{10}$ ratios as reported by Krause and Smith [64].

**Waste incineration plant** To estimate the emissions associated with heat sourcing from the waste incineration plant, published operations data from the plant operator has been used where available (*e.g.*, capacity, electricity generation, heat generation, flue gas temperature) [31]. Additionally, the literature has been screened for representative emission factors of waste incineration plants in central Europe or even Germany [36, 58, 60, 77, 78]. Data from the plant operator states a heat generation capacity of 37,000 MWh and a electricty generation capacity of 67,000 MWh p.a, at a total annual waste capacity of 180,000 ton. Correspondingly, specific emissions for $PM_{2.5}$, $PM_{10}$ and $NO_2$ have been calculated per MWh of produced heat, and are summarised in Table 7. It should be noted that the figures for $NO_2$ assume the plant to employ at least basic Selective Catalytic Reduction

[1]Not to be exceeded on more than 18 hours per year

[2]Not to be exceeded on more than 35 days per year

technology, which is in line with operator data.

**District heating plant** In order to estimate emissions associated with the operation of the gas boilers and CHP gas turbine at the municipal heating plant, emission factors for natural gas combustion have been collected from a range of sources, including scientific publications and government agency reports [5, 26, 56, 61, 70, 79, 103, 107]. To ensure consistency and accuracy, parameters related to natural gas, such as density and energy content, have been harmonized with the actual fuel composition used at the plant, as determined through a gas analysis conducted and shared by the plant operator. Consequently, specific emissions for $PM_{2.5}$, $PM_{10}$, and $NO_2$ have been calculated per MWh of natural gas burned (*i.e.*, with reference to the lower calorific value), and are summarised in Table 8.

**Table 7:** *Emission factors for pollutants from energy from waste plants (in* kg *of pollutant per* MWh *of produced heat).*

| Pollutant | Emission factors (kg/MWh) | | | | | |
|---|---|---|---|---|---|---|
| | IPCC [60] | EEA [78] | Energy Justice Network [36] | German Env. Agency [58] | European IPPC Bureau [77] | **Median** |
| $NO_2$ | 4.014 | 5.210 | 4.025 | 4.014 | 3.478 | **4.014** |
| $PM_{10}$ | | 0.015 | 0.131 | 0.134 | 0.161 | **0.133** |
| $PM_{2.5}$ | | 0.015 | 0.131 | 0.108 | 0.130 | **0.119** |

**Table 8:** *Emission factors for pollutants from district heating plant, i.e., gas boilers and gas turbines (in* kg *of pollutant per* MWh *of natural gas burned (referenced to lower calorific value)).*

| Pollutant | Emission factors (kg/MWh) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | IPCC [56] | EEA [79] | Medeiros et al. [70] | Sosa et al. [103] | Villela et al. [26] | Aste et al. [5] | German Env. Agency [61] | EPA [107] | **Median** |
| $NO_2$ | 0.9000 | 0.3204 | 0.2434 | 0.2963 | 0.0640 | 0.2052 | 0.1219 | 0.1559 | **0.2243** |
| $PM_{10}$ | | 0.0032 | | 0.0119 | 0.0227 | 0.0018 | 0.0004 | 0.0119 | **0.0075** |
| $PM_{2.5}$ | | 0.0032 | | 0.0119 | 0.0227 | 0.0018 | 0.0004 | 0.0089 | **0.0060** |

### A.4.3 City Energy Analyst agent

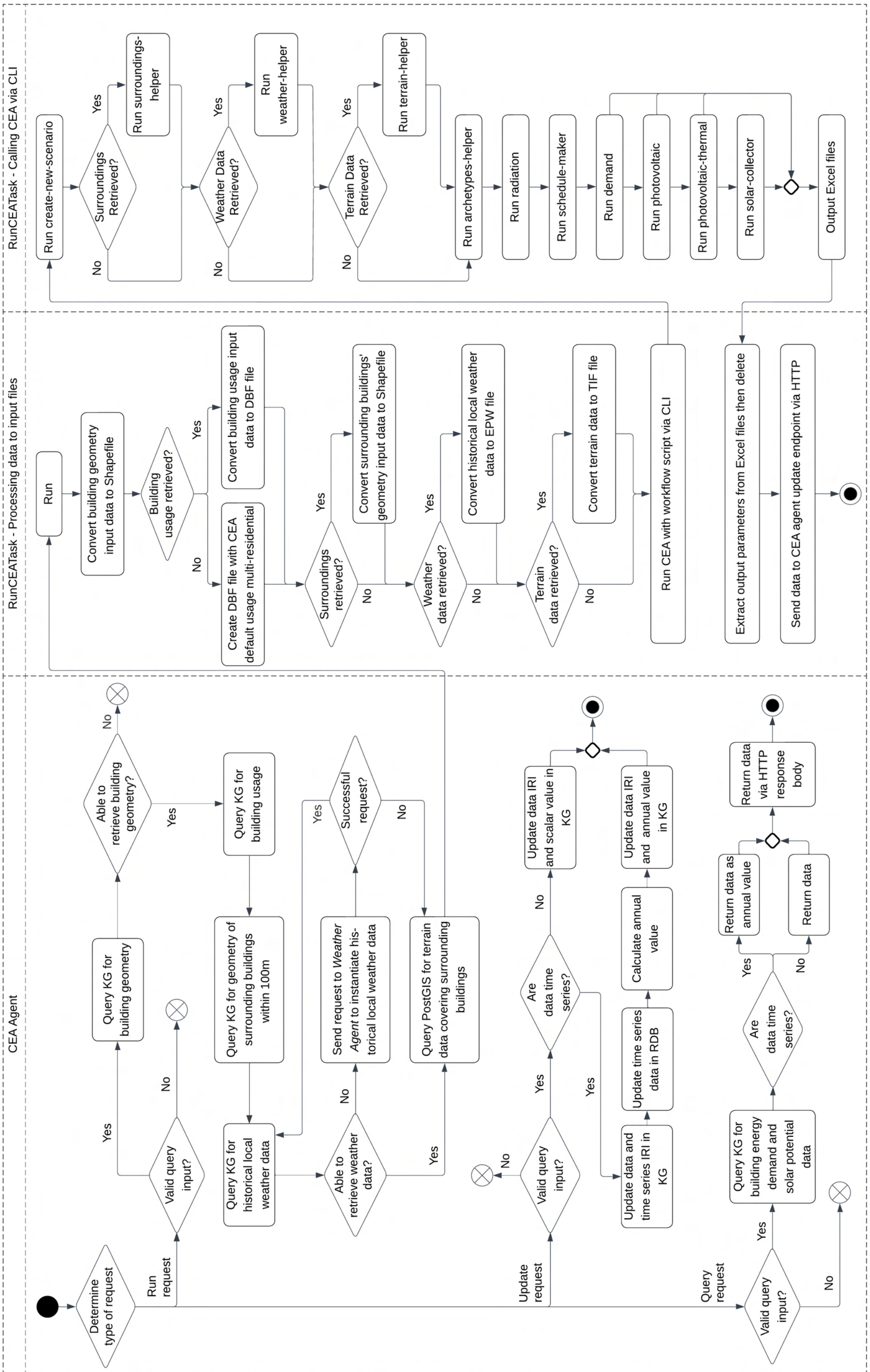



**Figure 24:** *CEA agent activity diagram. Overview of internal CEA agent logic for all three implemented routes: 1) run agent for new building analysis, 2) write/update analysis results into KG, 3) retrieve instantiated analysis results.*

