

# The National Digital Twin of the UK – a knowledge-graph approach

Jethro Akroyd<sup>1,2</sup>, Sebastian Mosbach<sup>1,2</sup>, Amit Bhawe<sup>3</sup>, Markus Kraft<sup>1,2,4</sup>

released: 18 Dec 2020

<sup>1</sup> Department of Chemical Engineering  
and Biotechnology  
University of Cambridge  
Philippa Fawcett Drive  
Cambridge, CB3 0AS  
United Kingdom

<sup>2</sup> CARES  
Cambridge Centre for Advanced  
Research and Education in Singapore  
1 Create Way  
CREATE Tower, #05-05  
Singapore, 138602

<sup>3</sup> CMCL Innovations  
Sheraton House  
Cambridge  
CB3 0AX  
United Kingdom

<sup>4</sup> School of Chemical  
and Biomedical Engineering  
Nanyang Technological University  
62 Nanyang Drive  
Singapore, 637459

Preprint No. 264



---

*Keywords:* Dynamic knowledge graph, digital twin, data, agents, interoperability, distributed architecture

**Edited by**

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

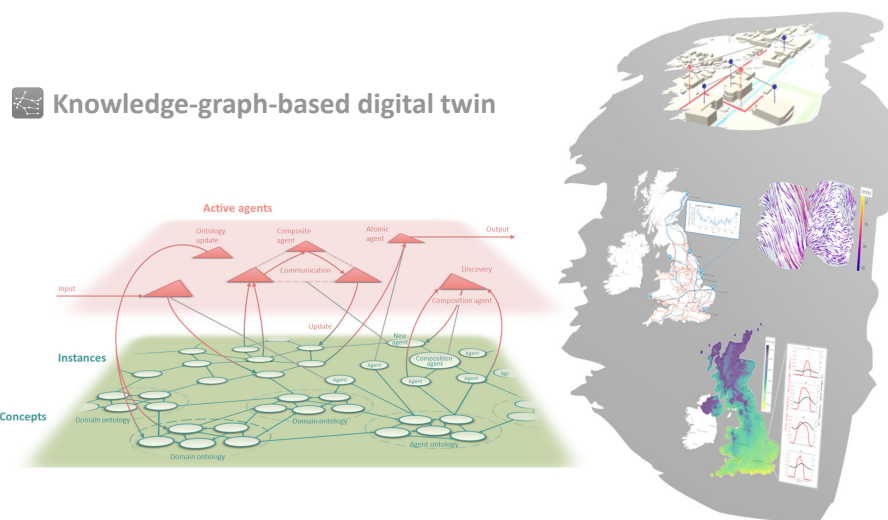
**E-Mail:** [mk306@cam.ac.uk](mailto:mk306@cam.ac.uk)

**World Wide Web:** <https://como.ceb.cam.ac.uk/>



## Abstract

This paper introduces a dynamic knowledge-graph approach for digital twins and demonstrates how it fulfils many of the technical requirements of the vision to create a National Digital Twin of the UK. The dynamic knowledge graph is implemented using technologies from the Semantic Web. It is composed of concepts and instances that are defined using ontologies, and of computational agents that operate on both the concepts and instances to update the dynamic knowledge graph. By construction, it is distributed, supports cross-domain interoperability, and ensures that data is connected, portable, discoverable and queryable via a uniform interface. The knowledge graph includes the notions of a ‘base world’ that describes the real world and that is maintained by agents that incorporate real-time data, and of ‘parallel worlds’ that support the intelligent exploration of alternative designs without affecting the base world. Use cases are presented that demonstrate the ability of the dynamic knowledge graph to host geospatial and chemical data, control chemistry experiments, perform cross-domain simulations and perform scenario analysis. The questions of how to make intelligent suggestions for alternative scenarios and how to ensure alignment between the scenarios considered by the knowledge graph and the goals of society are considered. Work to extend the dynamic knowledge graph to develop a digital twin of the UK to support the decarbonisation of the energy system is discussed. Important directions for future research are highlighted.



## Highlights

- Dynamic knowledge-graph approach proposed for National Digital Twin.
- Enables distributed architecture and supports cross-domain interoperability.
- Data is unambiguous, discoverable and accessible via uniform query interface.
- Base world maintained by computational agents acting on knowledge graph.
- Parallel worlds enable exploration of complex design spaces.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The Semantic Web stack</b>	<b>5</b>
2.1	Ontologies . . . . .	5
2.2	Linked Data . . . . .	5
2.3	Knowledge graphs . . . . .	6
2.4	Data storage . . . . .	6
2.5	Queries and updates . . . . .	6
2.6	Inference and rule engines . . . . .	7
2.7	Examples . . . . .	7
<b>3</b>	<b>The World Avatar dynamic knowledge graph</b>	<b>7</b>
3.1	Structure . . . . .	8
3.2	Agents . . . . .	9
3.3	Ontological coverage . . . . .	10
3.4	Use cases . . . . .	10
3.4.1	Urban planning . . . . .	10
3.4.2	Intelligent querying and generation of data . . . . .	11
3.4.3	Cross-domain calculation of air quality . . . . .	12
3.4.4	Parallel worlds for scenario analysis . . . . .	14
3.5	The knowledge-graph-based digital twin of the UK . . . . .	16
3.5.1	Base world . . . . .	16
3.5.2	Goal alignment for parallel worlds . . . . .	21
<b>4</b>	<b>Conclusions</b>	<b>22</b>
	<b>References</b>	<b>26</b>

# 1 Introduction

In the 2016 Paris Agreement, 196 nations committed to reduce their greenhouse gas emissions in a legally binding treaty on climate change. The obligations under the agreement are projected to require significant changes in infrastructure, ranging from a move to large-scale energy distribution [35, 98] to the adoption of distributed energy resources [71] and the development of intelligent infrastructure [112].

The challenges posed by climate change are inherently interdisciplinary. The development of solutions requires the consideration of economic, engineering, environmental, and social factors [44, 77, 96] over a range of geographic scales [104]. It may also require the development of new concepts, for example to facilitate the consideration of societal practices and patterns of consumption in the analysis of different solutions [95]. These factors are strongly interconnected and it is widely appreciated that interoperable [37], collaborative [45] models that span multiple disciplines [71] are necessary, perhaps in the form of digital twins supported by artificial intelligence (AI) [69].

This has been recognised in the UK [20], which in 2018 launched a National Digital Twin programme [38]. The vision of the programme is to lead the digital transformation of the construction and infrastructure sectors by creating a National Digital Twin that enables a data-centric approach to managing the built and natural environments. The National Digital Twin is not intended to be monolithic, but rather a system of interconnected entities that share data and that are able to combine to answer complex questions. For example, what infrastructure exists, and what is its capacity, location, condition and value? What are the environmental, social and economic impacts of proposed infrastructure projects?

The benefit of a data-centric approach, within which we include the availability of data from smart infrastructure and the sharing of data in a National Digital Twin, lies in the use of data to support better decision making. In mature economies such as the UK, the value of existing infrastructure far exceeds the value of infrastructure that is under development, such that the key benefits will derive from optimising the use of existing assets [19]. Examples from the UK include aviation and rail, where new capacity has been (or will be) created by optimising the air traffic and rail control systems [19].

In 2018, the Centre for Digital Built Britain (CDBB) published the Gemini Principles [31] to establish a basis for the development of the National Digital Twin. The principles state that each element of the digital twin must have a clear purpose, must be trustworthy and must function effectively. They raise challenging questions about the ownership, curation, quality and security of data, about the protection of intellectual property, and about the inter-operability of data and models, but intentionally avoid prescribing solutions.

CDBB subsequently published a plan to develop an Information Management Framework (IMF) for the National Digital Twin [68]. It was proposed that the IMF should include a Foundation Data Model to provide a consistent high-level ontology for the digital twin and for sharing and validating data, a Reference Data Library to provide a shared vocabulary for describing digital twins, and an Integration Architecture to support the design and connection of digital twins. It is implicit in the discussion that data should be stored using a distributed architecture as opposed to a centralised database, and that it must be easy to discover and query across multiple sources. The discussion further emphasises

the need to enable interactions between data and models from different organisations and sectors, including buildings, utilities, infrastructure and transport, as well as social and environmental models.

Despite the bold statement of intent by the National Digital Twin programme, it remains unclear how to implement such a digital twin. In order to achieve its full potential, it will be necessary, for example, to address the challenges to interoperability posed by the heterogeneity of data and services (software and tools), and the non-uniqueness (*i.e.* duplication and inconsistency) of data, all of which impede communication between domains.

Starting in 2013, researchers at CARES [2], the Cambridge Centre for Advanced Research and Education in Singapore, started work on the World Avatar project to investigate this type of problem. The idea behind the project is to connect data and computational agents to create a living digital ‘avatar’ of the real world. The digital world is intended to be ‘living’ in the sense that the computational agents will, amongst other things, incorporate real-time data such that the digital world remains up to date and that analyses based on the data remain self-consistent. The name ‘World Avatar’ seeks to convey the possibility of representing every aspect of the real world, extending the idea of digital twins to consider the possibility of representing any abstract concept or process.

The initial focus of the research at CARES was the decarbonisation of the chemical industry in Singapore. This resulted in the development of the J-Park Simulator [15, 52], which uses technologies from the Semantic Web [30] to create a knowledge-graph-based digital twin of the eco-industrial park on Jurong Island [73, 92, 93, 107]. The scope of World Avatar project has since broadened considerably. It has demonstrated the use of ontologies to facilitate interoperability between models and data from different technical domains [51, 60], developed an agent discovery and composition service to allow agents to be combined to perform more complex tasks [109], developed a blockchain-based smart contract system to enable an agent marketplace [110], and demonstrated a parallel world framework for scenario analysis [53]. It also now includes a Cities Knowledge Graph project with researchers at ETH Zurich to improve the planning and design of cities, and a Pharma Innovation Programme Singapore (PIPS) project to allow the World Avatar to control laboratory robots performing chemistry experiments.

The **purpose of this paper** is to demonstrate how Semantic Web technology can be used to implement a National Digital Twin as part of the World Avatar project. The paper is structured as follows. Section 2 discusses the Semantic Web and the technologies that are available as part of it. Section 3 explains the World Avatar and presents examples that demonstrate its capability, and discusses current work to develop a knowledge-graph-based digital twin of the UK to support the decarbonisation of the UK energy landscape. Conclusions are discussed in section 4.

## 2 The Semantic Web stack

The Semantic Web [30] provides a valid description logic [27] representation of data on the World Wide Web. Resources on the Semantic Web are all instances of ontological classes. The resources are identified using internationalised resource identifiers (IRIs), ensuring an unambiguous representation. Technologies that enable the architecture, development and existence of the Semantic Web are vertically orchestrated based on functional dependency, forming the Semantic Web stack. The interested reader is referred to Allemang and Hendler [25].

### 2.1 Ontologies

An ontology is a collection of classes, object properties and data properties expressing facts about and a semantic model of a domain of interest, for example, the built environment. An object property links an instance of a class (the domain of the property) to an instance of a class (the range of the property).<sup>1</sup> Object properties may be structured hierarchically. A data property links an instance of a class (the domain of the property) to a data element (the range of the property).<sup>2</sup> Similar to object properties, data properties may be structured hierarchically. A class acquires meaning via relationships to other classes.

