Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

# Universal Digital Twin – Integration of national-scale energy systems and climate data

Thomas Savage[1], Jethro Akroyd[1,2], Sebastian Mosbach[1,2],
Nenad Krdzavac[2], Michael Hillman[3], Markus Kraft[1,2,4]

released: August 11, 2021

[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

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

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

Preprint No. 279

UNIVERSITY OF CAMBRIDGE

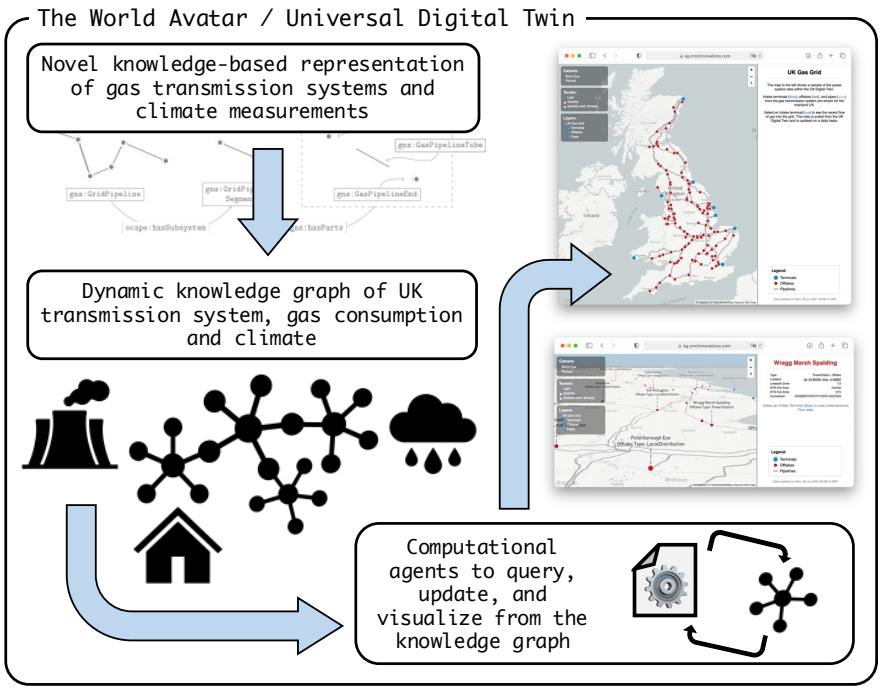
*Keywords:* Energy supply system, natural gas, climate data, ontology, knowledge graph, digital twin

**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 applies a knowledge graph-based approach to unify multiple heterogeneous domains inherent in climate and energy supply research. Existing approaches that rely on bespoke models with spreadsheet-type inputs are non-interpretable, static and make it difficult to combine existing domain specific models. The difficulties inherent to this approach become increasingly prevalent as energy supply models gain complexity while society pursues a net-zero future. In this work we develop new ontologies to extend the World Avatar knowledge graph to represent gas grids, gas consumption statistics, and climate data. Using a combination of the new and existing ontologies we construct a Universal Digital Twin that integrates data describing the systems of interest and specifies respective links between domains. We represent the UK gas transmission system, and HadUK-Grid climate data set as linked data for the first time, formally associating the data with the statistical output areas used to report governmental administrative data throughout the UK. We demonstrate how computational agents contained within the World Avatar can operate on the knowledge graph, incorporating live feeds of data such as instantaneous gas flow rates, as well as parsing information into interpretable forms such as interactive visualisations. Through this approach, we enable a dynamic, interpretable, modular and cross-domain representation of the UK that enables domain specific experts to contribute towards a national-scale digital twin.


## Highlights

- Ontologies created to represent gas networks, gas consumption and climate data.
- Gas and climate data integrated into Universal Digital Twin based on World Avatar.
- Computational agents used to provide live data feeds to digital twin.