## A.5 Derivation markup details

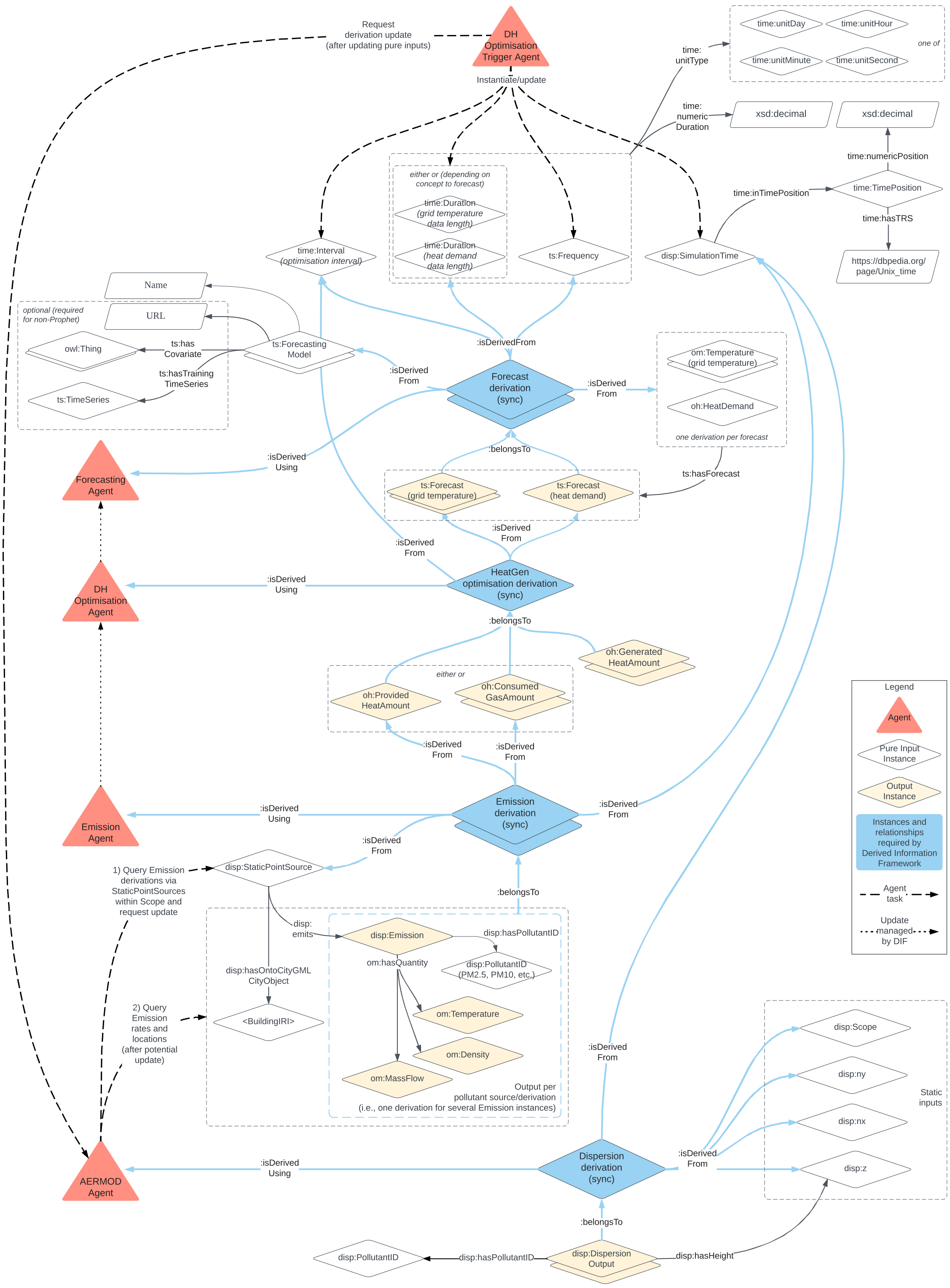



**Figure 25:** *Detailed illustration of actual derivation markup to chain agents.*

## A.6 Description logic

### A.6.1 Time series ontology

The latest version of the ontology is publicly available as OWL file on GitHub under: https://github.com/cambridge-cares/TheWorldAvatar/blob/main/JPS_Ontology/ontology/ontotimeseries/OntoTimeSeries.owl. A representation of the ontology in Description Logic is provided below:

**Classes:**

ts:InstantaneousTimeSeries $\sqsubseteq$ ts:TimeSeries
ts:AverageTimeSeries $\sqsubseteq$ ts:TimeSeries
ts:CumulativeTimeSeries $\sqsubseteq$ ts:TimeSeries
ts:CumulativeTotalTimeSeries $\sqsubseteq$ ts:CumulativeTimeSeries
ts:StepwiseCumulativeTimeSeries $\sqsubseteq$ ts:CumulativeTimeSeries
ts:Frequency $\sqsubseteq$ time:Duration

**Object properties:**

$\exists$ ts:hasTimeSeries.$\top$ $\sqsubseteq$ $\top$
$\exists$ ts:hasForecast.$\top$ $\sqsubseteq$ $\top$
$\exists$ ts:hasRDB.$\top$ $\sqsubseteq$ ts:TimeSeries
$\exists$ ts:hasTimeUnit.$\top$ $\sqsubseteq$ ts:TimeSeries
$\exists$ ts:hasAveragingPeriod.$\top$ $\sqsubseteq$ ts:AverageTimeSeries
$\exists$ ts:hasForecastingModel.$\top$ $\sqsubseteq$ ts:Forecast
$\exists$ ts:hasInputTimeInterval.$\top$ $\sqsubseteq$ ts:Forecast
$\exists$ ts:hasOutputTimeInterval.$\top$ $\sqsubseteq$ ts:Forecast
$\exists$ ts:hasCovariate.$\top$ $\sqsubseteq$ ts:ForecastingModel
$\exists$ ts:hasCheckpointURL.$\top$ $\sqsubseteq$ ts:ForecastingModel
$\exists$ ts:hasModelURL.$\top$ $\sqsubseteq$ ts:ForecastingModel
$\exists$ ts:hasTrainingTimeSeries.$\top$ $\sqsubseteq$ ts:ForecastingModel
$\exists$ ts:scaleData.$\top$ $\sqsubseteq$ ts:ForecastingModel
$\exists$ ts:resampleData.$\top$ $\sqsubseteq$ ts:Frequency
$\exists$ om:hasUnit.$\top$ $\sqsubseteq$ ts:Forecast
$\top$ $\sqsubseteq$ $\forall$ ts:hasTimeSeries.ts:TimeSeries
$\top$ $\sqsubseteq$ $\forall$ ts:hasForecast.ts:Forecast
$\top$ $\sqsubseteq$ $\forall$ ts:hasAveragingPeriod.time:Duration
$\top$ $\sqsubseteq$ $\forall$ ts:hasTrainingTimeSeries.ts:TimeSeries
$\top$ $\sqsubseteq$ $\forall$ ts:hasInputTimeInterval.time:Interval
$\top$ $\sqsubseteq$ $\forall$ ts:hasOutputTimeInterval.time:Interval
$\top$ $\sqsubseteq$ $\forall$ ts:hasForecastingModel.ts:ForecastingModel
$\top$ $\sqsubseteq$ $\forall$ ts:hasCovariate.$\top$

**Data properties:**

$\top \sqsubseteq \forall$ ts:hasCheckpointURL.xsd:string
$\top \sqsubseteq \forall$ ts:hasModelURL.xsd:string
$\top \sqsubseteq \forall$ ts:hasRDB.xsd:string
$\top \sqsubseteq \forall$ ts:hasTimeUnit.xsd:string
$\top \sqsubseteq \forall$ ts:resampleData.xsd:boolean
$\top \sqsubseteq \forall$ ts:scaleData.xsd:boolean

### A.6.2 District heating network ontology

The latest version of the ontology is publicly available as OWL file on GitHub under: `https://github.com/cambridge-cares/TheWorldAvatar/blob/main/JPS_Ontology/ontology/ontoheatnetwork/ontoheatnetwork.owl`. A representation of the ontology in Description Logic is provided below:

**Classes:**

oh:HeatingNetwork $\sqsubseteq = 1$ oh:hasMinFlowRate.om:VolumetricFlowRate
oh:Consumer $\sqsubseteq = 1$ oh:hasHeatDemand.oh:HeatDemand
oh:HeatProvider $\sqsubseteq = 1$ oh:hasDownstreamGridConnection.oh:GridConnection
oh:HeatProvider $\sqsubseteq = 1$ oh:hasUpstreamGridConnection.oh:GridConnection
oh:MunicipalUtility $\sqsubseteq$ oh:HeatProvider
oh:IncinerationPlant $\sqsubseteq$ oh:HeatProvider
oh:HeatGenerator $\sqsubseteq = 1$ oh:hasConsumedGasAmount.oh:ConsumedGasAmount
oh:HeatGenerator $\sqsubseteq = 1$ oh:hasGeneratedHeatAmount.oh:GeneratedHeatAmount
oh:HeatGenerator $\sqsubseteq = 1$ oh:hasOperatingTime.oh:DurationInTimeInterval
oh:HeatGenerator $\sqsubseteq = 1$ oh:hasRatedThermalPower.om:Power
oh:HeatBoiler $\sqsubseteq$ oh:HeatGenerator
oh:GasTurbine $\sqsubseteq$ oh:HeatGenerator
oh:GasTurbine $\sqsubseteq = 1$ oh:hasCoGenElectricityAmount.oh:CoGenElectricityAmount
oh:GasTurbine $\sqsubseteq = 1$ oh:hasMinimumIdleTime.om:Duration
oh:GasTurbine $\sqsubseteq = 1$ oh:hasMinimumThermalLoad.om:Power
oh:GasTurbine $\sqsubseteq = 1$ oh:hasRatedElectricalPower.om:Power
oh:EnergyInTimeInterval $\sqsubseteq$ om:Energy
oh:HeatDemand $\sqsubseteq$ oh:EnergyInTimeInterval
oh:ProvidedHeatAmount $\sqsubseteq$ oh:EnergyInTimeInterval
oh:GeneratedHeatAmount $\sqsubseteq$ oh:EnergyInTimeInterval
oh:ConsumedGasAmount $\sqsubseteq$ oh:EnergyInTimeInterval
oh:CoGenElectricityAmount $\sqsubseteq$ oh:EnergyInTimeInterval
oh:CO2EmissionInTimeInterval $\sqsubseteq$ om:Quantity
oh:CoGenRevenueInTimeInterval $\sqsubseteq$ om:AmountOfMoney
oh:DurationInTimeInterval $\sqsubseteq$ om:Duration
oh:NaturalGas $\sqsubseteq$ OntoChemPlant:FuelType
oh:CO2Factor $\sqsubseteq$ om:Quantity

oh:CalorificValue $\sqsubseteq$ om:Quantity
oh:HigherCalorificValue $\sqsubseteq$ oh:CalorificValue
oh:LowerCalorificValue $\sqsubseteq$ oh:CalorificValue
OntoChemPlant:FuelType $\sqsubseteq$ $=1$ oh:hasCO2Factor.oh:CO2Factor
OntoChemPlant:FuelType $\sqsubseteq$ $=1$ oh:hasUnitPrice.oh:FuelUnitCost
OntoChemPlant:FuelType $\sqsubseteq$ $=1$ oh:hasHigherCalorificValue.oh:HigherCalorificValue
OntoChemPlant:FuelType $\sqsubseteq$ $=1$ oh:hasLowerCalorificValue.oh:LowerCalorificValue
OntoCAPE_Phase_System:ThermodynamicStateProperty $\sqsubseteq$ OntoCAPE_System:Property
om:Density $\sqsubseteq$ OntoCAPE_Phase_System:ThermodynamicStateProperty
om:Pressure $\sqsubseteq$ OntoCAPE_Phase_System:ThermodynamicStateProperty
om:SpecificHeatCapacity $\sqsubseteq$ OntoCAPE_Phase_System:ThermodynamicStateProperty
om:Temperature $\sqsubseteq$ OntoCAPE_Phase_System:ThermodynamicStateProperty
FIBO_Contracts:Contract $\sqsubseteq$ $=1$ oh:hasCurrentUnitPrice.oh:UnitPrice
FIBO_Contracts:Contract $\sqsubseteq$ $=1$ oh:hasMaxAnnualPurchaseVolume.om:Energy
FIBO_Contracts:Contract $\sqsubseteq$ $=1$ oh:hasMinAnnualPurchaseVolume.om:Energy
oh:Tier $\sqsubseteq$ $=1$ oh:hasCumulativeEnergyCap.om:Energy
oh:Tier $\sqsubseteq$ $=1$ oh:hasUnitPrice.oh:UnitPrice
oh:UnitRate $\sqsubseteq$ om:Quantity
oh:UnitPrice $\sqsubseteq$ oh:UnitRate
oh:CHPBonus $\sqsubseteq$ oh:UnitRate
oh:CO2CertificatePrice $\sqsubseteq$ oh:UnitRate
oh:ElectricitySpotPrice $\sqsubseteq$ oh:UnitRate
oh:FuelUnitCost $\sqsubseteq$ oh:UnitRate
oh:GasUnitCost $\sqsubseteq$ oh:FuelUnitCost
oh:GridCharges $\sqsubseteq$ oh:UnitRate
oh:HourlyLabourCost $\sqsubseteq$ oh:UnitRate
oh:HourlyWearCost $\sqsubseteq$ oh:UnitRate
oh:DemandDrivenWearCost $\sqsubseteq$ oh:UnitRate
om:Cost $\equiv$ OntoCAPE_Economics:Costs
oh:CostInTimeInterval $\sqsubseteq$ om:Cost
oh:FixedCost $\sqsubseteq$ oh:CostInTimeInterval
oh:VariableCost $\sqsubseteq$ oh:CostInTimeInterval
oh:LabourCost $\sqsubseteq$ oh:FixedCost
oh:FixedWearCost $\sqsubseteq$ oh:FixedCost
oh:FuelCost $\sqsubseteq$ oh:VariableCost
oh:EmissionCost $\sqsubseteq$ oh:VariableCost
oh:VariableWearCost $\sqsubseteq$ oh:VariableCost
oh:SwitchingCost $\sqsubseteq$ oh:VariableCost
oh:ShutDownCost $\sqsubseteq$ oh:SwitchingCost
oh:StartUpCost $\sqsubseteq$ oh:SwitchingCost
oh:isPublicHoliday $\sqsubseteq$ oh:CalendarEffect
oh:isSchoolVacation $\sqsubseteq$ oh:CalendarEffect

**Object properties:**

$\exists$ OntoMeta:hasPart.$\top \sqsubseteq$ oh:HeatingNetwork
$\exists$ oh:suppliesHeatTo.$\top \sqsubseteq$ oh:HeatingNetwork
$\exists$ oh:hasMinFlowRate.$\top \sqsubseteq$ oh:HeatingNetwork
$\exists$ oh:hasHeatDemand.$\top \sqsubseteq$ oh:Consumer
$\exists$ oh:providesHeatTo.$\top \sqsubseteq$ oh:HeatProvider
$\exists$ oh:hasUpstreamGridConnection.$\top \sqsubseteq$ oh:HeatProvider
$\exists$ oh:hasDownstreamGridConnection.$\top \sqsubseteq$ oh:HeatProvider
$\exists$ oh:hasObservableProperty.$\top \sqsubseteq$ oh:GridConnection
$\exists$ OntoCAPE_System:isOwnerOf.$\top \sqsubseteq$ oh:MunicipalUtility
$\exists$ oh:hasPurchaseAgreement.$\top \sqsubseteq$ oh:MunicipalUtility
$\exists$ oh:operates.$\top \sqsubseteq$ oh:MunicipalUtility
$\exists$ OntoCAPE_Economics:hasCost.$\top \sqsubseteq$ (oh:HeatGenerator $\sqcup$ oh:MunicipalUtility)
$\exists$ OntoChemPlant:hasFuelType.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ oh:hasRatedThermalPower.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ oh:applicableOPEXComponent.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ oh:hasGeneratedHeatAmount.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ oh:hasConsumedGasAmount.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ oh:hasOperatingTime.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ OntoChemPlant:hasIndividualCO2Emission.$\top \sqsubseteq$ oh:HeatGenerator
$\exists$ oh:hasOperatingAvailability.$\top \sqsubseteq$ (oh:HeatGenerator $\sqcup$ oh:IncinerationPlant)
$\exists$ oh:hasMinimumThermalLoad.$\top \sqsubseteq$ oh:GasTurbine
$\exists$ oh:hasMinimumIdleTime.$\top \sqsubseteq$ oh:GasTurbine
$\exists$ oh:hasRatedElectricalPower.$\top \sqsubseteq$ oh:GasTurbine
$\exists$ oh:hasCoGenElectricityAmount.$\top \sqsubseteq$ oh:GasTurbine
$\exists$ OntoCAPE_Economics:hasRevenue.$\top \sqsubseteq$ oh:GasTurbine
$\exists$ oh:hasCO2Factor.$\top \sqsubseteq$ OntoChemPlant:FuelType
$\exists$ oh:hasHigherCalorificValue.$\top \sqsubseteq$ OntoChemPlant:FuelType
$\exists$ oh:hasLowerCalorificValue.$\top \sqsubseteq$ OntoChemPlant:FuelType
$\exists$ oh:hasUnitPrice.$\top \sqsubseteq$ OntoChemPlant:FuelType
$\exists$ oh:hasMaxHourlySupply.$\top \sqsubseteq$ oh:IncinerationPlant
$\exists$ oh:hasMinHourlySupply.$\top \sqsubseteq$ oh:IncinerationPlant
$\exists$ oh:hasProvidedHeatAmount.$\top \sqsubseteq$ oh:IncinerationPlant
$\exists$ oh:isFulfilledBy.$\top \sqsubseteq$ FIBO_Contracts:Contract
$\exists$ oh:hasMaxAnnualPurchaseVolume.$\top \sqsubseteq$ FIBO_Contracts:Contract
$\exists$ oh:hasMinAnnualPurchaseVolume.$\top \sqsubseteq$ FIBO_Contracts:Contract
$\exists$ oh:hasCurrentUnitPrice.$\top \sqsubseteq$ FIBO_Contracts:Contract
$\exists$ oh:hasTieredUnitPrice.$\top \sqsubseteq$ FIBO_Contracts:Contract
$\exists$ oh:hasTier.$\top \sqsubseteq$ oh:TieredUnitPrice
$\exists$ oh:hasCumulativeEnergyCap.$\top \sqsubseteq$ oh:Tier
$\exists$ oh:hasUnitPrice.$\top \sqsubseteq$ oh:Tier
$\exists$ oh:applicableLocation.$\top \sqsubseteq$ oh:CalendarEffect
$\top \sqsubseteq \forall$ oh:operates.oh:HeatingNetwork
$\top \sqsubseteq \forall$ oh:providesHeatTo.oh:HeatingNetwork
$\top \sqsubseteq \forall$ oh:hasMinFlowRate.om:VolumetricFlowRate