An ontology is also known as a TBox (Terminological Component), where a class hierarchy is formed and associative relations are established through user-defined object properties, allowing cross-domain linking and modelling. Data properties allow the representation of different types of data. An ABox (Assertional Component) is an instantiation of a TBox representing entities, together with data and metadata about the entities. Inconsistencies can be determined at the level of data and relationships between the entities in an ABox via the assertion of data types, logical constraints, and cardinality (*i.e.* how many of something are allowed) in the TBox.

### 2.2 Linked Data

Linked Data [29] is Semantic Web data that is linked to other Semantic Web data to facilitate exploration and discovery (as opposed to semantic, but otherwise isolated data). Linked Data uses RDF (Resource Description Framework) [74] as the standard language to store data in the form of *subject*, *predicate*, *object* triples. Related entities across datasets are connected via IRIs which are meant to be resolvable and provide meaningful information when looked up. Linked Data use ontologies to create and link entities and enrich the entities with data and metadata. Linked Open Data (LOD) refers to Linked Data that is published with an open license, such that it offers the possibility of finding and using related data, for example to resolve issues with naming and consistency.

---

<sup>1</sup> *e.g.* <person> (the domain) <bornIn> (the object property) <place> (the range).

<sup>2</sup> *e.g.* <person> (the domain) <hasName> (the data property) <string> (the range).

## 2.3 Knowledge graphs

A knowledge graph is a network of data expressed as a directed graph, where the nodes of the graph are concepts or their instances (data items) and the edges of the graph are links between related concepts or instances. Knowledge graphs are often built using the principles of Linked Data. They provide a powerful means to host, query and traverse data, and to find and retrieve related information.

## 2.4 Data storage

Triple stores and quad stores provide solutions for the storage and management of RDF data. As the names suggest, triple stores use a 3-column (subject, predicate, object) statement format, whereas quad stores use a 4-column (subject, predicate, object, context) statement format. The context element can provide additional information about the data, for example a time coordinate, a reference or a name. When the context contains a name (which could be codified by an IRI) the statement is called a named graph. Many statements can belong to the same named graph, forming a subunit within the store and allowing the possibility of querying only the named graph. Most stores are able to accommodate both triples and quads and, for historic reasons, are often simply referred to as ‘triple stores’. A number of storage solutions exist. For example, RDF4J [13] and Fuseki [8] at the lab scale, and Blazegraph [1] and Virtuoso [17] at the enterprise scale. In addition to triple stores, Ontology-Based Data Access (OBDA) enables ontological queries to be performed over relational and NoSQL databases [33]. Triple stores and OBDA provide endpoints, known as SPARQL endpoints, that are identified by IRIs and that can be used to query and update data. Authentication and authorisation solutions to control security and access exist and are discussed elsewhere [102].

## 2.5 Queries and updates

SPARQL [101] is a semantic-aware standard language to query and/or update data via a SPARQL endpoint, with or without activating an inference engine. Queries can be performed over named graphs, over an entire endpoint or (using a federated query) over multiple endpoints. Federated queries offer the possibility to retrieve data from and reason over multiple data sources at the same time, including the possibility of combining data from public and private data sources owned by any number of different parties. GeoSPARQL [87] is a standard defined by the Open Geospatial Consortium (OGC) as a domain-specific extension of SPARQL, allowing spatial reasoning and computations on data to reply to geospatial queries.

Each triple or quad store provides its own Application Programming Interface (API) to support SPARQL query and update operations. However, each API is only compatible with its native store. Jena-JDBC [11] provides a highly scalable store-agnostic API that can be used to query and update virtually any store.



## 2.6 Inference and rule engines

Inference engines (*e.g.* HermiT [10]) can operate on data expressed using RDF to reason about existing knowledge and infer new knowledge [25], for example using the domain and range of object and data properties. Rule engines (*e.g.* Drools [7]) can generate new knowledge through the codification and analysis of ‘if-then-else’ conditions to calculate new values for data properties.

## 2.7 Examples

Google, Microsoft, IBM, Facebook and eBay recently published a paper [86] to explain how they use knowledge graphs. Google and Microsoft both use knowledge graphs to power their search engines. IBM uses one to support the IBM Watson question and answer engine. Facebook uses one to help understand what people care about. eBay is building one to connect what a seller is offering with what a buyer wants.

The article provides a number of insights and identifies opportunities and challenges, particularly relating to the scalability and query-response time of triple stores. Google built its knowledge graph using a set of low-level structures and associated reasoning mechanisms that are replicated at increasing levels of abstraction. This gives advantages in terms of scalability because new abstractions are built on tried and tested low-level units, and because it makes the design easier to understand. Facebook addresses the challenge of conflicting information by either dropping information if confidence in it was deemed to be too low, or by retaining the conflicting information and the associated provenance data. In this manner, their knowledge graph is able to explain why certain assertions are made, even if it cannot identify which assertion is objectively ‘correct’. IBM Watson uses polymorphic stores to offer the ability to support multiple indices, database structures, in-memory and graph stores. Data is split (often redundantly) into one or more of these stores, allowing each store to address specific requirements and workloads.

## 3 The World Avatar dynamic knowledge graph

The World Avatar project aims to create a digital ‘avatar’ of the real world. The digital world is composed of a dynamic knowledge graph that contains concepts and data that describe the world, and an ecosystem of autonomous computational agents that simulate the behaviour of the world and that update the concepts and data so that the digital world remains current in time. In this manner, it is proposed that it is possible to create digital twins that are able to describe the behavior of complex systems, and by doing so provide the ability to make data-driven decisions about how to optimise the systems.

Within the World Avatar, there exists the notions of a ‘*base world*’ that describes the current state of the real world, and ‘*parallel worlds*’ that describe the effect of what-if changes to the base world. The base world enables the World Avatar to address control problems. The parallel worlds enable the World Avatar to perform scenario analysis to address design problems.

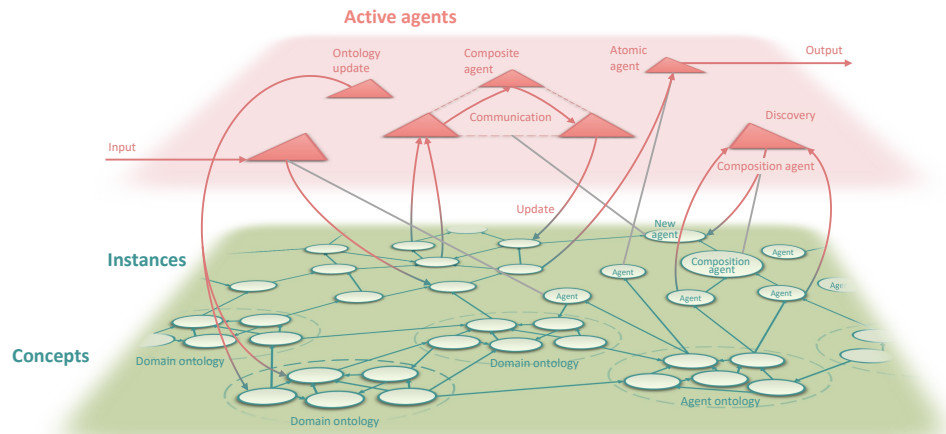
A central tenet of the World Avatar is the requirement to share data so that all the computational agents experience a self-consistent view of the world, and the requirement that data and computational agents must be interoperable across different technical and social domains. This interoperability, defined in the sense of the ability of tools and systems to understand and use the functionalities of other tools [40], is considered to be one of the key aspects of Industry 4.0 [78] and will become increasingly important as the complexity and choice of tools inevitably increases. For example, considerations raised by recent reports concerning the possible role of hydrogen in the future UK energy landscape [21–24] span the electrical power system, the gas grid, the potential of biomass, solar and wind energy, the transport system, heating of commercial and domestic buildings, in addition to considerations relating to the acceptance and adoption of new technologies by the public.

The following sections present an overview of the World Avatar and give examples of how it has been used to date. Section 3.1 describes the structure of the World Avatar. Section 3.2 discusses the computational agents that operate within the World Avatar. Section 3.4 presents example use cases. Section 3.5 describes the extension of the World Avatar towards a knowledge-graph-based digital twin of the UK.

### 3.1 Structure

The World Avatar [15, 52] represents information in a dynamic knowledge graph using technologies from the Semantic Web stack. Unlike a traditional database, the World Avatar contains an ecosystem of autonomous computational agents that continuously update it.

The main components of the dynamic knowledge graph are illustrated in Fig. 1. The dynamic knowledge graph contains concepts and instances that are linked to form a directed graph. The agents are described ontologically [109] as part of the knowledge graph, and are able to perform actions on both concepts and instances. This confers versatility because it combines a design that enables agents to update and restructure the knowledge graph, with a structure that allows agents to discover and compose other agents simply by reading from and writing to the knowledge graph.



**Figure 1:** *The main components of the World Avatar dynamic knowledge graph.*

The links between the contents of the dynamic knowledge graph take the form of Internationalised Resource Identifiers (IRIs), which can be thought of as generalised web-addresses [103]. This ensures an unambiguous representation and enables the efficient traversal of the graph, both in the sense of traversing related content and in the sense of traversing whatever hardware is used to host the content.

## 3.2 Agents

At any given time, instances of the computational agents described in the knowledge graph may be activated to perform tasks. In this context, ‘agent’ refers to software that uses Semantic Web technologies to operate on the knowledge graph to fulfil specific objectives. Agents may be *atomic* agents (*i.e.* not composed of other agents), or *composite* agents composed of multiple atomic agents operating in a coordinated way.

Agents can perform a variety of activities and can operate on both the concepts and instances in the knowledge graph [52]:

- **Input/output.** Agents can update **instances** with data from the real world, for example data from sensors or smart infrastructure. Agents can retrieve data from the knowledge graph and send signals back to real world, for example to visualise data or to control an actuator.
- **Update.** Agents can query the knowledge graph, calculate derived quantities and update **instances** with the new data. An example could be querying sources of emissions and the prevailing wind conditions to update estimates of air quality by simulating the dispersion of the emissions [52, 109].
- **Restructure.** Agents can restructure the knowledge graph by:
  - Adding **instances** (*i.e.* items in ABoxes), for example, to explore the optimal placement of new infrastructure [50, 53].
  - Adding **concepts** (*i.e.* items in TBoxes) to improve the description of data, for example, to add classes relating to *dispatchable* and *non-dispatchable* resources to an ontology describing energy sources.
  - Adding **relationships** between **instances** or between **concepts**, for example, using ontology matching to support the creation of relationships between equivalent concepts.
- **Create new agents.** Agents can provide services to facilitate agent discovery and to create composite agents to coordinate the activities of agents to perform complex tasks, for example, the estimation of air quality [109]. The possibility of introducing an agent marketplace and quality of service (QoS) measures for agents has been demonstrated using a blockchain-based smart contract system [110].

The calculations performed by agents can take any form, including physics-based models with a theoretical structure, grey box models that combine some theoretical structure with data-driven components, and pure data-driven models. See [105] for examples of

these approaches in the context of emissions modelling. The availability of semantically-structured machine-queryable data makes the World Avatar well-suited to be coupled to machine learning applications, where it well known that the time spent on data preparation and curation typically adds a significant cost to machine learning.