# Contents

# 1 Introduction

The gas grid in the UK is responsible for the distribution of gas from intake terminals on the coast to domestic and industrial end users. The grid has existed in a near constant state of flux since its construction in the 19$^{th}$ century [34]. The grid is currently used to distribute natural gas, which is responsible for 52% of carbon dioxide emissions from the UK (2020) [14]. It is possible that the next evolution of the grid may see it adapted to deliver hydrogen to mitigate carbon emissions whilst ensuring energy security [5, 17]. The role hydrogen may have in a net zero UK has been outlined by the Committee for Climate Change [9] which deems hydrogen promising in low-regret short-term scenarios such as blending with natural gas, as well as longer term scenarios in providing peak energy alongside heat pumps, taking advantage of the flexibility of the gas grid to smooth out fluctuations between energy supply and demand.

Assessing how best to use the gas grid to support net zero requires models that describe the interactions and dependencies between technologies included in the energy mix. As the energy mix becomes increasingly varied, the scenarios considered by the models will necessarily increase in complexity [50]. Inevitably, such analyses will build upon diverse heterogeneous data sets and will likely include sub-models that consider a range of factors, for example including more detailed geospatial and temporal descriptions of renewables, social and environmental factors. As the complexity increases, it is likely that modelling will transition from single-institution teams to distributed, collaborative teams, so that multiple domain experts are able to contribute to a given analysis [11, 54]. O'Dwyer et al. [42] demonstrate a Sustainable Energy Management System (SEMS) 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 unrealized.

The sub-optimal organisation of complex models and data creates problems. With respect to energy, it is important to ensure models and assumptions are clearly understood, and that data are transparent [48]. The types of data particularly relevant to energy scenarios are time-series, geographic, and tabular data [48]. Current energy policy research lacks open data and modelling transparency, impeding the ability not only to reproduce results, but to adapt and combine existing models [47]. The popular MARKAL and TIMES United Kingdom energy models [23] are highlighted by DeCarolis et al. [10] as examples of models that would benefit from increased interpretability in how they handle the large quantities of data required by the models. The authors describe a typical workflow of entering data into a series of spreadsheets, with all changes and edits being performed manually. The problems exemplified by this type of workflow are widespread, where for example Delmelle [12] notes that '*fusing a multitude of types of data together in creative ways remains a challenge*' in the context of geospatial data. It is clear that future tools will have to incorporate different types of data from a variety of domains.

Knowledge graphs are a promising technology to describe a broad range of domain specific information in an interpretable and modular manner. The information is represented using ontologies expressed as a directed graph, where the nodes of the graph represent concepts and instances, and the edges between nodes represent the respective relations between nodes. By specifying the relationships between data, the information becomes more accessible, making it easier for computational agents to interpret, query, and update

the data. The World Avatar project [18, 19] is exploring the use of *dynamic* knowledge graphs to enable interoperability between models and data from different domains. The dynamic knowledge graph is operated on by computational agents that read, manipulate, and update the nodes and edges of the knowledge graph, including adding new data, new concepts and new relations. The computational agents are themselves described in the dynamic knowledge graph. This forms a critical part of the design because it confers the ability to discover agents by reading from the knowledge graph, and the ability to create new agents, for example by combining existing agents to perform composite tasks, by writing to the the knowledge graph. Each node in the knowledge graph has a unique identifier, allowing multiple agents and data sets to refer unambiguously to the same entity. Given a suitable ontologies, it is possible to represent anything. Therefore, temporal, dynamic, and geospatial data can be integrated, facilitating the complex representation of systems starting from simple sets of rules. The ability of computational agents to input data, simulate the behaviour of systems and provide output has led to the suggestion of dynamic knowledge graph technology providing a suitable architecture for implementing a Universal Digital Twin [2].

The **purpose of this paper** is to extend the World Avatar by developing ontologies to describe gas transmission systems, gas consumption data and climate observations. The ontologies are used to integrate climate observations for the first time with information relating to the UK gas transmission system, gas consumption and statistical output areas used to report governmental administrative data into large-scale Universal Digital Twin within the World Avatar. Computational agents are used to incorporate live data so that the resulting dynamic knowledge graph remains current in time. The paper is structured as follows. Section 2 provides background about the World Avatar and the systems we represent in this paper. Section 3 details the methodology used to develop the ontologies, and to instantiate and query the knowledge graph. Section 4 presents a use case that outlines the instantiation of the knowledge graph, and demonstrates the use of agents to create data pipelines and query geospatial data. Finally, Section 5 draws conclusions and discusses future work.