$\top \sqsubseteq \forall$ oh:suppliesHeatTo.oh:Consumer
$\top \sqsubseteq \forall$ oh:hasHeatDemand.oh:HeatDemand
$\top \sqsubseteq \forall$ OntoCAPE_System:isOwnerOf.(OntoPowSys:ThermalEnergyStorage $\sqcup$ oh:HeatGenerator)
$\top \sqsubseteq \forall$ OntoMeta:hasPart.oh:GridConnection
$\top \sqsubseteq \forall$ oh:hasUpstreamGridConnection.oh:GridConnection
$\top \sqsubseteq \forall$ oh:hasDownstreamGridConnection.oh:GridConnection
$\top \sqsubseteq \forall$ oh:hasObservableProperty.OntoCAPE_Phase_System: ThermodynamicStateProperty
$\top \sqsubseteq \forall$ oh:hasPurchaseAgreement.FIBO_Contracts:Contract
$\top \sqsubseteq \forall$ oh:hasMaxAnnualPurchaseVolume.om:Energy
$\top \sqsubseteq \forall$ oh:hasMinAnnualPurchaseVolume.om:Energy
$\top \sqsubseteq \forall$ oh:isFulfilledBy.oh:HeatProvider
$\top \sqsubseteq \forall$ oh:hasMaxHourlySupply.om:Power
$\top \sqsubseteq \forall$ oh:hasMinHourlySupply.om:Power
$\top \sqsubseteq \forall$ oh:hasTieredUnitPrice.oh:TieredUnitPrice
$\top \sqsubseteq \forall$ oh:hasCurrentUnitPrice.oh:UnitPrice
$\top \sqsubseteq \forall$ oh:hasTier.oh:Tier
$\top \sqsubseteq \forall$ oh:hasUnitPrice.oh:UnitPrice
$\top \sqsubseteq \forall$ oh:hasCumulativeEnergyCap.om:Energy
$\top \sqsubseteq \forall$ oh:hasRatedThermalPower.om:Power
$\top \sqsubseteq \forall$ oh:hasRatedElectricalPower.om:Power
$\top \sqsubseteq \forall$ oh:hasMinimumThermalLoad.om:Power
$\top \sqsubseteq \forall$ oh:hasMinimumIdleTime.om:Duration
$\top \sqsubseteq \forall$ oh:hasGeneratedHeatAmount.oh:GeneratedHeatAmount
$\top \sqsubseteq \forall$ oh:hasConsumedGasAmount.oh:ConsumedGasAmount
$\top \sqsubseteq \forall$ oh:hasCoGenElectricityAmount.oh:CoGenElectricityAmount
$\top \sqsubseteq \forall$ oh:hasProvidedHeatAmount.oh:ProvidedHeatAmount
$\top \sqsubseteq \forall$ oh:hasOperatingAvailability.oh:Availability
$\top \sqsubseteq \forall$ oh:hasOperatingTime.oh:DurationInTimeInterval
$\top \sqsubseteq \forall$ OntoCAPE_Economics:hasCost.om:Cost
$\top \sqsubseteq \forall$ OntoCAPE_Economics:hasRevenue.oh:CoGenRevenueInTimeInterval
$\top \sqsubseteq \forall$ oh:applicableOPEXComponent.oh:UnitRate
$\top \sqsubseteq \forall$ oh:hasUnitPrice.oh:FuelUnitCost
$\top \sqsubseteq \forall$ OntoChemPlant:hasFuelType.OntoChemPlant:FuelType
$\top \sqsubseteq \forall$ oh:hasCO2Factor.oh:CO2Factor
$\top \sqsubseteq \forall$ oh:hasHigherCalorificValue.oh:HigherCalorificValue
$\top \sqsubseteq \forall$ oh:hasLowerCalorificValue.oh:LowerCalorificValue
$\top \sqsubseteq \forall$ OntoChemPlant:hasIndividualCO2Emission.oh:CO2EmissionInTimeInterval
$\top \sqsubseteq \forall$ oh:applicableLocation.http://purl.org/dc/terms/Location

### A.6.3 Emission dispersion ontology

The latest version of the ontology is publicly available as OWL file on GitHub under: https://github.com/cambridge-cares/TheWorldAvatar/blob/main/JPS_O ntology/ontology/ontodispersion/OntoDispersion.owl. A representation of the ontology in Description Logic is provided below:

**Classes:**

disp:CO $\sqsubseteq$ disp:PollutantID
disp:CO2 $\sqsubseteq$ disp:PollutantID
disp:NO2 $\sqsubseteq$ disp:PollutantID
disp:NOx $\sqsubseteq$ disp:PollutantID
disp:PM10 $\sqsubseteq$ disp:PollutantID
disp:PM2.5 $\sqsubseteq$ disp:PollutantID
disp:SO2 $\sqsubseteq$ disp:PollutantID
disp:uHC $\sqsubseteq$ disp:PollutantID
disp:SimulationTime $\sqsubseteq$ time:Instant
disp:nx $\sqsubseteq$ om:Quantity
disp:ny $\sqsubseteq$ om:Quantity
disp:z $\sqsubseteq$ om:Height
disp:Scope $\sqsubseteq$ geo:Feature
disp:StaticPointSource $\sqsubseteq$ disp:PointSource
disp:DynamicPointSource $\sqsubseteq$ disp:PointSource

**Object properties:**

$\exists$ om:hasQuantity.$\top$ $\sqsubseteq$ disp:Emission
$\exists$ disp:hasPollutantID.$\top$ $\sqsubseteq$ disp:Emission
$\exists$ disp:emits.$\top$ $\sqsubseteq$ disp:PointSource
$\exists$ disp:hasOntoCityGMLCityObject.$\top$ $\sqsubseteq$ disp:StaticPointSource
$\exists$ disp:hasPollutantID.$\top$ $\sqsubseteq$ disp:DispersionOutput
$\exists$ disp:hasHeight.$\top$ $\sqsubseteq$ disp:DispersionOutput
$\exists$ disp:hasDispersionMatrix.$\top$ $\sqsubseteq$ disp:DispersionOutput
$\exists$ disp:hasDispersionRaster.$\top$ $\sqsubseteq$ disp:DispersionOutput
$\exists$ disp:hasDispersionColourBar.$\top$ $\sqsubseteq$ disp:DispersionOutput
$\exists$ disp:hasValue.$\top$ $\sqsubseteq$ disp:DispersionPolygon
$\exists$ disp:hasName.$\top$ $\sqsubseteq$ disp:OntoCityGMLNamespace
$\top$ $\sqsubseteq$ $\forall$ disp:hasOntoCityGMLCityObject.$\top$
$\top$ $\sqsubseteq$ $\forall$ disp:emits.disp:Emission
$\top$ $\sqsubseteq$ $\forall$ disp:hasPollutantID.disp:PollutantID
$\top$ $\sqsubseteq$ $\forall$ om:hasQuantity.om:Density
$\top$ $\sqsubseteq$ $\forall$ om:hasQuantity.om:MassFlow
$\top$ $\sqsubseteq$ $\forall$ om:hasQuantity.om:Temperature
$\top$ $\sqsubseteq$ $\forall$ disp:hasHeight.disp:z
$\top$ $\sqsubseteq$ $\forall$ disp:hasDispersionMatrix.disp:DispersionMatrix
$\top$ $\sqsubseteq$ $\forall$ disp:hasDispersionRaster.disp:DispersionRaster
$\top$ $\sqsubseteq$ $\forall$ disp:hasDispersionColourBar.disp:DispersionColourBar