The potential of the World Avatar arises from the generic functionalities offered by agents that integrate ontologies and create new agents. Both would serve to raise the level of semantic interoperability and allow for the automation of processes to include new data and capabilities, and maintain interoperability as the quantity and complexity of the data and agents increases.

### 3.3 Ontological coverage

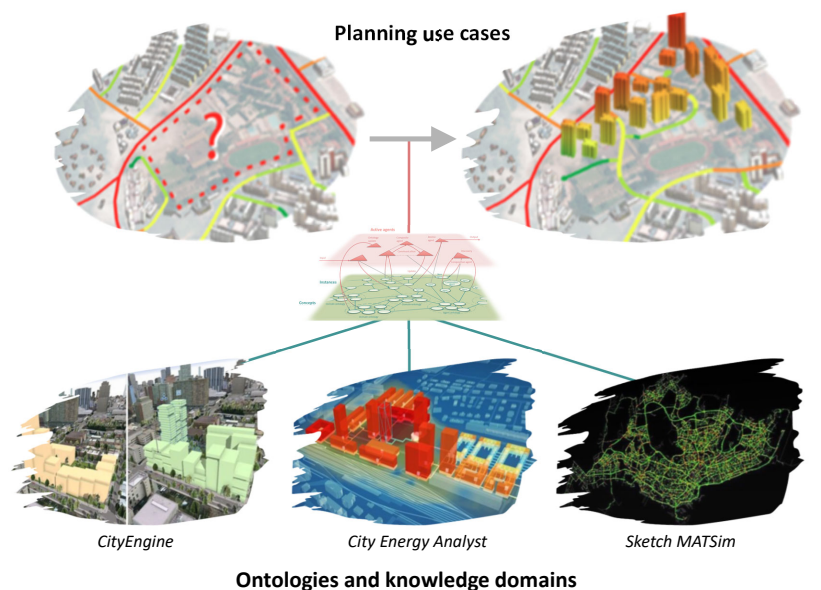
The current ontological coverage of the World Avatar includes 3D city data [OntoCityGML, 52], weather [Weather Ontology, 70], process engineering [OntoCAPE, 80], eco-industrial parks [OntoEIP, 106–108] and electrical power systems [OntoPowerSys, 51]. It includes ontologies for quantum chemistry [OntoCompChem, 75], chemical species [OntoSpecies, 59], chemical kinetic reaction mechanisms [OntoKin, 58] and combustion experiments [OntoChemExp 28]. Selected subgraphs of the Linked Open Data (LOD) Cloud [12], in particular DBpedia [6, 76] are also connected to the knowledge graph.

### 3.4 Use cases

This section presents use cases that demonstrate different aspects of the World Avatar. The use cases are from the Cities Knowledge Graph, PIPS and J-Park Simulator projects so are mainly associated with Singapore. However, the application of the World Avatar is not restricted to these contexts, as we shall discuss in Section 3.5.

#### 3.4.1 Urban planning

The Cities Knowledge Graph project [3] will integrate urban design and planning tools developed by the Future Cities Laboratory at the Singapore-ETH Centre with the World Avatar. The tools include the Collaborative Interactive Visualisation and Analysis Laboratory design informatics platform [63], the Multi-Agent Transport Simulation Singapore model [64], the City Energy Analyst toolbox [62] and the ESRI CityEngine [57] 3D modelling software. The concept is shown in **Fig. 2**.



**Figure 2:** *The Cities Knowledge Graph project will develop a pilot for a comprehensive knowledge management platform that provides interoperability between different types of city-relevant data to improve the precision of planning instruments and bridge the gap between planning use cases and knowledge domains.*

Reproduced from <https://www.cares.cam.ac.uk/research/cities/> with permission.

The resulting system will enable interoperability between building data, transport flows, underground infrastructure, humidity and temperature data to improve the design and planning of cities by automating aspects of data processing, and by integrating concepts and targets from different planning arenas. Work to date has demonstrated the capability to incorporate live building management system (BMS) data into the dynamic knowledge graph [16] and has integrated the World Avatar with OntoCityGML [39], an ontology that describes 3D models of cities and landscapes based on the CityGML 2.0 standard [65]. Tools have been developed to convert data from CityGML to OntoCityGML format. Work to find the best way to host this data within the dynamic knowledge graph is ongoing.

### 3.4.2 Intelligent querying and generation of data

The World Avatar contains chemical data, including kinetic reaction mechanisms [58] and computational chemistry [75]. It is able to use computational chemistry data to calculate thermodynamic quantities required by the reaction mechanisms, and link the resulting quantities to the computational chemistry calculations used to derive them [59]. It is able to identify inconsistencies in the data [60] and interact with high performance computing (HPC) facilities in the real world to perform additional computational chemistry calculations to generate the data required to resolve the problems [81]. It includes experimental data (for combustion experiments) and is able to automate the process of calibrating reaction mechanisms versus experiment data [28].

Research in collaboration with CMCL [5] has investigated the use of natural language processing (NLP) to develop an intelligent query interface for the chemistry data available via the World Avatar [111]. **Figure 3** shows a screenshot of a demonstration website [43]. It has been shown that the use of intent recognition and topic modelling, and the structure of the domain ontologies has a strong impact on the accuracy of the queries, with shallow ontologies performing better than deep ontologies. This poses a number of open challenges because scientific research typically leads to hierarchical classification systems that do not lend themselves to shallow ontologies.

The availability of intelligent query interfaces has the potential to improve the discovery and interoperability of data. Work is ongoing to use such interfaces in conjunction with the World Avatar to develop quantitative structure activity/property relationships (QSAR/QSPR) that allow the prediction of the properties of a chemical from its structure, for example its toxicity [91] or its photovoltaic performance [54, and references therein]. This is the focus of the PIPS [4] project. PIPS will develop the capability of the World Avatar not only to automate calculations [81], but to automate *experiments* by allowing agents to control a laboratory robot. This capability will close the loop between analysis and experiment to automate the generation of reaction and synthesis data, and establish an ontology framework for AI reasoning agents to generate chemical knowledge from data.

The screenshot shows the Marie website interface. At the top, there is a header with a portrait of Marie Curie, the name 'Marie', and navigation links 'About', 'Partners', and 'Contact'. Below the header is a search bar containing the text 'Molecular model of aromatic hydrocarbons'. The search results are displayed under the heading 'Results (from the World Avatar)'. There are five results listed, each with a number, a name, and a molecular model:

Results (from the World Avatar)
1 naphthalene
2 naphthalene
3 corannulene
4 coronene
5 phenanthrene

**Figure 3:** Screenshot of Marie website (with enlarged text).

### 3.4.3 Cross-domain calculation of air quality

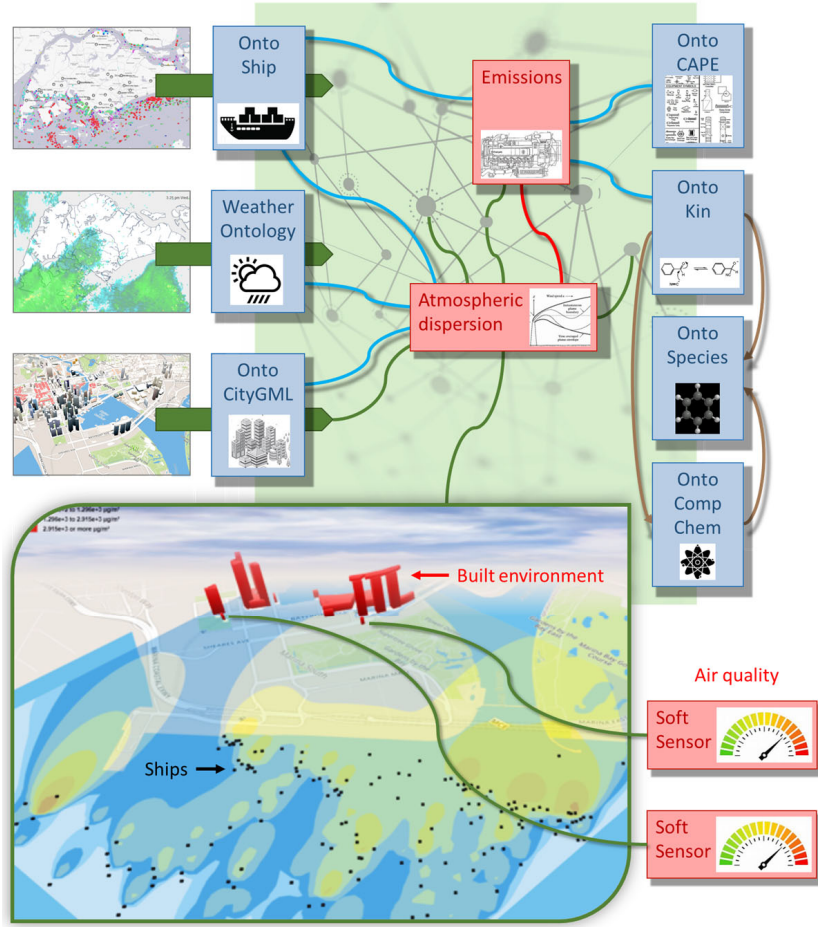
In recent years there has been increasing concern over the effect of emissions from combustion on air quality [47, 82], in particular, the emission of  $\text{NO}_x$  [48, 49, 56, 66] and particulate matter (PM) [55], and the effect of black carbon (BC) emissions on the climate [32].

Combustion accounts for approximately 99% of Singapore’s  $\text{CO}_2$  emissions [83, pages 134–137]. In addition to being one of the most densely populated countries [97], it hosts one of the world’s busiest ports with arrivals exceeding 138,000 vessels<sup>3</sup> and a throughput of more than 37 million twenty-foot equivalent units in 2019 [79]. There naturally arises the question of how emissions from hard-to-abate sectors like shipping influence factors such as air quality in different regions of Singapore. The interoperability of the World

<sup>3</sup> Includes only vessels with gross tonnage exceeding 75 GT.



Avatar provides a method to answer these questions. **Figure 4** illustrates how this was achieved. See references [52] and [60] for more details of the implementation and results.



**Figure 4:** Real-time cross-domain estimation of the contribution of emissions from shipping to air quality in Singapore. Agents operating on the World Avatar update the knowledge graph with real-time information about the weather and about ships in the vicinity of Singapore. An emissions agent is able to use the information about the ships to estimate the emissions of unburned hydrocarbons, CO, NO<sub>2</sub>, NO<sub>x</sub>, O<sub>3</sub>, SO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> from each ship. An atmospheric dispersion agent is able to use the information about the weather, the emissions from each ship and the built environment in Singapore to simulate the dispersion of the emissions. Virtual sensor agents report the resulting air quality estimates at different locations. Adapted from [60].

The choice of which agents to use in the calculation illustrated in Fig. 4 is controlled by a composition agent [109]. The emissions can be calculated either by agents that use surrogate models developed using the SRM Engine Suite [42] and Model Development Suite [41], or by agents that directly use the SRM Engine Suite. The atmospheric dispersion can be simulated either by agents that use the Atmospheric Dispersion Modelling System (ADMS) [36] or by agents that use the EPISODE 3D CFD code [72]. The agents

wrap around existing software, including commercial software, with the possibility of an agent marketplace [110] offering an alternative to traditional licensing models.