# 2 Background

## 2.1 Introduction to knowledge graphs

A knowledge graph $\mathscr{G}$ consists of a set of triples. Each triple $t \in \mathscr{G}$ contains a subject $s \in C$, predicate $p \in P$ and object $o \in C$ where $C$ is a set of concepts and $P$ is the set of possible relations between concepts. The subject and object define nodes within the knowledge graph, with predicates defining the connections between these nodes. Similar to how web pages are assigned URLs, subjects predicates and objects are each given Internationalized Resource Identifiers (IRIs). However, IRIs do not need to be informative themselves as they are designed to be machine readable. Rather, for a human to understand triples, subjects predicates and objects are assigned additional 'label' predicates in order to provide context to human users. An example set of triples within a knowledge

graph is as follows:

$$\underbrace{\begin{array}{c}\texttt{<example.org/29b4667e>}\\ \texttt{<example.org/a8350827>}\\ \texttt{<example.org/75e3e506>}\end{array}}_{\text{Example machine-readable triple}} \begin{array}{c}\xrightarrow{\texttt{RDFS:label}}\\ \xrightarrow{\texttt{RDFS:label}}\\ \xrightarrow{\texttt{RDFS:label}}\end{array} \underbrace{\begin{array}{c}\texttt{Hartlepool 005D}\\ \texttt{Has Closest NTS Offtake}\\ \texttt{Ganstead Offtake}\end{array}}_{\text{Respective human-readable URI labels}}$$

Where `RDFS:` denotes the Resource Description Framework Schema namespace `http://www.w3.org/2000/01/rdf-schema/#`. For the remainder of this work we reference human-readable labels of classes and instances unless otherwise stated. The namespaces used in the rest of this paper are defined in the nomenclature.

Knowledge graphs can be divided into two sets of triples. The first $\mathscr{G}_\mathrm{A} \in \mathscr{G}$ contains *assertive* relations and the second $\mathscr{G}_\mathrm{T} \in \mathscr{G}$ contains *terminological* relations. A set of terminological triples $\mathscr{G}_\mathrm{T}$ is also known as an ontology. An ontology defines the triples that can appear within a knowledge-graph, originating from the philosophical idea of *what is known*. Typical triples within an ontology consist of the definition of classes, relations, and the domain and ranges over which relations can take. For example:

```
<GasGridOfftake, Type, Class>
<HasConnectedPipeline, Type, ObjectProperty>
<HasConnectedPipeline, Domain, GasGridOfftake>
<HasConnectedPipeline, Range, GasPipeline>
```

defines `Gas Grid Offtake` as a class within the knowledge graph, stating that this represents a subject or object of a triple, `Has Connected Pipeline` is defined as an object property stating that it should relate a subject and object within the knowledge graph. Lastly this object property is assigned a domain and range that specifies what subjects and objects it should relate. It can be seen that by building up a representation of the systems we wish to represent using a collection of basic triples within an ontology, we specify precisely what is known, providing interpretability.

The second aspect of a knowledge graph $\mathscr{G}_\mathrm{A}$ concerns the assertional triples. This is where concrete examples of classes such as physical entities are defined and is where data exists within the knowledge graph. For example in the following three triples

```
<GansteadOfftake, Class, GasGridOfftake>
<Ganstead-Asselby, Class, GasPipeline>
<GansteadOfftake, HasConnectedPipeline, Ganstead-Asselby>
```

we specify instances of the `Gas GridOfftake` and `GasPipeline` class which are subsequently 'connected' using the semantic relation previously defined. Using a logical *reasoner* we can check whether the triples defined within $\mathscr{G}_\mathrm{A}$ follow the rules defined in $\mathscr{G}_\mathrm{T}$. If they do not, then the knowledge-graph is deemed *inconsistent*.

## 2.2 The World Avatar project and a Universal Digital Twin

The World Avatar project seeks to investigate how a dynamic knowledge graph can be used to integrate multi-scale cross-domain knowledge to create a world model [18]. The dynamic knowledge graph is operated on by computational agents. The agents are themselves described in the

knowledge graph so they can be discovered by reading from the knowledge graph, and can be combined to create new agents with composite functionality by writing to the knowledge graph. The computational agents can perform a wide variety of tasks including updating the knowledge graph with new data, simulating the behaviour of systems described in the knowledge graph, and analysing the results of such simulations. These capabilities form the basis of the notion that the dynamic knowledge graph contains a *base world* that provides a model of the world that remains current in time, and *parallel worlds* where alternative scenarios can be hypothesised based on the current base world, and agents used to simulate the behaviour of the parallel world to perform what-if scenario analysis to support enhanced decision making [19].