**Data properties:**

$\top \sqsubseteq \forall$ disp:hasName.xsd:string
$\top \sqsubseteq \forall$ disp:hasValue.xsd:string

# References


[1] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal Digital Twin - A Dynamic Knowledge Graph. *Data-Centric Engineering*, 2:e14, 2021. doi:10.1017/dce.2021.10.

[2] S. B. Almobydeen, J. R. Viqueira, and M. Lama. GeoSPARQL query support for scientific raster array data. *Computers & Geosciences*, 159:105023, 2022. doi:10.1016/j.cageo.2021.105023.

[3] O. Angelidis, A. Ioannou, D. Friedrich, A. Thomson, and G. Falcone. District heating and cooling networks with decentralised energy substations: Opportunities and barriers for holistic energy system decarbonisation. *Energy*, 269:126740, 2023. doi:10.1016/j.energy.2023.126740.

[4] C. B. Aranda, O. Corby, S. Das, L. Feigenbaum, P. Gearon, B. Glimm, S. Harris, S. Hawke, I. Herman, N. Humfrey, N. Michaelis, C. Ogbuji, M. Perry, A. Passant, A. Polleres, E. Prud'hommeaux, A. Seaborne, and G. T. Williams. SPARQL 1.1 Overview, W3C Recommendation 21 March 2013. World Wide Web Consortium (W3C), 2013. Available at `https://www.w3.org/TR/sparql11-overview/` (accessed Dec 2023).

[5] N. Aste, R. Adhikari, J. Compostella, and C. D. Pero. Energy and environmental impact of domestic heating in Italy: Evaluation of national $NO_x$ emissions. *Energy Policy*, 53:353–360, 2013. doi:10.1016/j.enpol.2012.10.064.

[6] J. Atherton, J. Akroyd, F. Farazi, S. Mosbach, M. Q. Lim, and M. Kraft. British wind farm ESS attachments: curtailment reduction vs. price arbitrage. *Energy & Environmental Science*, 16(9):4020–4040, 2023. doi:10.1039/d3ee01355c.

[7] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook*. Cambridge University Press, Cambridge, UK, 2nd edition, 2007. ISBN 978-0-521-87625-4.

[8] J. Bai, K. F. Lee, M. Hofmeister, S. Mosbach, J. Akroyd, and M. Kraft. A derived information framework for a dynamic knowledge graph and its application to smart cities. *Future Generation Computer Systems*, 152:112–126, 2024. doi:10.1016/j.future.2023.10.008.

[9] T. Becker, C. Nagel, and T. H. Kolbe. *Semantic 3D Modeling of Multi-Utility Networks in Cities for Analysis and 3D Visualization*, page 41–62. Springer Berlin Heidelberg, Oct. 2012. doi:10.1007/978-3-642-29793-9_3.

[10] K. Bereta, G. Xiao, and M. Koubarakis. Ontop-spatial: Ontop of geospatial databases. *Journal of Web Semantics*, 58:100514, 2019. doi:10.1016/j.websem.2019.100514.

[11] A. Berjawi, S. Walker, C. Patsios, and S. Hosseini. An evaluation framework for future integrated energy systems: A whole energy systems approach. *Renewable and Sustainable Energy Reviews*, 145:111163, July 2021. doi:10.1016/j.rser.2021.111163.

[12] T. Berners-Lee. Linked data - design issues, 2006. Available at `http://www.w3.org/DesignIssues/LinkedData.html` (accessed Dec 2023).

[13] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):28–37, 2001.

[14] C. Bizer, T. Heath, and T. Berners-Lee. Linked data: The story so far. In A. Sheth, editor, *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI Global, 2011. doi:10.4018/978-1-60960-593-3.ch008.

[15] Blazegraph, 2020. Available at `https://blazegraph.com` (accessed Dec 2023).

[16] E. Botoeva, D. Calvanese, B. Cogrel, M. Rezk, and G. Xiao. OBDA beyond relational DBs: A study for MongoDB. In M. Lenzerini and R. Peñaloza, editors, *Proc. of the 29th Int. Workshop on Description Logics (DL 2016)*, volume 1577 of *CEUR Workshop Proceedings*, `http://ceur-ws.org`, 2016.

[17] D. J. Carruthers, C. A. McHugh, E. Vanvyve, and E. Solazzo. Comparison of ADMS and AERMOD meteorological preprocessor and dispersion algorithms. *Proceedings of Air & Waste Management Association, Air Quality Models: Next Generation of Models*, pages 28–30, 2009.

[18] D. J. Carruthers, M. D. Seaton, C. A. McHugh, X. Sheng, E. Solazzo, and E. Vanvyve. Comparison of the complex terrain algorithms incorporated into two commonly used local-scale air pollution dispersion models (adms and aermod) using a hybrid model. *Journal of the Air & Waste Management Association*, 61(11): 1227–1235, 2011. doi:10.1080/10473289.2011.609750.

[19] 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, Dec. 2021. doi:10.1016/j.egyai.2021.100106.

[20] A. Chadzynski, S. Li, A. Grisiute, J. Chua, M. Hofmeister, J. Yan, H. Y. Tai, E. Lloyd, Y. K. Tsai, M. Agarwal, J. Akroyd, P. Herthogs, and M. Kraft. Semantic 3D city interfaces—Intelligent interactions on dynamic geospatial knowledge graphs. *Data-Centric Engineering*, 4:e20, 2023. doi:10.1017/dce.2023.14.

[21] A. J. Cimorelli, S. G. Perry, A. Venkatram, J. C. Weil, R. J. Paine, R. B. Wilson, R. F. Lee, W. D. Peters, and R. W. Brode. AERMOD: A Dispersion Model for Industrial Source Applications. Part I: General Model Formulation and Boundary Layer Characterization. *Journal of Applied Meteorology*, 44(5):682–693, 2005. doi:10.1175/jam2227.1.

[22] S. Cox, C. Little, J. R. Hobbs, and F. Pan. Time Ontology in OWL, 2020. Available at `https://www.w3.org/TR/owl-time/` (accessed Dec 2023).

[23] J. Cuenca, F. Larrinaga, and E. Curry. A Unified Semantic Ontology for Energy Management Applications. In *WSP/WOMoCoE@ISWC*, 2017. URL `https://api.semanticscholar.org/CorpusID:19853342`.

[24] J. Cuenca, F. Larrinaga, and E. Curry. DABGEO: A reusable and usable global energy ontology for the energy domain. *Journal of Web Semantics*, 61–62:100550, Mar. 2020. doi:10.1016/j.websem.2020.100550.

[25] Data & Knowledge Group. HermiT OWL Reasoner, 2019. Available at `http://www.hermit-reasoner.com/` (accessed Apr 2023).

[26] I. A. de Castro Villela and J. L. Silveira. Ecological efficiency in thermoelectric power plants. *Applied Thermal Engineering*, 27(5-6):840–847, 2007. doi:10.1016/j.applthermaleng.2006.09.019.

[27] A. Devanand, G. Karmakar, N. Krdzavac, R. Rigo-Mariani, E. Y. S. Foo, I. A. Karimi, and M. Kraft. OntoPowSys: A power system ontology for cross domain interactions in an eco industrial park. *Energy and AI*, 1:100008, 2020. doi:10.1016/j.egyai.2020.100008.