Whilst the emissions from ships can be described using data-driven models, trained using measurement data for example, this would preclude the ability to assess the impact of different fuels. The chemical data (species properties, reactions and reaction rates) that are required to assess the impact of different fuels can be described in the knowledge graph and any unknown quantities can, in principle, be calculated (although the calculation of reaction rates remains challenging). The ability of the World Avatar to enable interoperability between the different components of such calculations has been demonstrated in terms of calculating thermodynamic data for chemical species from computational chemistry data [81], linking the resulting species data with reaction mechanisms [59], and performing parameter estimation to improve the agreement with experimental data [28]. This is depicted on the right-hand side of Fig. 4.

This use case illustrates how the World Avatar may be used to calculate the consequences of current activities in the base world. It demonstrates what can be achieved in terms of interoperability, both between models and data from different domains (weather, the built environment, ships, chemistry, emissions, atmospheric dispersion) and covering different length scales (from the atomic length scales of the computational chemistry calculations to the kilometre length scales of the atmospheric dispersion simulations). Of course, this is unnecessarily elaborate if we are only interested in Singapore. The point, however, is that the World Avatar allows such calculations to be performed for any location.

#### 3.4.4 Parallel worlds for scenario analysis

The parallel worlds capability of the World Avatar supports decision making in a complex environment by allowing the exploration of different scenarios and their associated outcomes. The parallel worlds exploit the structure of the dynamic knowledge graph to use scenario agents to group new instances of any entities that are modified in a scenario and store them in a scenario-specific part of the knowledge graph. Access, queries and updates to the scenario-specific instances are mediated via a scenario agent. The scenario-specific information in the parallel world can be thought of as overlaying the base world, so unchanged entities remain connected to the base world and any changes to the base world are reflected in the scenario. Conversely, any changes in the parallel world remain isolated within the scenario-specific part of the knowledge graph and so do not interfere with the base world.

**Figure 5** illustrates the parallel world concept. In this example, the parallel world is used to explore the effect of levying a carbon tax on the emissions from power generation processes to motivate a transition from fossil fuels to clean energy technologies. The following questions are addressed:

- What value of carbon tax is required to make the transition to small modular nuclear reactors (SMRs) profitable for a given set of conditions (*e.g.* project life span, depreciation rate, electrical load profiles and generator characteristics)?
- Which plant(s) should be replaced and where should the new SMR(s) be located?





**Figure 5:** *Parallel world concept for what-if scenario analysis. The panels at the top illustrate the electrical network from the real world (left) and an optimised network that is subject to a carbon tax (right). The network from the real world is described in the base world. The modifications to the network in the parallel world are described in a scenario-specific part of the knowledge graph. The pink triangles denote natural gas generators that are present in both the base world and the parallel world. The blue square denotes an oil generator that is only present in the base world. The radiation symbol denotes a small modular nuclear reactor that is only present in the parallel world. Adapted from [53].*

The parallel world shows that oil generators are replaced with SMRs as the carbon tax is increased from S\$5 to S\$170. The types of generator present and the corresponding estimates of the CO<sub>2</sub> emissions are automatically updated in the parallel world to reflect these changes, and an optimal power flow (OPF) agent is invoked to minimise the overall operating cost of the scenario in the parallel world. See references [53] and [50] for more details of the implementation and results.

This use case illustrates how digital twinning and what-if scenario analysis using a dynamic knowledge graph can support decision makers to understand the effect of different design choices and policy instruments. In this example, the problem addressed by the parallel world is simple enough to be formulated as a classical optimisation problem with a well-defined objective function. However, many scenarios will be too complex for this to be the case. How to address this is discussed in the context of the knowledge-graph-based digital twin of the UK in Section 3.5.

### 3.5 The knowledge-graph-based digital twin of the UK

The knowledge-graph-based digital twin of the UK is an effort within the World Avatar project to create a semantic digital twin of the UK. The aim is to provide a comprehensive live distributed platform to support the optimal use, planning and development of infrastructure in the UK and to support scenario analysis to help the UK achieve its development goals, for example by supporting the decarbonisation of the energy landscape.

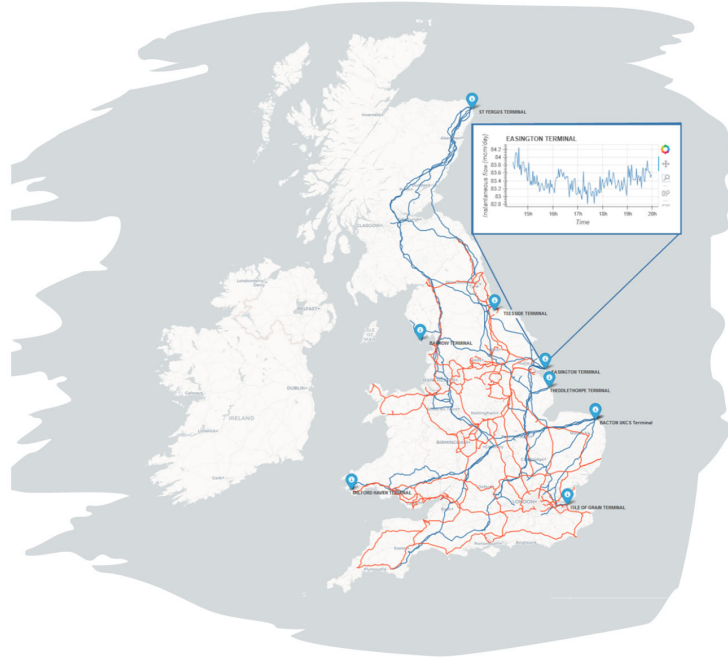
By construction, the design of the World Avatar ensures that all data is connected and has an unambiguous representation, and is discoverable and queryable through a uniform interface. The use of ontologies promotes portability, making it easy to reuse data and results and support interoperability. It is inherently multi-domain and can be extended by adding new ontologies and defining equivalence, or other relationships as appropriate, between semantically related terms to maintain, and even improve, interoperability as it grows. It is designed to allow the data to evolve over time, with the agents ensuring that it remains up to date. The architecture is distributed, offering the possibility of hosting different types of data in different places and providing links to the requisite provenance and licence information. This includes the option for data owners to retain control over the hosting of their data and to control access as required by the nature of the data, ranging from heavily restricted access for sensitive data to open access for open data.

The following sections describe progress towards the creation of the base world in a knowledge-graph-based digital twin of the UK, and discuss a route to meet the challenges posed by goal alignment in the context of parallel worlds.

#### 3.5.1 Base world

This section describes the progress to date and the immediate next steps for the creation of the base world in a knowledge-graph-based digital twin of the UK. The immediate motivation for doing this is to help address questions relating to how best to help the UK achieve net zero carbon emissions. The energy sector is the largest contributor to emissions in the UK and was responsible for 82% of total emissions in 2018 [34, page 40]. Work to date has therefore focused on data relating to the electric power system, the gas grid, the potential for photovoltaic and wind power, biomass and the built environment, all of which are critical to the future of the energy landscape [21].

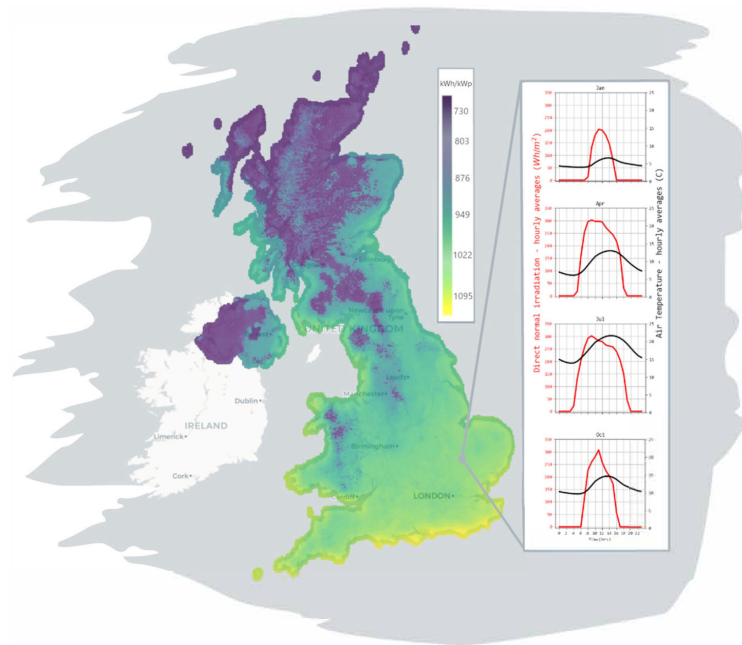
**Figure 6** shows the geospatial configuration of the National Grid Electricity Transmission and National Grid Gas Transmission systems [84, 85]. Work to extend the ontological coverage of the World Avatar to include this data is ongoing. In addition to the configuration of the networks, data for all the regional generators over 30 MW in the UK [46] have been added to the World Avatar and used in a preliminary analysis of how a carbon tax on existing generation infrastructure would effect the composition of power generation [26]. It was shown that the carbon tax motivated a transition from coal to combined cycle gas turbines (CCGT), with nuclear and renewable sources being fully utilised. When regional capacities, demands and (found to be complementary) transmission losses were considered, this transition was found to occur first in London and the South East, where northern coal imports were displaced by local CCGT generation.



**Figure 6:** *Geospatial configuration of the National Grid Electricity Transmission (red) and National Grid Gas Transmission (blue) systems. The markers (blue) show the locations of the intake terminals for the National Grid Gas Transmission system. The inset shows real-time data for the gas flow into the network from the Easington North Sea gas terminal.*

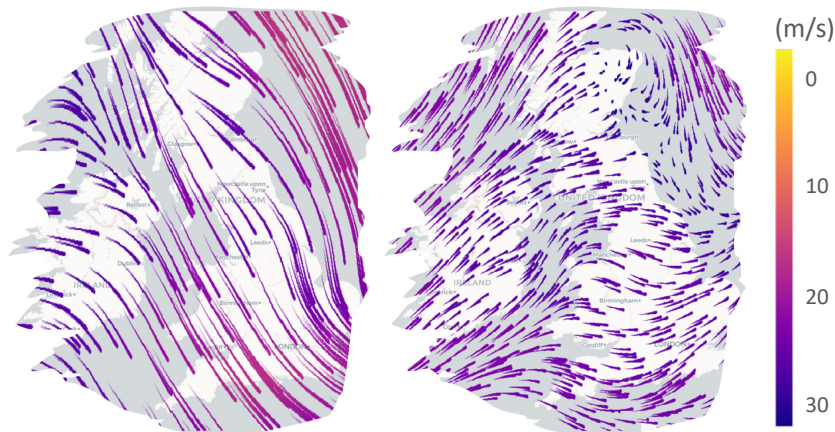
*Data obtained from the National Grid [84, 85].*

**Figures 7 and 8** shows the photovoltaic power potential and wind speed and direction data in the UK. The photovoltaic data are published by SolarGIS [14] on behalf of the World Bank as part of the Global Solar Atlas [9]. The wind data are part of the ERA5 data set published by the EU Climate Data Service [67]. Work to extend the ontological coverage of the World Avatar to include this type of data is an immediate priority.



**Figure 7:** Map showing the average annual photovoltaic power potential in the UK from 1994–2018. The inset shows the seasonal variation in the average daily photovoltaic potential and temperature in the vicinity of Cambridge, UK.