The World Avatar is implemented using technology based on the Semantic Web [52]. This choice is intended to ensure that the data in the dynamic knowledge graph is findable, accessible, interoperable and reusable as per the FAIR Guiding Principles for scientific data [53]. The expressive power of ontologies means that the knowledge graph can represent and integrate data for almost anything. The applications of the World Avatar to date have focused on decarbonisation [15, 16, 26, 43, 44, 55], city planning [7, 51] and chemistry [20, 20, 27, 32]. These examples illustrate the ability of the World Avatar to integrate models and data across different length scales and technical domains, ranging from the sub-atomic length scales of quantum chemistry calculations [21] to the application of the results of these calculations in city-scale atmospheric dispersion calculations [32]. The ability of the Semantic Web to support a distributed architecture and to represent and integrate heterogeneous data and models in a form that is discoverable and queryable via a uniform interface, combined with the ability of computational agents to input data, simulate the behaviour of systems and provide output has led to the suggestion of dynamic knowledge graph technology providing a suitable architecture for implementing a Universal Digital Twin [2]. Recent work building on this idea has investigated the effect of a carbon tax on the power system [3] and developed a description of land use [1] in the UK.

## 2.3 Domain specific knowledge

This paper demonstrates the modularity of a knowledge graph-based digital twin by combining existing sources of Linked Data (*i.e.* data that is already expressed in triples) with new semantic representations of the gas transmission system, and climate throughout the UK. In this section we outline the sources of information we consider in the construction of a dynamic knowledge graph.

### 2.3.1 ONS linked geography data

The Office for National Statistics (ONS) publishes Geography Linked Data [40]. First issued in October 2018 [41], this collection of triples provides a geospatial representation of output areas within the UK. Through the use of a `within` relation, areas of different size are related to each other. These range from the entire UK down to areas containing on average 1500 people, known as lower super output areas (LSOA). This relation enables data that is associated to the smallest output areas to be easily aggregated to larger regions. **Fig.** 1 highlights the structure of this aspect of the knowledge graph, omitting relations that we do not make use of such as `LandHectarage` and `OperativeDate`. For a complete list of relations within this data set see [40].

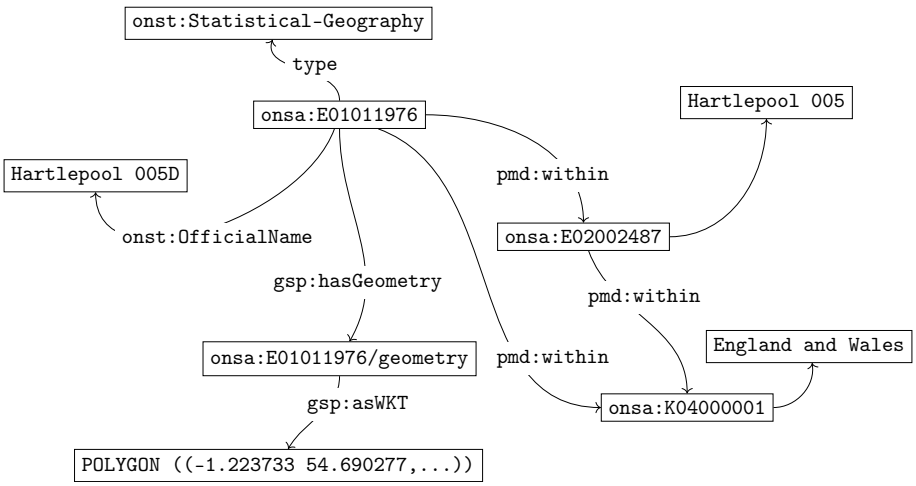



**Figure 1:** *Representation of statistical output areas as linked data, or assertional triples $\mathscr{G}_A$. The example shown is for the Hartlepool 005D output area with respective code E01011976.*

Delmelle [12] demonstrates the risks of performing data-driven geography with samples of uneven population size. Output areas as reported by the Office for National Statistics [39] are designed in a manner to approximately cover areas of equal population, social demographic and built environment based on census data. Cockings et al. [8] outlines the methodology for the construction of the output areas utilised by the ONS.