[28] Digital Twin Hub. Climate Resilience Demonstrator, 2022. Available at `https://digitaltwinhub.co.uk/credo/credo/` (accessed Dec 2023).

[29] Digital4Grids for the European Commission. Landscape report on energy and flexibility data models and interoperability across the sectors of energy, mobility and buildings, 2023. Available at `https://digital-strategy.ec.europa.eu/en/library/landscape-report-energy-and-flexibility-data-models-and-interoperability-across-sectors-energy` (accessed Jan 2024).

[30] L. Ding, G. Xiao, A. Pano, M. Fumagalli, D. Chen, Y. Feng, D. Calvanese, H. Fan, and L. Meng. Integrating 3D City Data through Knowledge Graphs, 2023. URL `https://arxiv.org/abs/2310.11555`. Submitted for publication. Available from `https://arxiv.org/abs/2310.11555`.

[31] EEW Energy from Waste Saarbrücken GmbH. Taking the lead at the Pirmasens site, 2023. Available at `https://www.eew-energyfromwaste.com/fileadmin/content/infomaterial/standorte_-_eng/eew_pirmasens_gb.pdf` (accessed Nov 2023).

[32] EEW Energy from Waste Saarbrücken GmbH. Emissionen des Müllheizkraftwerkes Pirmasens, 2023. Available at `https://www.eew-energyfromwaste.com/fileadmin/content/infomaterial/emissionsdaten/pirmasens/2023-03-22_veroeffentlichung_emissionen_mhkw_pir_2022.pdf` (accessed Jan 2024).

[33] A. Eibeck, M. Q. Lim, and M. Kraft. J-Park Simulator: An ontology-based platform for cross-domain scenarios in process industry. *Computers & Chemical Engineering*, 131:106586, 2019. doi:10.1016/j.compchemeng.2019.106586.

[34] T. El-Diraby and H. Osman. A domain ontology for construction concepts in urban infrastructure products. *Automation in Construction*, 20(8):1120–1132, Dec. 2011. doi:10.1016/j.autcon.2011.04.014.

[35] M. El-Harbawi. Air quality modelling, simulation, and computational methods: a review. *Environmental Reviews*, 21(3):149–179, 2013. doi:10.1139/er-2012-0056.

[36] Energy Justice Network. Trash Incineration More Polluting than Coal, 2019. Available at `https://www.energyjustice.net/incineration/worsethancoal` (accessed Aug 2023).

[37] ETSI. SAREF: the Smart Applications REFerence ontology, 2020. Available at `https://saref.etsi.org/core/` (accessed Dec 2023).

[38] ETSI. SAREF4EHAW: an extension of SAREF for eHealth Ageing Well domain, 2020. Available at `https://saref.etsi.org/saref4ehaw/v1.1.1/` (accessed Dec 2023).

[39] European Environment Agency. Air quality standards, 2021. Available at `https://www.eea.europa.eu/themes/air/air-quality-concentrations/air-quality-standards` (accessed Aug 2023).

[40] European Environment Agency. AQ eReporting - Annual statistics, 2023. Available at `https://eeadmz1-cws-wp-air02.azurewebsites.net/index.php/users-corner/statistical-viewer/` (accessed Aug 2023).

[41] M. Fodstad, P. Crespo del Granado, L. Hellemo, B. R. Knudsen, P. Pisciella, A. Silvast, C. Bordin, S. Schmidt, and J. Straus. Next frontiers in energy system modelling: A review on challenges and the state of the art. *Renewable and Sustainable Energy Reviews*, 160:112246, May 2022. doi:10.1016/j.rser.2022.112246.

[42] J. A. Fonseca, N. Thuy-An, 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, Feb. 2016. doi:10.1016/j.enbuild.2015.11.055.

[43] Y. Hamdani, G. Xiao, L. Ding, and D. Calvanese. An Ontology-Based Framework for Geospatial Integration and Querying of Raster Data Cube Using Virtual Knowledge Graphs. *ISPRS International Journal of Geo-Information*, 12(9):375, 2023. doi:10.3390/ijgi12090375.

[44] P. D. Hamer, S.-E. Walker, G. Sousa-Santos, M. Vogt, D. Vo-Thanh, S. Lopez-Aparicio, P. Schneider, M. O. P. Ramacher, and M. Karl. The urban dispersion model EPISODE v10.0 – Part 1: An Eulerian and sub-grid-scale air quality model and its application in Nordic winter conditions. *Geoscientific Model Development*, 13(9):4323–4353, 2020. doi:10.5194/gmd-13-4323-2020.

[45] G. Happle, J. A. Fonseca, and A. Schlueter. Effects of air infiltration modeling approaches in urban building energy demand forecasts. *Energy Procedia*, 122:283–288, Sept. 2017. doi:10.1016/j.egypro.2017.07.323.

[46] G. Happle, Z. Shi, S. Hsieh, B. Ong, J. A. Fonseca, and A. Schlueter. Identifying carbon emission reduction potentials of BIPV in high-density cities in South-east Asia. *Journal of Physics: Conference Series*, 1343(1):012077, Nov. 2019. doi:10.1088/1742-6596/1343/1/012077.

[47] J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. V. Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, and G. Grosch. Darts: User-Friendly Modern Machine Learning for Time Series. *Journal of Machine Learning Research*, 23(124):1–6, 2022. URL http://jmlr.org/papers/v23/21-1177.html.

[48] J.-L. Hippolyte, Y. Rezgui, H. Li, B. Jayan, and S. Howell. Ontology-driven development of web services to support district energy applications. *Automation in Construction*, 86:210–225, 2018. doi:10.1016/j.autcon.2017.10.004.

[49] L. Hofbauer, W. McDowall, and S. Pye. Challenges and opportunities for energy system modelling to foster multi-level governance of energy transitions. *Renewable and Sustainable Energy Reviews*, 161:112330, June 2022. doi:10.1016/j.rser.2022.112330.

[50] M. Hofmeister, S. Mosbach, J. Hammacher, M. Blum, G. Röhrig, C. Dörr, V. Flegel, A. Bhave, and M. Kraft. Resource-Optimised Generation Dispatch Strategy for District Heating Systems Using Dynamic Hierarchical Optimisation. *Applied Energy*, 305:117877, 2022. doi:10.1016/j.apenergy.2021.117877.

[51] M. Hofmeister, J. Bai, G. Brownbridge, S. Mosbach, K. F. Lee, F. Farazi, M. Hillman, M. Agarwal, S. Ganguly, J. Akroyd, and M. Kraft. Semantic Agent Framework for Automated Flood Assessment Using Dynamic Knowledge Graphs, 2023. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/309/.

[52] M. Hofmeister, G. Brownbridge, M. Hillman, S. Mosbach, J. Akroyd, K. F. Lee, and M. Kraft. Cross-domain flood risk assessment for smart cities using dynamic knowledge graphs. *Sustainable Cities and Society*, page 105113, Dec. 2023. doi:10.1016/j.scs.2023.105113.

[53] N. Holmes and L. Morawska. A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment*, 40(30):5902–5928, 2006. doi:10.1016/j.atmosenv.2006.06.003.

[54] Hotmaps project. Hotmaps, 2020. Available at https://www.hotmaps.eu/map (accessed Dec 2023).

[55] O. Inderwildi, C. Zhang, X. Wang, and M. Kraft. The impact of intelligent cyber-physical systems on the decarbonization of energy. *Energy & Environmental Science*, 13(3):744–771, 2020. doi:10.1039/c9ee01919g.

[56] Intergovernmental Panel on Climate Change. IPCC Emission Factor Database, 2017. Available at https://www.ipcc-nggip.iges.or.jp/EFDB/find_ef.php (accessed Aug 2023).

[57] int:net consortium. The Interoperability Network for the Energy Transition (int:net), 2022. Available at https://intnet.eu/ (accessed Dec 2023).

[58] M. Ittershagen. The role of waste incineration in Germany. Umweltbundesamt, 2008. Available at https://www.umweltbundesamt.de/sites/default/files/medien/publikation/long/3872.pdf (accessed Aug 2023).

[59] K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, and M. Lefrançois. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, 56:1–10, 2019. doi:10.1016/j.websem.2018.06.003.