Contains data licensed by The World Bank under the Creative Commons Attribution license (CC BY 4.0) with the mandatory and binding addition presented in Global Solar Atlas terms (<https://globalsolaratlas.info/support/terms-of-use>).



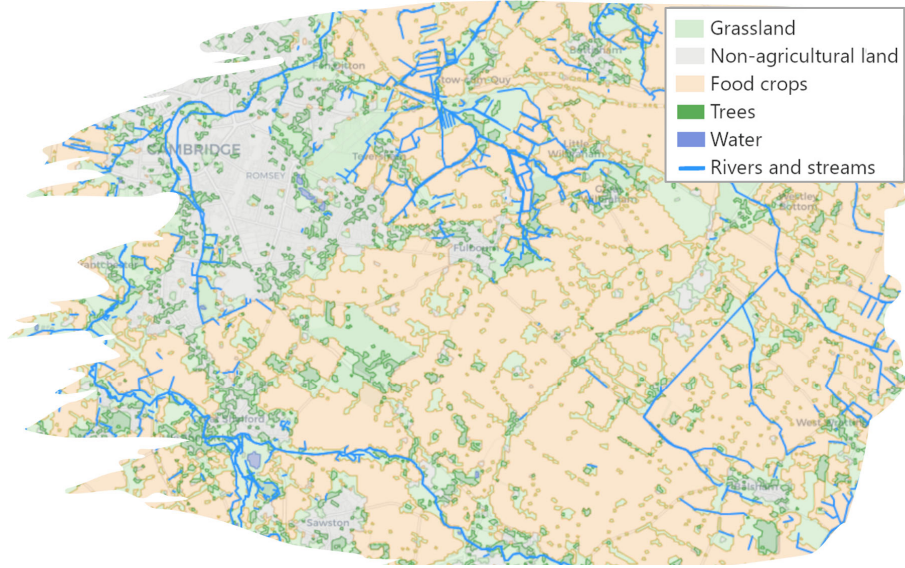
**Figure 8:** Map showing wind data for the UK at selected points in time. The data show the horizontal component of the velocity 100 m above sea level. Left: 19 Nov 2020 12:00. Right: 01 Feb 2020 15:00.

Contains modified Copernicus Climate Change Service information [2020].

The data is available under an open licence from Copernicus Products.

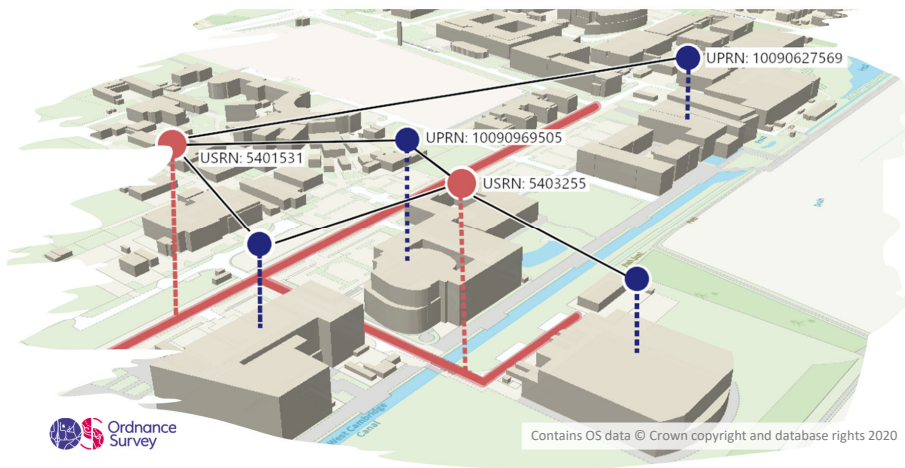


**Figures 9 and 10** show an example of land use data in the vicinity of Cambridge, UK and data about the built environment in the vicinity of the Department of Chemical Engineering and Biotechnology at the University of Cambridge. The land use data are available via the Crop Map of England [94], which provides geospatial data about the types of biomass grown throughout the whole of England. The data about the rivers and streams in Fig. 9 are available using the Ordnance Survey (OS) Features API [88]. The base map and building data in Fig. 10 are available using the OS Vector Tile API [90]. Every building with a postal address in the UK has an associated Unique Property Reference Number (UPRN),



**Figure 9:** Land use in the vicinity of Cambridge, UK.

*Contains data from the Crop Map of England (CROME) 2019 [94] licensed under an Open Government Licence.*

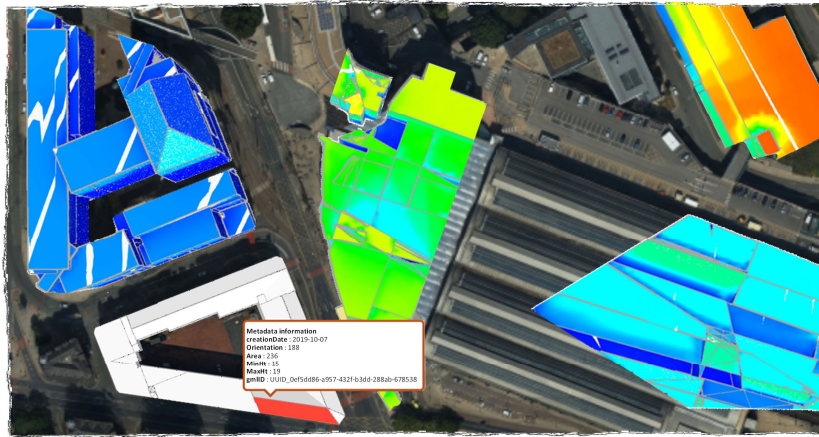


**Figure 10:** Ordnance Survey building data in the vicinity of the Department of Chemical Engineering and Biotechnology (UPRN:10090969505) and the Centre for Digital Built Britain (UPRN:10090627569) in Cambridge, UK.

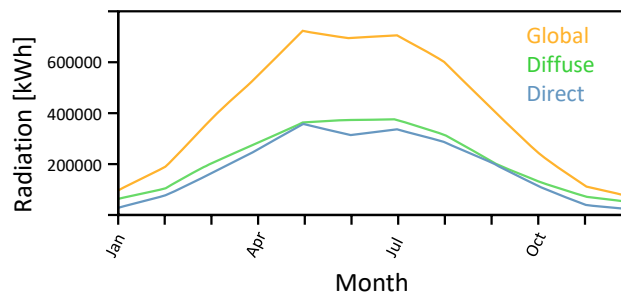
*Contains public sector information licensed under the Open Government Licence v3.0.*

likewise every road has a Unique Street Reference Number (USRN). These identifiers and items linked to them are available using the OS Linked Identifiers API [89]. Work to extend the World Avatar to include the biomass data is ongoing. Work to include information about the built environment using OntoCityGML [39] is underway as part of the Cities Knowledge Graph project (see Section 3.4.1).

**Figure 11** shows the built environment in the vicinity of Manchester Piccadilly Railway Station overlaid with data about the solar radiation incident on the roof of each building. The data is published as part of a virtual city map of Manchester [100]. This illustrates just one way in which the types of data in Figs. 6–10 can be combined. Making such data available as part of a knowledge-graph-based digital twin of the UK presents the opportunity to include it in parallel world scenario analyses. For example, to investigate the building-by-building potential of installing solar panels and battery storage, either locally within each building, at a street or at a district level.



(a) Solar Radiation on buildings. Clockwise, from bottom left. CityGML representation of building data, coloured by solar radiation in January, April, July, October. Background: Building view.



(b) Annual variation in solar radiation data for the building with UUID\_0ef5dd86-a957-432f-b3dd-288ab-678538 (the building with the metadata in the CityGML view).

**Figure 11:** *Modified screenshots of city data in the vicinity of Manchester Piccadilly Railway Station [100].*

*Contains public sector information licensed under the Open Government Licence v3.0.*

### 3.5.2 Goal alignment for parallel worlds

The parallel worlds capability of the World Avatar offers a powerful tool to explore the design space of complex problems. However, the problems of interest, for example how best to develop the UK infrastructure, rapidly become too complex to be treated in the manner of classical optimisation problems. This is the case both in the sense that the problems are not amenable to defining a straightforward objective function and in the sense that the design space is too complex to be explored without some additional insight to suggest propitious scenarios. In an ideal world, computational agents would suggest scenarios and automate the exploration of the design space. In order to attempt this, it will be necessary to equip the agents with an understanding of our values and what we, as a society, want to achieve. In other words, we need to equip the agents with *goals*.

There are many ways to define goals. The Sustainable Development Goals (SDGs) [99] defined by the UN offer one possible starting point. There are 17 SDGs, which are further specified in terms of 169 targets, each of which is associated with a set of proposed indicators. Each level in this hierarchy is increasingly specific, so whilst the goals are abstract, the indicators are specific and measurable. Table 1 shows an example for SDG 9.

**Table 1:** *Sustainable Development Goal 9, Target 9.4 and Indicator 9.4.1.*

SDG	Description
Goal 9	<i>Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation.</i>
Target 9.4	<i>By 2030, upgrade infrastructure and retrofit industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes, with all countries taking action in accordance with their respective capabilities.</i>
Indicator 9.4.1	<i>CO<sub>2</sub> emission per unit of value added.</i>

A Sustainable Development Goals Interface Ontology (SDGIO) already exists [18]. Work is underway to link the World Avatar to this ontology and to create agents to evaluate selected SDG Indicators. The current focus is on Indicator 9.4.1. Data for all the UK power stations (and in fact power stations worldwide) is available via the World Avatar, and agents are being developed to estimate the emissions intensity of power generation in the UK and worldwide. In the first instance, this will be calculated as CO<sub>2</sub> emissions per unit electrical power produced, as opposed to per unit value added. The rationale for this is that the power produced is a measurable physical quantity, whereas value added is less straightforward to define. The ability to evaluate these indicators and learn how actions investigated in parallel worlds translate to progress towards things like the SDGs provides a possible route for how to develop the insight required to suggest new scenarios.

Another possible route forward could be to calculate goals based on environmental, social and governance (ESG) metrics, either for individual companies or for specific industrial

sectors. Such metrics are projected to be incorporated into future financial regulations and will be strongly linked to the risk and growth profiles of the companies involved. The use of the dynamic knowledge graph could help address the issues of ambiguity and consistency that currently affects such metrics because all metrics would be calculated on the basis of the same data set. For example, environmental factors such as emissions, waste management, water and energy usage could be assessed based on real-world data from the knowledge graph. Further, the parallel worlds capability would allow industries to explore the impact of different strategies, for example investment in different energy technologies, and use this to guide their corporate social responsibility (CSR) strategy.

Inevitably, there will arise scenarios where goals conflict with each other [61]. For example, building more roads may enable better access to schools and therefore better education, so lifting people out of poverty and reducing hunger. However, this would also lead to more road use which could (depending on the type of transport) lead to more pollution. It is therefore necessary to consider *goal alignment*, that is to say, how to ensure that the scenarios considered by the World Avatar support the goals of humanity. This could potentially be achieved by using the goals not only to inform, but also constrain the suggestion of scenarios. The questions of how to suggest scenarios and achieve goal alignment remain important topics of research.

## 4 Conclusions

This paper demonstrates for the first time how a comprehensive digital twin can be implemented as a dynamic knowledge graph built using technologies from the Semantic Web stack. Examples are presented from the World Avatar project.