### 2.3.2 UK gas transmission system and gas consumption

The UK gas transmission system, also known within technical documents as the National Transmission System (NTS), consists of pipelines that transport high pressure gas from intakes near the coast to major industrial users such as power stations or local distribution offtakes where gas is further distributed to low pressure domestic gas networks. Compression stations and valves are situated throughout the NTS in order to maintain adequate pressure across the entire system based on fluctuating supply and demand. A key advantage of the gas transmission system is the flexibility it provides in energy supply. The quantity of gas contained within the grid at any one time is referred to as the *linepack*. Short term fluctuations in demand such as daily load changes can be met by pressurising the grid in the evening, therefore increasing the total linepack. Likewise long-term seasonal changes in demand can be met through the decompression of liquefied natural gas (LNG) imports or the storage of gas within underground caverns.

Broadly, information regarding the NTS does not exist in a single location, and as such, key pieces of information that define what the NTS *is* must be identified to be parsed into relational triples. Apart from physical infrastructure itself such as pipes and compression stations, we also consider statistics associated with gas consumption, flow rates of gas at points throughout the grid, and additional knowledge regarding infrastructure.

The sources of information relating to the NTS that we represent as linked data within the knowledge graph are shown in Table 1.

**Table 1:** *Sources of information as they relate to the UK gas transmission system including both static and dynamic data over a variety of file formats.*

| Data source | Description | Reference | File format | Size | Frequency |
|---|---|---|---|---|---|
| Gas Grid Route Map | Locations of all gas pipelines throughout the UK as well as corresponding information regarding ownership | [36] | .shp | 5.3 MB | N/A |
| Gas Grid Site Map | Locations of all gas infrastructure sites throughout the UK, measured in Million Cubic Meters per Day (mcm/day). | [36] | .shp | 257 KB | N/A |
| Gas Grid Site Information | Information regarding NTS offtakes such as NTS exit area and zone as well as linepack zone. | [33] | .csv | 25 KB | N/A |
| Instantaneous Flow Rates | Flow rate of gas entering major gas terminals throughout the UK | [35] | .csv | 30 KB | 2 minutes |
| Sub-national consumption statistics | Yearly mean and median gas consumption across Lower Layer Super Output Areas (LSOAs). Measured in kWh per gas meter, also containing information regarding the number of meters per LSOA. | [13] | .csv | 32.5 MB | Annual |

It should be noted that the information in Table 1 is at this stage disjoint. That is, despite the gas grid site map and instantaneous flow rate data referring to the same physical gas terminals, the information is not cross-referenced in a consistent way. As a result it becomes increasingly difficult to keep track of sources of information as they relate to the same physical entities. This issue is common-place in energy systems modelling wherein systems such as gas and electricity overlap. Currently approaches are bespoke and often complex such as the development of new management tools to support cross-domain interactions [42]. By instantiating the concept of each gas terminal as a node within the knowledge graph, which each disjoint data set can link to, we can unify this information allowing computational agents to infer links between sources of information, therein providing a complete representation. Moreover, should additional information come-to-light, for example a hypothetical gas terminal operating condition data set, this can be easily appended to the knowledge graph by referring to the original concept of the specific gas terminal in question without requiring knowledge of existing data. Through this brief example we demonstrate the flexibility, modularity and scalablity of a knowledge graph-based solution.

### 2.3.3 HadUK grid climate observations

The HadUK-Grid climate data set [38, 45] created by the Met Office, consists of values for various climate variables over a 1 km by 1 km grid covering the entire of the United Kingdom. The variables available within the data set include: minimum, maximum and mean air temperature, precipitation, hours of sunshine, mean sea level pressure, wind speed, relative humidity, vapour pressure, snow cover and frost cover. Values are calculated through the interpolation of measurements at approximately 540 weather stations. Perry and Hollis [45] outlines the specific regression procedure used to generate monthly values.

Each grid point contains a discrete climate variable value for each month dating back to 1862. It should be noted that whilst the number of weather stations has changed since 1862 the grid over which interpolated values are presented is constant.