[60] B. Johnke, R. Hoppaus, E. Lee, B. Irving, T. Martinsen, and K. Mareckova. Emissions from waste incineration. In *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories*, 2000. Available at https://www.ipcc-nggip.iges.or.jp/public/gp/bgp/5_3_Waste_Incineration.pdf (accessed Aug 2023).

[61] K. Juhrich and R. Beckers. Aktualisierung der Emissionsfaktoren für Großfeuerungsanlagen . Umweltbundesamt, 2019. Available at https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-11-29_texte_140-2019_emissionsfaktoren-grossfeuerungsanlagen.pdf (accessed Aug 2023).

[62] S. Khan and Q. Hassan. Review of developments in air quality modelling and air quality dispersion models. *Journal of Environmental Engineering and Science*, 16 (1):1–10, 2020. doi:10.1680/jenes.20.00004.

[63] G. Klyne and J. J. Carroll. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 February 2004. World Wide Web Consortium (W3C), 2004. Available at http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ (accessed Dec 2023).

[64] M. Krause and S. Smith. Methodology to Calculate Particulate Matter (PM) 2.5 and PM 2.5 Significance Thresholds. Technical report, South Coast Air Quality Management District, 2006. Available at http://www.aqmd.gov/docs/default-source/ceqa/handbook/localized-significance-thresholds/particulate-matter-(pm)-2.5-significance-thresholds-and-calculation-methodology/final_pm2_5methodology.pdf (accessed Aug 2023).

[65] Landesamts für Vermessung und Geobasisinformation Rheinland-Pfalz. Rheinland-Pfalz in 3D, 2023. Available at http://www.rheinland-pfalz-in-3d.rlp.de/ (accessed Feb 2023).

[66] H. Li and T. Hong. A semantic ontology for representing and quantifying energy flexibility of buildings. *Advances in Applied Energy*, 8:100113, 2022. doi:10.1016/j.adapen.2022.100113.

[67] Y. Li, Y. Rezgui, and S. Kubicki. An intelligent semantic system for real-time demand response management of a thermal grid. *Sustainable Cities and Society*, 52:101857, 2020. doi:10.1016/j.scs.2019.101857.

[68] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021. doi:10.1016/j.ijforecast.2021.03.012.

[69] M. Q. Lim, X. Wang, O. Inderwildi, and M. Kraft. *The World Avatar—A World Model for Facilitating Interoperability*, pages 39–53. Springer International Publishing, Cham, 2022. doi:10.1007/978-3-030-86215-2_4.

[70] A. S. S. Medeiros, G. Calderaro, P. C. Guimarães, M. R. Magalhaes, M. V. B. Morais, S. A. A. Rafee, I. O. Ribeiro, R. V. Andreoli, J. A. Martins, L. D. Martins, S. T. Martin, and R. A. F. Souza. Power plant fuel switching and air quality in a tropical, forested environment. *Atmospheric Chemistry and Physics*, 17(14):8987–8998, 2017. doi:10.5194/acp-17-8987-2017.

[71] P. Mehta, D. Griego, A. Nunez-Jimenez, and A. Schlueter. The Impact of self-consumption regulation on individual and community solar PV adoption in Switzerland: an agent-based model. *Journal of Physics: Conference Series*, 1343: 012143, Nov. 2019. doi:10.1088/1742-6596/1343/1/012143.

[72] C. Metral, G. Falquet, and K. Karatzas. Ontologies for the integration of air quality models and 3D city models. *arXiv preprint arXiv:1201.6511*, 2012. doi:10.48550/arXiv.1201.6511.

[73] Ministerium für Klimaschutz, Umwelt, Energie und Mobilität. Landesweites Solarkataster Rheinland-Pfalz, 2023. Available at `https://solarkataster.rlp.de/` (accessed Jan 2024).

[74] M. M. Mokhtar, M. H. Hassim, and R. M. Taib. Health risk assessment of emissions from a coal-fired power plant using AERMOD modelling. *Process Safety and Environmental Protection*, 92(5):476–485, 2014. doi:10.1016/j.psep.2014.05.008.

[75] M. Mosteiro-Romero, J. A. Fonseca, and A. Schlueter. Seasonal effects of input parameters in urban-scale building energy simulation. *Energy Procedia*, 122: 433–438, Sept. 2017. doi:10.1016/j.egypro.2017.07.459.

[76] M. Mosteiro-Romero, D. Maiullari, F. Collins, A. Schlueter, and A. V. Timmeren. District-scale energy demand modeling and urban microclimate: A case study in The Netherlands. *Journal of Physics: Conference Series*, 1343(1):012003, Nov. 2019. doi:10.1088/1742-6596/1343/1/012003.

[77] F. Neuwahl, G. Cusano, J. G. Benavides, S. Holbrook, and S. Roudier. Best Available Techniques (BAT) Reference Document for Waste Incineration. European Commission, Joint Research Centre, 2019. doi:10.2760/761437.

[78] O.-K. Nielsen, M. Nielsen, K. Hjelgaard, P. Coleman, O. Rentz, D. Oertel, H. Jones, M. Wenborn, and M. Woodfield. Incineration of domestic or municipal wastes (without energy recovery). In *EMEP/EEA air pollutant emission inventory guidebook 2019*. European Environment Agency, 2019. Available at `https://www.eea.europa.eu/publications/emep-eea-guidebook-`

2019/part-b-sectoral-guidance-chapters/5-waste/5-c-1-a-municipal/view (accessed Aug 2023).

[79] O.-K. Nielsen, M. Plejdrup, O. Rentz, D. Oertel, M. Woodfield, and R. Stewart. Combustion in energy and transformation industries. In *EMEP/EEA air pollutant emission inventory guidebook 2019*. European Environment Agency, 2019. Available at https://www.eea.europa.eu/publications/emep-eea-guidebook-2019/part-b-sectoral-guidance-chapters/1-energy/1-a-combustion/1-a-1-energy-industries/view (accessed Aug 2023).

[80] OGC. OGC GeoSPARQL - A Geographic Query Language for RDF Data, 2012. Available at https://www.ogc.org/standards/geosparql/ (accessed Dec 2023).

[81] OGC. CityGML, 2021. Available at https://www.ogc.org/standard/citygml/ (accessed Dec 2023).

[82] G. Oliveri Conti, B. Heibati, I. Kloog, M. Fiore, and M. Ferrante. A review of AirQ Models and their applications for forecasting the air pollution health outcomes. *Environmental Science and Pollution Research*, 24:6426–6445, 2017. doi:10.1007/s11356-016-8180-1.

[83] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

[84] E. O'Dwyer, I. Pan, R. Charlesworth, S. Butler, and N. Shah. Integration of an energy management tool and digital twin for coordination and control of multi-vector smart energy systems. *Sustainable Cities and Society*, 62:102412, Nov. 2020. doi:10.1016/j.scs.2020.102412.

[85] S. G. Perry, A. J. Cimorelli, R. J. Paine, R. W. Brode, J. C. Weil, A. Venkatram, R. B. Wilson, R. F. Lee, and W. D. Peters. AERMOD: A Dispersion Model for Industrial Source Applications. Part II: Model Performance against 17 Field Study Databases. *Journal of Applied Meteorology*, 44(5):694–708, 2005. doi:10.1175/jam2228.1.

[86] R. L. Petersen, S. A. Guerra, and A. S. Bova. Critical review of the building downwash algorithms in AERMOD. *Journal of the Air & Waste Management Association*, 67(8):826–835, 2017. doi:10.1080/10962247.2017.1279088.

[87] S. Pfenninger, A. Hawkes, and J. Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33: 74–86, May 2014. doi:10.1016/j.rser.2014.02.003.

[88] S. Pfenninger, J. DeCarolis, L. Hirth, S. Quoilin, and I. Staffell. The importance of open data and software: Is energy research lagging behind? *Energy Policy*, 101: 211–215, Feb. 2017. doi:10.1016/j.enpol.2016.11.046.

[89] S. Pfenninger, L. Hirth, I. Schlecht, E. Schmid, F. Wiese, T. Brown, C. Davis, M. Gidden, H. Heinrichs, C. Heuberger, S. Hilpert, U. Krien, C. Matke, A. Nebel, R. Morrison, B. Müller, G. Pleßmann, M. Reeg, J. C. Richstein, A. Shivakumar, I. Staffell, T. Tröndle, and C. Wingenbach. Opening the black box of energy modelling: Strategies and lessons learned. *Energy Strategy Reviews*, 19:63–71, Jan. 2018. doi:10.1016/j.esr.2017.12.002.