The World Avatar dynamic knowledge graph is composed of Linked Data and an ecosystem of computational agents operating on the knowledge graph. The agents enable the automated incorporation of new data and the calculation of quantities of interest, such that the dynamic knowledge graph is able to evolve in time and remain up-to-date.

The contents of the dynamic knowledge graph are instantiated using ontologies that provide semantic models of the domains of interest and of computational agents. This design confers versatility because it enables agents to update and restructure the knowledge graph, and to discover and compose other agents simply by reading from and writing to the knowledge graph. Further, the knowledge graph could, in principle, be extended to accommodate any concept that is required in the future. It has a distributed architecture that supports a uniform SPARQL query and update interface, and that supports the same access control technologies that already secure electronic data. The use of Linked Data helps address the issues of ambiguity, and enables the linking of provenance information and related data, supporting the discovery and reuse of data and results.

Use cases from the World Avatar demonstrate key aspects of the dynamic knowledge graph. The Cities Knowledge Graph project is being used to develop a pilot for a comprehensive knowledge management platform that provides interoperability between different types of city-relevant data to improve the planning of cities. The PIPS project demonstrates the versatility of the dynamic knowledge graph both in terms of the breadth of



domains that it can accommodate, and that it can interact with systems in the real world. Use cases from the J-Park Simulator were used to introduce the concepts of the base world and parallel worlds. The base world describes the real world as it is and allows for intelligent control of infrastructure. The parallel worlds allow scenario analysis to support the intelligent exploration of alternative designs whilst remaining linked to, but without affecting the base world.

The use case from the base world demonstrates the cross-domain interoperability of the dynamic knowledge graph. Agents calculate the impact of shipping on the air quality by combining live feeds of weather and shipping data, fuel chemistry and combustion simulations with atmospheric dispersion simulations. The agents wrap around commercial software and third party APIs, and offer black box, gray box and physics-based models. A composition agent is used to select which agents to deploy. A smart contract system enables agent selection to be performed using quality of service (QoS) measures, and offers the possibility of a software-as-a-service licensing model.

The use case from the parallel world demonstrates the use of the dynamic knowledge graph to perform scenario analysis. The example explores how the design of possible energy systems changes when a carbon tax is levied on emissions from power generation. It identifies the tax level that is required to motivate a change to cleaner energy technologies, and calculates the optimal design and operating conditions of the new system. The parallel worlds capability raises a number of important research questions relating to how to explore design spaces that are too complex to be addressed as conventional optimisation problems. For example, how to assess whether a scenario is beneficial, how to make good suggestions for alternative scenarios and how to ensure alignment between the scenarios considered by the digital twins and the goals of society? One option is to take inspiration from the Sustainable Development Goals (SDGs) defined by the UN and from metrics for environmental, social, and governance (ESG) in the context of corporate decision making. Work to evaluate progress towards the SDGs has been started.

Work to develop a knowledge-graph-based digital twin of the UK is ongoing. The next iteration of the digital twin will be applied in the context of supporting the decarbonisation of the UK energy landscape and will enable the semantic annotation of data describing the power system and the electrical and gas transmission systems, of data describing solar radiation and wind, and of data describing what biomass is grown where. The Semantic Web stack provides tools that confer a number of advantages on knowledge-graph-based digital twins. However, the implementation of a digital twin that meets the needs of the National Digital Twin remains hugely challenging and poses a number of open questions. Nonetheless, there are good reasons to be optimistic. The achievements of Google, Microsoft, IBM, Facebook and eBay [86] show that these types of technical challenges can be addressed and that the results have the potential to be extraordinary.

## Acknowledgements

Thank you to the members of the Computational Modelling (CoMo) Group in the Department of Chemical Engineering and Biotechnology at the University of Cambridge, the members of the Cambridge Centre for Advanced Research and Education in Singapore (CARES) and the members of Computational Modelling Cambridge Ltd. (CMCL) who have contributed to the research underlying the ideas described in this paper. Special thanks to Tom Savage who helped prepare a number of the figures and to Dr Feroz Farazi who helped check some of the text.

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. Part of the research was also funded by the European Commission, Horizon 2020 Programme, DOME 4.0 Project, GA 953163. Additional support for a number of PhD studentships was provided by Computational Modelling Cambridge Ltd.

## Nomenclature

<b>ABox</b>	Assertional Component (of an ontology)
<b>ADMS</b>	Atmospheric Dispersion Modelling System
<b>AI</b>	Artificial Intelligence
<b>API</b>	Application Programming Interface
<b>BC</b>	Black Carbon
<b>BMS</b>	Building Management System
<b>CCGT</b>	Combined Cycle Gas Turbine
<b>CFD</b>	Computational Fluid Dynamics
<b>CSR</b>	Corporate Social Responsibility
<b>DUKES</b>	Digest of UK Energy Statistics
<b>EPISODE</b>	Eulerian Urban Dispersion Model
<b>ESG</b>	Environmental, Social and Governance
<b>ESRI</b>	Environmental Systems Research Institute
<b>GeoSPARQL</b>	Geographic Query Language for RDF Data
<b>HPC</b>	High Performance Computing
<b>IMF</b>	Information Management Framework
<b>IRI</b>	Internationalised Resource Identifier
<b>LOD</b>	Linked Open Data
<b>NLP</b>	Natural Language Processing
<b>NO<sub>x</sub></b>	Nitrogen oxides
<b>OBDA</b>	Ontology-Based Data Access
<b>OGC</b>	Open Geospatial Consortium
<b>OPF</b>	Optimal Power Flow
<b>OS</b>	Ordnance Survey
<b>PIPS</b>	Pharma Innovation Programme Singapore
<b>PM</b>	Particulate Matter
<b>PM<sub>10</sub></b>	Particulate Matter less than 10 $\mu\text{m}$ in diameter
<b>PM<sub>2.5</sub></b>	Particulate Matter less than 2.5 $\mu\text{m}$ in diameter
<b>QoS</b>	Quality of Service
<b>QSAR</b>	Quantitative Structure Activity Relationship
<b>QSPR</b>	Quantitative Structure Property Relationship
<b>RDF</b>	Resource Description Framework
<b>SDGIO</b>	Sustainable Development Goal Interface Ontology
<b>SDG</b>	Sustainable Development Goal
<b>SMR</b>	Small Modular Nuclear Reactor
<b>SPARQL</b>	SPARQL Protocol and RDF Query Language
<b>SQL</b>	Structured Query Language
<b>SRM</b>	Stochastic Reactor Model
<b>TBox</b>	Terminological Component (of an ontology)
<b>UPRN</b>	Unique Property Reference Number
<b>USRN</b>	Unique Street Reference Number

## References

- [1] Blazegraph. URL <https://blazegraph.com>. Accessed Dec 2020.
- [2] Cambridge Centre for Advanced Research and Education in Singapore. URL <https://www.cares.cam.ac.uk>. Accessed Jun 2020.
- [3] Cities Knowledge Graph. URL <http://www.cares.cam.ac.uk/research/cities>. Accessed Jun 2020.
- [4] PIPS: Development of Multi-Step Processes in Pharma. URL <https://www.cares.cam.ac.uk/research/pips>. Accessed Dec 2020.
- [5] Computational Modelling Cambridge Ltd. URL <https://cmclinnovations.com>. Accessed Dec 2020.
- [6] DBpedia. URL <https://wiki.dbpedia.org/>. Accessed Dec 2020.
- [7] Drools Business Rules Management System. URL <https://www.drools.org>. Accessed Dec 2020.
- [8] Apache Jena Fuseki. URL <https://jena.apache.org/documentation/fuseki2>. Accessed Dec 2020.
- [9] Global Solar Atlas. URL <https://globalsolaratlas.info/map>. Accessed Dec 2020.
- [10] HermiT OWL Reasoner. URL <http://www.hermit-reasoner.com>. Accessed Dec 2020.
- [11] Jena JDBC – A SPARQL over JDBC driver framework. URL <https://jena.apache.org/documentation/jdbc>. Accessed Dec 2020.
- [12] The Linked Open Data Cloud. URL <https://lod-cloud.net/>. Accessed Dec 2020.
- [13] Eclipse rdf4j. URL <https://rdf4j.org>. Accessed Dec 2020.
- [14] SolarGIS. URL <https://solargis.com>. Accessed Dec 2020.
- [15] The World Avatar. URL <http://theworldavatar.com>. Accessed Jun 2020.
- [16] The World Avatar, Building Management System data. URL <http://www.theworldavatar.com:82/visualizeBms>. Accessed Jun 2020.
- [17] Virtuoso Universal Server. URL <https://virtuoso.openlinksw.com>. Accessed Dec 2020.
- [18] Sustainable Development Goals Interface Ontology (SDGIO), 2015. URL <https://github.com/SDG-InterfaceOntology/sdgio>. Accessed Oct 2020.

- [19] Smart Infrastructure: Getting more from strategic assets. Cambridge Centre for Smart Infrastructure and Construction (CISL), 2016. URL <https://www-smartinfrastucture.eng.cam.ac.uk/files/the-smart-infrastructure-paper>. Accessed Nov 2020.
- [20] Data for the public good. National Infrastructure Commission, 2017. URL <https://www.nic.org.uk/wp-content/uploads/Data-for-the-Public-Good-NIC-Report.pdf>. Accessed Nov 2020.
- [21] Hydrogen in a low-carbon economy. Committee on Climate Change, 2018. URL <https://www.theccc.org.uk/publication/hydrogen-in-a-low-carbon-economy/>. Accessed Jul 2020.
- [22] Establishing a hydrogen economy: The future of energy 2035. Arup, 2019. URL <https://www.arup.com/perspectives/publications/research/section/establishing-a-hydrogen-economy-the-future-of-energy-2035>. Accessed Jul 2020.
- [23] The role of hydrogen in delivering net zero. Hydrogen Taskforce, 2020. URL [http://www.hydrogentaskforce.co.uk/wp-content/uploads/2020/03/Hydrogen-Taskforce-Report-Feb2020\\_web.pdf](http://www.hydrogentaskforce.co.uk/wp-content/uploads/2020/03/Hydrogen-Taskforce-Report-Feb2020_web.pdf). Accessed Jul 2020.
- [24] Economic impact assessment: Hydrogen is ready to power the UK’s green recovery. Hydrogen Taskforce, 2020. URL <https://www.hydrogentaskforce.co.uk/wp-content/uploads/2020/08/6-EIA-report.pdf>. Accessed Jul 2020.
- [25] D. Allemang and J. Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2011. ISBN 9780123859655.
- [26] J. Atherton, W. Xie, L. K. Aditya, X. Zhou, G. Karmakar, J. Akroyd, S. Mosbach, M. Q. Lim, , and M. Kraft. How does a carbon tax affect Britain’s power generation composition? c4e-Preprint Series, Technical Report 260, 2020. ISSN 1473–4273. URL <https://como.ceb.cam.ac.uk/preprints/260/>. Manuscript submitted for publication.
- [27] 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.
- [28] J. Bai, R. M. Geeson, F. Farazi, S. Mosbach, J. Akroyd, E. J. Bringley, and M. Kraft. Automated calibration of a poly(oxymethylene) dimethyl ether oxidation mechanism using knowledge-graph technology. c4e-Preprint Series, Technical Report 262, 2020. ISSN 1473–4273. URL <https://como.ceb.cam.ac.uk/preprints/262/>. Manuscript submitted for publication.
- [29] T. Berners-Lee. Linked data - design issues, 2006. URL <http://www.w3.org/DesignIssues/LinkedData.html>. Accessed Dec 2020.

- [30] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Sci. Am.*, 284(5): 28–37, 2001.
- [31] A. Bolton, L. Butler, I. Dabson, M. Enzer, M. Evans, T. Fenemore, F. Harradence, E. Keaney, A. Kemp, A. Luck, N. Pawsey, S. Saville, J. Schooling, M. Sharp, T. Smith, J. Tennison, J. Whyte, A. Wilson, and C. Makri. Gemini Principles. Centre for Digital Built Britain (CDBB), 2018. doi:10.17863/CAM.32260.
- [32] T. C. Bond, S. J. Doherty, D. W. Fahey, P. M. Forster, T. Berntsen, B. J. DeAngelos, M. G. Flanner, S. Ghan, B. Kärcher, D. Koch, S. Kinne, Y. Kondo, P. K. Quinn, M. C. Sarofim, M. G. Schultz, M. Schulz, C. Venkataraman, H. Zhang, S. Zhang, N. Bellouin, S. K. Guttikunda, P. K. Hopke, M. Z. Jacobson, J. W. Kaiser, Z. Klimont, U. Lohmann, J. P. Schwarz, D. Shindell, T. Storelvmo, S. G. Warren, and C. S. Zender. Bounding the role of black carbon in the climate system: A scientific assessment. *J. Geophys. Res.: Atmos.*, 118(11):5380–5552, 2013. doi:10.1002/jgrd.50171.
- [33] 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.
- [34] P. Brown, L. Cardenas, S. Choudrie, L. Jones, E. Karagianni, J. MacCarthy, N. Pasant, B. Richmond, H. Smith, G. Thistlethwaite, A. Thomson, L. Turtle, D. W. S. Bradley, M. Broomfield, G. Buys, H. Clilverd, M. Gibbs, S. Gilhespy, M. Glending, R. Gluckman, P. Henshall, M. Hobson, N. Lambert, H. Malcolm, A. Manning, R. Matthews, K. May, A. Milne, A. Misra, T. Misselbrook, T. Murrells, Y. Pang, B. Pearson, J. Raoult, J. Richardson, R. Stewart, C. Walker, J. Watterson, N. Webb, and H. Zhang. UK Greenhouse Gas Inventory, 1990 to 2018: Annual Report for Submission under the Framework Convention on Climate Change, 2020. URL [https://naei.beis.gov.uk/reports/reports?report\\_id=998](https://naei.beis.gov.uk/reports/reports?report_id=998). Accessed Dec 2020.
- [35] P. R. Brown and A. Botterud. The value of inter-regional coordination and transmission in decarbonizing the US electricity system. *Joule*, 2020. doi:10.1016/j.joule.2020.11.013. In press.
- [36] Cambridge Environmental Research Consultants. Atmospheric Dispersion Modelling System (ADMS), 2020. URL <https://cerc.co.uk/environmental-software/ADMS-model.html>. Accessed Mar 2020.
- [37] I. Capellán-Pérez, I. de Blas, J. Nieto, C. de Castro, L. J. Miguel, Ó. Carpintero, M. Mediavilla, L. F. Lobejón, N. Ferreras-Alonso, P. Rodrigo, F. Frechoso, and D. Álvarez-Antelo. MEDEAS: A new modeling framework integrating global biophysical and socioeconomic constraints. *Energy Environ. Sci.*, 13:986–1017, 2020. doi:10.1039/C9EE02627D.
- [38] Centre for Digital Built Britain. National Digital Twin programme, 2018. URL <https://www.cdbb.cam.ac.uk/what-we-do/national-digital-twin-programme>. Accessed Oct 2020.

- [39] Centre Universitaire d’Informatique at University of Geneva, 2012. URL <http://cui.unige.ch/isi/onto/citygml2.0.owl>. Accessed Dec 2018.
- [40] D. Chen, G. Doumeingts, and F. Vernadat. Architectures for enterprise integration and interoperability: Past, present and future. *Comput. Ind.*, 59(7):647–659, 2008. doi:10.1016/j.compind.2007.12.016.
- [41] CMCL Innovations. Model Development Suite, version 2020.1, 2020. URL <https://cmclinnovations.com/solutions/products/mods>. Accessed Apr 2020.
- [42] CMCL Innovations. SRM Engine Suite, version 2020.1, 2020. URL <https://cmclinnovations.com/solutions/products/srm>. Accessed Apr 2020.
- [43] Computational Modelling Cambridge Ltd. Marie. URL [https://kg.cmclinnovations.com/chemistry\\_chatbot](https://kg.cmclinnovations.com/chemistry_chatbot). Accessed Dec 2020.
- [44] K. B. Debnath and M. Mourshed. Challenges and gaps for energy planning models in the developing-world context. *Nat. Energy*, 3:172–184, 2018. doi:10.1038/s41560-018-0095-2.
- [45] J. F. DeCarolis, P. Jaramillo, J. X. Johnson, D. L. McCollum, E. Trutnevyte, D. C. Daniels, G. Akın-Olcum, J. Bergerson, S. Cho, J.-H. Choi, M. T. Craig, A. R. de Queiroz, H. Eshraghi, C. S. Galik, T. G. Gutowski, K. R. Haapala, B.-M. Hodge, S. Hoque, J. D. Jenkins, A. Jenn, D. J. Johansson, N. Kaufman, J. Kiviluoma, Z. Lin, H. L. MacLean, E. Masanet, M. S. Masnadi, C. A. McMillan, D. S. Nock, N. Patankar, D. Patino-Echeverri, G. Schively, S. Siddiqui, A. D. Smith, A. Venkatesh, G. Wagner, S. Yeh, and Y. Zhou. Leveraging open-source tools for collaborative macro-energy system modeling efforts. *Joule*, 4:2523–2526, 2020. doi:10.1016/j.joule.2020.11.002.
- [46] Department for Business, Energy & Industrial Strategy. Digest of UK Energy Statistics (DUKES): Electricity, 2020. URL <https://www.gov.uk/government/statistics/electricity-chapter-5-digest-of-united-kingdom-energy-statistics-dukes>. Accessed Dec 2020.
- [47] Department for Environment Food and Rural Affairs (United Kingdom). Clean air strategy 2019, 2019. URL <https://www.gov.uk/government/publications/clean-air-strategy-2019>. Accessed Dec 2020.
- [48] Department of Transport (United Kingdom). Vehicle emissions testing programme: Data and conclusions, 2016. URL <https://www.gov.uk/government/publications/vehicle-emissions-testing-programme-conclusions>. Accessed Dec 2020.
- [49] Department of Transport (United Kingdom). Clean maritime plan, 2019. URL <https://www.gov.uk/government/publications/clean-maritime-plan-maritime-2050-environment-route-map>. Accessed Dec 2020.

- [50] A. Devanand, M. Kraft, and I. A. Karimi. Optimal site selection for modular nuclear power plants. *Comput. Chem. Eng.*, 125:339–350, 2019. doi:10.1016/j.compchemeng.2019.03.024.
- [51] 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.
- [52] A. Eibeck, M. Q. Lim, and M. Kraft. J-Park Simulator: An ontology-based platform for cross-domain scenarios in process industry. *Comput. Chem. Eng.*, 131: 106586, 2019. doi:10.1016/j.compchemeng.2019.106586.
- [53] A. Eibeck, A. Chadzynski, M. Q. Lim, L. K. Aditya, L. Ong, A. Devanand, G. Karmakar, S. Mosbach, R. Lau, I. A. Karimi, E. Y. S. Foo, and M. Kraft. A parallel world framework for scenario analysis in knowledge graphs. *Data-Centric Engineering*, 1:e6, 2020. doi:10.1017/dce.2020.6.
- [54] A. Eibeck, D. Nurkowski, A. Menon, J. Bai, L. Zhou, S. Mosbach, J. Akroyd, and M. Kraft. Predicting power conversion efficiency of organic photovoltaics with machine learning: models and data analysis. c4e-Preprint Series, Technical Report, 2020. ISSN 1473–4273. Manuscript in preparation.
- [55] Environmental Committee, London Assembly. Driving away from diesel: Reducing air pollution from diesel vehicles, 2015. URL <https://www.london.gov.uk/about-us/london-assembly/london-assembly-publications/driving-away-diesel-reducing-air-pollution>. Accessed Dec 2020.
- [56] Environmental Protection Agency (United States). Control of Air Pollution From New Motor Vehicles: Heavy-Duty Engine Standards, 2020. URL <https://www.federalregister.gov/documents/2020/01/21/2020-00542/control-of-air-pollution-from-new-motor-vehicles-heavy-duty-engine-standards>. Accessed Dec 2020.
- [57] Environmental Systems Research Institute. ArcGIS CityEngine. URL <https://www.esri.com/en-us/arcgis/products/arcgis-cityengine/overview>. Accessed Dec 2020.
- [58] F. Farazi, J. Akroyd, S. Mosbach, P. Buerger, D. Nurkowski, M. Salamanca, and M. Kraft. OntoKin: An ontology for chemical kinetic reaction mechanisms. *J. Chem. Inf. Model.*, 60(1):108–120, 2020. doi:10.1021/acs.jcim.9b00960.
- [59] F. Farazi, N. Krdzavac, J. Akroyd, S. Mosbach, A. Menon, D. Nurkowski, and M. Kraft. Linking reaction mechanisms and quantum chemistry: An ontological approach. *Comput. Chem. Eng.*, 137:106813, 2020. doi:10.1016/j.compchemeng.2020.106813.



- [60] F. Farazi, M. Salamanca, S. Mosbach, J. Akroyd, A. Eibeck, L. K. Aditya, A. Chadzynski, K. Pan, X. Zhou, S. Zhang, M. Q. Lim, and M. Kraft. Knowledge graph approach to combustion chemistry and interoperability. *ACS Omega*, 5 (29):18342–18348, 2020. doi:10.1021/acsomega.0c02055.
- [61] F. Fuso Nerini, J. Tomei, L. S. To, I. Bisaga, P. Parikh, M. Black, A. Borrion, C. Spataru, V. Castán Broto, G. Anandarajah, B. Milligan, and Y. Mulugetta. Mapping synergies and trade-offs between energy and the sustainable development goals. *Nat. Energy*, 3:10–15, 2018. doi:10.1038/s41560-017-0036-5.
- [62] Future Cities Laboratory. City Energy Analyst. URL <https://fcl.ethz.ch/research/high-density-cities/multi-scale-energy-systems/cea.html>. Accessed Dec 2020.
- [63] Future Cities Laboratory. Collaborative Interactive Visualisation and Analysis Laboratory. URL <https://fcl.ethz.ch/research/collaboration/cival.html>. Accessed Dec 2020.
- [64] Future Cities Laboratory. Engaging Mobility. URL <https://fcl.ethz.ch/research/responsive-cities/engaging-mobility.html>. Accessed Dec 2020.
- [65] G. Gröger, T. H. Kolbe, C. Nagel, and K.-H. Häfele. OGC City Geography Markup Language (CityGML) Encoding Standard, Version 2.0.0, 2012. URL <http://www.opengeospatial.org/standards/citygml>. Accessed Dec 2018.
- [66] A. Q. E. Group. Report: Impacts of shipping on uk air quality, 2017. URL [https://uk-air.defra.gov.uk/library/reports?report\\_id=934](https://uk-air.defra.gov.uk/library/reports?report_id=934). Accessed Dec 2020.
- [67] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thépaut. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2018. doi:10.24381/cds.adbb2d47.
- [68] J. Hetherington and M. West. The pathway towards an Information Management Framework - A ‘Commons’ for Digital Built Britain. Centre for Digital Built Britain (CDBB), 2020. doi:10.17863/CAM.52659.
- [69] O. Inderwildi, C. Zhang, X. Wang, and M. Kraft. The impact of intelligent cyber-physical systems on the decarbonization of energy. *Energy Environ. Sci.*, 13:744–771, 2020. doi:10.1039/C9EE01919G.
- [70] Institute of Computer Engineering at Technical University of Vienna, 2014. URL <http://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl>. Accessed Dec 2018.
- [71] R. K. Jain, J. Qin, and R. Rajagopal. Data-driven planning of distributed energy resources amidst socio-technical complexities. *Nat. Energy*, 2:17112, 2017. doi:10.1038/nenergy.2017.112.