[90] H. Y. Quek, M. Hofmeister, S. D. Rihm, J. Yan, J. Lai, G. Brownbridge, M. Hillman, S. Mosbach, W. Ang, Y.-K. Tsai, D. N. Tran, S. Kang, W. Tan, and M. Kraft. BIM-GIS Integration: Knowledge graphs in a world of data silos, 2023. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/311/.

[91] H. Y. Quek, F. Sielker, J. Akroyd, A. N. Bhave, A. von Richthofen, P. Herthogs, C. v. d. L. 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. doi:10.1017/dap.2023.1.

[92] B. Rana, Y. Singh, and P. K. Singh. A systematic survey on internet of things: Energy efficiency and interoperability perspective. *Transactions on Emerging Telecommunications Technologies*, 32(8):e4166, Dec. 2021. doi:10.1002/ett.4166.

[93] rasdaman GmbH & Constructor University. rasdaman: raster data manager, 2023. Available at http://www.rasdaman.org/ (accessed Jan 2024).

[94] V. Reif, T. I. Strasser, J. Jimeno, M. Farre, O. Genest, A. Gyrard, M. McGranaghan, G. Lipari, J. Schütz, M. Uslar, S. Vogel, A. Bytyqi, R. Dornmair, A. Corusa, G. Roy, F. Ponci, A. Dognini, and A. Monti. Towards an interoperability roadmap for the energy transition. *e & i Elektrotechnik und Informationstechnik*, 140(5):478–487, July 2023. doi:10.1007/s00502-023-01144-2.

[95] C. Reinisch, M. J. Kofler, and W. Kastner. ThinkHome: A smart home as digital ecosystem. In *4th IEEE International Conference on Digital Ecosystems and Technologies*. IEEE, Apr. 2010.

[96] S. D. Rihm, M. K. Kovalev, A. A. Lapkin, J. W. Ager, and M. Kraft. On the role of C4 and C5 products in electrochemical CO2 reduction via copper-based catalysts. *Energy & Environmental Science*, 16(4):1697–1710, 2023. doi:10.1039/d2ee03752a.

[97] S. D. Rihm, Y. R. Tan, W. Ang, M. Hofmeister, X. Deng, M. T. Laksana, H. Y. Quek, J. Bai, L. Pascazio, S. C. Siong, J. Akroyd, S. Mosbach, and M. Kraft. The Digital Lab Manager: Automating Research Support, 2024. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/318/.

[98] H. Rijgersberg, D. Willems, and J. Top. Ontology of units of Measure, 2020. Available at http://www.ontology-of-units-of-measure.org/page/om-2 (accessed Dec 2023).

[99] T. Savage, J. Akroyd, S. Mosbach, M. Hillman, F. Sielker, and M. Kraft. Universal digital twin – the impact of heat pumps on social inequality. *Advances in Applied Energy*, 5:100079, 2022. doi:10.1016/j.adapen.2021.100079.

[100] T. Savage, J. Akroyd, S. Mosbach, N. Krdzavac, M. Hillman, and M. Kraft. Universal digital twin: Integration of national-scale energy systems and climate data. *Data-Centric Engineering*, 3, 2022. doi:10.1017/dce.2022.22.

[101] K. Seangkiatiyuth, V. Surapipith, K. Tantrakarnapa, and A. W. Lothongkum. Application of the AERMOD modeling system for environmental impact assessment of $NO_2$ emissions from a cement complex. *Journal of Environmental Sciences*, 23 (6):931–940, 2011. doi:10.1016/S1001-0742(10)60499-8.

[102] P. Sicard, E. Agathokleous, A. D. Marco, E. Paoletti, and V. Calatayud. Urban population exposure to air pollution in Europe over the last decades. *Environmental Sciences Europe*, 33(1), 2021. doi:10.1186/s12302-020-00450-2.

[103] R. Sosa, E. Vega, A. Wellens, M. Jaimes, G. Fuentes, E. Granados, A. L. Alarcón, M. del Carmen Torres, P. Sánchez, S. Rosas, and E. Mateos. Reduction of atmospheric emissions due to switching from fuel oil to natural gas at a power plant in a critical area in Central Mexico. *Journal of the Air & Waste Management Association*, 70(10):1043–1059, 2020. doi:10.1080/10962247.2020.1808113.

[104] S. J. Taylor and B. Letham. Forecasting at scale. *The American Statistician*, 72(1): 37–45, 2018. doi:10.1080/00031305.2017.1380080.

[105] A. Tzachor, S. Sabri, C. E. Richards, A. Rajabifard, and M. Acuto. Potential and limitations of digital twins to achieve the Sustainable Development Goals. *Nature Sustainability*, 5(10):822–829, July 2022. doi:10.1038/s41893-022-00923-7.

[106] UK Government. Information Exchange Standard 4, 2023. Available at `https://github.com/dstl/IES4` (accessed Dec 2023).

[107] United States Environmental Protection Agency. *AP-42, Fifth Edition Compilation of Air Pollutant Emissions Factors, Volume 1: Stationary Point and Area Sources*, chapter External Combustion Sources, Natural Gas Combustion. United States Government, 2019. Available at `https://www3.epa.gov/ttnchie1/ap42/ch01/` (accessed Aug 2023).

[108] U. ur Rehman, P. Faria, L. Gomes, and Z. Vale. Future of energy management systems in smart cities: A systematic literature review. *Sustainable Cities and Society*, 96:104720, 2023. doi:10.1016/j.scs.2023.104720.

[109] W3C. OWL Web Ontology Language Overview, 2012. Available at `https://www.w3.org/TR/owl-features/` (accessed Dec 2023).

[110] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, and et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1): 160018, Mar 2016. doi:10.1038/sdata.2016.18.

[111] World Health Organization. WHO global air quality guidelines, 2021. Available at https://apps.who.int/iris/bitstream/handle/10665/345329/9789240034228-eng.pdf (accessed Aug 2023).

[112] World Wide Web Consortium. Semantic Sensor Network Ontology, 2017. Available at https://www.w3.org/TR/vocab-ssn/ (accessed Dec 2023).

[113] J. Wu, J. Yan, H. Jia, N. Hatziargyriou, N. Djilali, and H. Sun. Integrated Energy Systems. *Applied Energy*, 167:155–157, 2016. doi:10.1016/j.apenergy.2016.02.075.

[114] N. Wu, F. Zhang, J. Wang, X. Wang, J. Wu, J. Huang, J. Tan, R. Jing, J. Lin, S. Xie, and Y. Zhao. A review on modelling methods, tools and service of integrated energy systems in China. *Progress in Energy*, 5(3):032003, 2023. doi:10.1088/2516-1083/acef9e.

[115] G. Xiao, D. Lanti, R. Kontchakov, S. Komla-Ebri, E. Güzel-Kalaycı, L. Ding, J. Corman, B. Cogrel, D. Calvanese, and E. Botoeva. The virtual knowledge graph system ontop. In J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, and L. Kagal, editors, *The Semantic Web – ISWC 2020*, pages 259–277. Springer International Publishing, 2020.

[116] X. Xu and H. Cai. Domain Ontology for Utility Infrastructure: Coupling the Semantics of CityGML Utility Network ADE and Domain Glossaries. *Journal of Computing in Civil Engineering*, 35(5):04021011, Sept. 2021. doi:10.1061/(asce)cp.1943-5487.0000977.

[117] X. Zhou, A. Eibeck, M. Q. Lim, N. Krdzavac, and M. Kraft. An agent composition framework for the J-Park Simulator – a knowledge graph for the process industry. *Computers & Chemical Engineering*, 130:106577, 2019. doi:10.1016/j.compchemeng.2019.106577.

[118] X. Zhou, S. Zhang, M. Agarwal, J. Akroyd, S. Mosbach, and M. Kraft. Marie and BERT - A Knowledge Graph Embedding Based Question Answering System for Chemistry. *ACS Omega*, 8(36):33039–33057, 2023. doi:10.1021/acsomega.3c05114.

[119] B. Zou, F. Benjamin Zhan, J. Gaines Wilson, and Y. Zeng. Performance of AERMOD at different time scales. *Simulation Modelling Practice and Theory*, 18(5):612–623, 2010. doi:10.1016/j.simpat.2010.01.005.