- [72] M. Karl, S.-E. Walker, S. Solberg, and M. O. P. Ramacher. The Eulerian urban dispersion model EPISODE – Part 2: Extensions to the source dispersion and photochemistry for EPISODE–CityChem v1.2 and its application to the city of Hamburg. *Geosci. Model Dev.*, 12(8):3357–3399, 2019. doi:10.5194/gmd-12-3357-2019.
- [73] M. J. Kleinlanghorst, L. Zhou, J. Sikorski, E. Y. S. Foo, K. Aditya, S. Mosbach, I. Karimi, R. Lau, and M. Kraft. J-Park Simulator: Roadmap to smart eco-industrial parks. In *Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing*, 2017. doi:10.1145/3018896.3025155.
- [74] G. Klyne and J. J. Carroll. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, 2004. URL <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>. Accessed Dec 2020.
- [75] N. Krdzavac, S. Mosbach, D. Nurkowski, P. Buerger, J. Akroyd, J. Martin, A. Menon, and M. Kraft. An ontology and semantic web service for quantum chemistry calculations. *J. Chem. Inf. Model.*, 59(7):3154–3165, 2019. doi:10.1021/acs.jcim.9b00227.
- [76] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer. DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. *Semant. Web*, 6(2):167–195, 2015. doi:10.3233/SW-140134.
- [77] P. J. Levi, S. D. Kurland, M. Carbajales-Dale, J. P. Weyant, A. R. Brandt, and S. M. Benson. Macro-energy systems: Toward a new discipline. *Joule*, 3:2282–2286, 2019. doi:10.1016/j.joule.2019.07.017.
- [78] Y. Lu. Industry 4.0: A survey on technologies, applications and open research issues. *J. Ind. Inf. Integr.*, 6:1–10, 2017. doi:10.1016/j.jii.2017.04.005.
- [79] Maritime and Port Authority of Singapore. Port Statistics, 2020. URL <https://www.mpa.gov.sg/web/portal/home/maritime-singapore/port-statistics>. Accessed Dec 2020.
- [80] W. Marquardt, J. Morbach, A. Wiesner, and A. Yang. *OntoCAPE – A Re-Usable Ontology for Chemical Process Engineering*. Springer-Verlag Berlin Heidelberg, 1st edition, 2010. doi:10.1007/978-3-642-04655-1.
- [81] S. Mosbach, A. Menon, F. Farazi, N. Krdzavac, X. Zhou, J. Akroyd, and M. Kraft. A multi-scale cross-domain thermochemical knowledge-graph. *J. Chem. Inf. Model.*, 2020. doi:10.1021/acs.jcim.0c01145. In press.
- [82] National Climate Change Secretariat (Singapore). Singapore’s long-term low-emissions development strategy, 2020. URL <https://www.nccs.gov.sg/media/publications/singapores-long-term-low-emissions-development-strategy>. Accessed Dec 2020.

- [83] National Environment Agency. Singapore’s Fourth National Communication and Third Biennial Update Report under the United Nations Framework Convention on Climate Change, 2018. URL <https://www.nccs.gov.sg/docs/default-source/default-document-library/singapore’s-fourth-national-communication-and-third-biennial-update-repo.pdf>. Accessed Dec 2020.
- [84] National Grid. Instantaneous flows into the NTS, 2020. URL <https://mip-prd-web.azurewebsites.net/InstantaneousView>. Accessed Nov 2020.
- [85] National Grid. Network route maps, 2020. URL <https://www.nationalgrid.com/uk/gas-transmission/uk/gas-transmission/land-and-assets/network-route-maps>. Accessed Nov 2020.
- [86] N. F. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor. Industry-scale knowledge graphs: Lessons and challenges. *Commun. ACM*, 62(8):36–43, 2019. doi:10.1145/3331166.
- [87] OGC. OGC GeoSPARQL - A Geographic Query Language for RDF Data, 2012. URL <https://www.ogc.org/standards/geosparql/>. Accessed Dec 2020.
- [88] Ordnance Survey. OS Features API. URL <https://osdatahub.os.uk/docs/wfs/overview>. Accessed Dec 2020.
- [89] Ordnance Survey. OS Linked Identifiers API. URL <https://osdatahub.os.uk/docs/linkedIdentifiers/overview>. Accessed Dec 2020.
- [90] Ordnance Survey. OS Vector Tile API. URL <https://osdatahub.os.uk/docs/vts/overview>. Accessed Dec 2020.
- [91] Organisation for Economic Co-operation and Development. Introduction to (quantitative) structure activity relationships, 2020. URL <https://www.oecd.org/env/ehs/risk-assessment/introductiontoquantitativestructureactivityrelationships.htm>. Accessed Dec 2020.
- [92] M. Pan, J. Sikorski, C. A. Kastner, J. Akroyd, S. Mosbach, R. Lau, and M. Kraft. Applying Industry 4.0 to the Jurong Island eco-industrial park. *Energy Procedia*, 75:1536–1541, 2015. doi:10.1016/j.egypro.2015.07.313.
- [93] M. Pan, J. Sikorski, J. Akroyd, S. Mosbach, R. Lau, and M. Kraft. Design technologies for eco-industrial parks: From unit operations to processes, plants and industrial networks. *Applied Energy*, 175:305–323, 2016. doi:10.1016/j.apenergy.2016.05.019.
- [94] Rural Payments Agency. Crop Map of England (CROME), 2019. URL <https://data.gov.uk/dataset/8c5b635f-9b23-4f32-b12a-c080e3f455d0/crop-map-of-england-crome-2019>. Accessed Oct 2020.
- [95] E. Shove. Time to rethink energy research. *Nat. Energy*, 2020. doi:10.1038/s41560-020-00739-9. In press.

- [96] E. Spyrou, B. F. Hobbs, M. D. Bazilian, and D. Chattopadhyay. Planning power systems in fragile and conflict-affected states. *Nat. Energy*, 4:300–310, 2019. doi:10.1038/s41560-019-0346-x.
- [97] The World Bank Group. Population Density (people per sq. km of land area), 2018. URL <https://data.worldbank.org/indicator/EN.POP.DNST>. Accessed 22 Oct 2020.
- [98] T. Tröndle, J. Lilliestam, S. Marelli, and S. Pfenninger. Trade-offs between geographic scale, cost, and infrastructure requirements for fully renewable electricity in europe. *Joule*, 4:1929–1948, 2020. doi:10.1016/j.joule.2020.07.018.
- [99] United Nations. Sustainable Development Goals, 2015. URL <https://sdgs.un.org/goals>. Accessed Jun 2020.
- [100] Virtual City Systems. Manchester 3D Data Viewer, 2020. URL <https://manchester.virtualcitymap.de/#/legend>. Accessed Nov 2020.
- [101] W3C. SPARQL 1.1 Overview, 2013. URL <https://www.w3.org/TR/sparql11-overview/>. Accessed Dec 2020.
- [102] W3C. Web Authentication: An API for accessing Public Key Credentials Level 1, 2019. URL <https://www.w3.org/TR/webauthn/>. Accessed Dec 2020.
- [103] World Wide Web Consortium. An Introduction to Multilingual Web Addresses, 2008. URL <https://www.w3.org/International/articles/idn-and-iri/>. Accessed Oct 2020.
- [104] S. G. Yalew, M. T. H. van Vliet, D. E. H. J. Gernaat, F. Ludwig, A. Miara, C. Park, E. Byers, E. De Cian, F. Piontek, G. Iyer, I. Mouratiadou, J. Glynn, M. Hejazi, O. Dessens, P. Rochedo, R. Pietzcker, R. Schaeffer, S. Fujimori, S. Dasgupta, S. Mima, S. R. S. da Silva, V. Chaturvedi, R. Vautard, and D. P. van Vuuren. Impacts of climate change on energy systems in global and regional scenarios. *Nat. Energy*, 5:794–802, 2020. doi:10.1038/s41560-020-0664-z.
- [105] C. Yu, M. Seslija, G. Brownbridge, S. Mosbach, M. Kraft, M. Parsi, M. Davis, V. Page, and A. Bhawe. Deep kernel learning approach to engine emissions modeling. *Data-Centric Engineering*, 1:e4, 2020. doi:10.1017/dce.2020.4.
- [106] C. Zhang, A. Romagnoli, L. Zhou, and M. Kraft. Knowledge management of eco-industrial park for efficient energy utilization through ontology-based approach. *Applied Energy*, 204:1412–1421, 2017. doi:10.1016/j.apenergy.2017.03.130.
- [107] L. Zhou, M. Pan, J. J. Sikorski, S. Garud, L. K. Aditya, M. J. Kleinlanghorst, I. A. Karimi, and M. Kraft. Towards an ontological infrastructure for chemical process simulation and optimization in the context of eco-industrial parks. *Applied Energy*, 204:1284–1298, 2017. doi:j.apenergy.2017.05.002.
- [108] L. Zhou, C. Zhang, I. A. Karimi, and M. Kraft. An ontology framework towards decentralized information management for eco-industrial parks. *Comput. Chem. Eng.*, 118:49–63, 2018. doi:10.1016/j.compchemeng.2018.07.010.

- [109] 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. *Comput. Chem. Eng.*, 130:106577, 2019. [doi:10.1016/j.compchemeng.2019.106577](https://doi.org/10.1016/j.compchemeng.2019.106577).
- [110] X. Zhou, M. Q. Lim, and M. Kraft. A smart contract-based agent marketplace for the J-Park Simulator – a knowledge graph for the process industry. *Comput. Chem. Eng.*, 139:106896, 2020. [doi:10.1016/j.compchemeng.2020.106896](https://doi.org/10.1016/j.compchemeng.2020.106896).
- [111] X. Zhou, D. Nurkowski, S. Mosbach, J. Akroyd, and M. Kraft. A question answering system for chemistry. c4e-Preprint Series, Technical Report, 2020. ISSN 1473–4273. Manuscript in preparation.
- [112] H. Ziar, P. Manganiello, O. Isabella, and M. Zeman. Photovoltatronics: Intelligent PV-based devices for energy and information applications. *Energy Environ. Sci.*, 2021. [doi:10.1039/D0EE02491K](https://doi.org/10.1039/D0EE02491K). In press.