One of the key advantages of climate data as published as a uniform grid is that '*regional values can be produced for any arbitrary area with greater accuracy and consistency*' [30]. This advantage in the aggregation of climate data provides a benefit when considering its addition to the knowledge graph and subsequent linking to other, previously outlined aspects.

# 3 Methodology

## 3.1 Ontology development

In this section ontologies are created to ensure that entities within the knowledge graph can be described using an appropriate vocabulary. These ontologies specify the rules as to which triples can logically exist and which cannot. For example a pipe segment has a single input and output, if triples were created allocating two outputs to a single pipe segment this would be reasoned as logically inconsistent within the rules of the ontology.
A guiding principle of ontology creation is that concepts should be reused from existing ontologies as much as possible to facilitate links across domains [37]. In this work we define two new ontologies that are used alongside concepts from existing ontologies including OntoCAPE [31], and the Ontology of Units of Measure [49].

### 3.1.1 OntoGasGrid

To represent the concept of a gas grid ontologically we decompose the system into its parts and the whole they form. By decomposing a gas grid as such, we produce the set of rules that a gas grid must abide by in the form of an ontology, $\mathscr{G}_{\mathrm{T}}$. We base the ontology on the vocabulary used to describe systems, respective subsystems and their parts defined by the OntoCAPE [29], an ontology to represent chemical processes. Specifically we make use of the upper-layer ontology.

The first aspect we represent are physical gas pipelines. **Fig.** 2 outlines how gas transmission pipelines are represented as triples. A mereological approach is taken, that is decomposing a system into respective parts and the whole they form. In this case, a `GridPipeline` is described as a combination of `GridPipelineSegment` instances. Each `GridPipelineSegment` is a system containing three parts: the start of the pipe, end of the pipe and connecting tube. **Fig.** 3 illustrates how these discrete pipe segments are subsequently connected to form a complete grid pipeline.

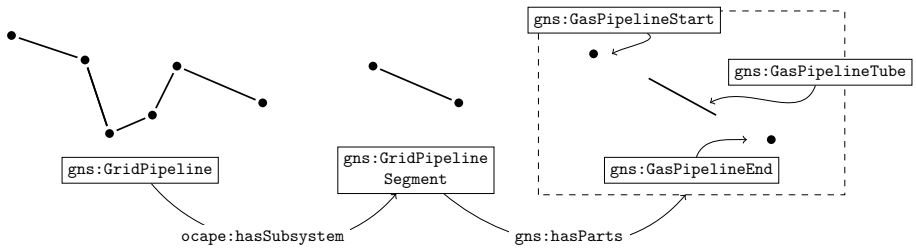



**Figure 2:** *Outline of how pipelines are decomposed into respective segments and their parts within OntoGasGrid.*

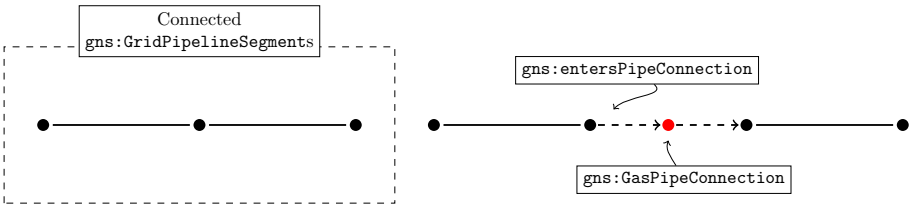



**Figure 3:** *Example of how two connected pipe segments are related, specifying their connection.*

As shown in Fig. 3, the concept of a `GasPipeConnection` is introduced, allowing the end of one `GridPipelineSegment` to be specified as connected to the start of another. A longitude and latitude is assigned to each `GasPipeConnection`, as opposed to specifying the coordinates of the start and end of a pipe segment. This ensures that two pipe segments that start and end in different locations respectively cannot be deemed 'connected'.

Aside from physical gas pipelines and their connectivity, OntoGasGrid also describes connected grid infrastructure including gas terminals and offtakes such as industrial users or power stations. The main class within this aspect of the ontology is that of a `GridComponent`, which consists of four main subclasses themselves decomposed into specific classes of infrastructure. This hierarchy is shown in **Fig.** 4.

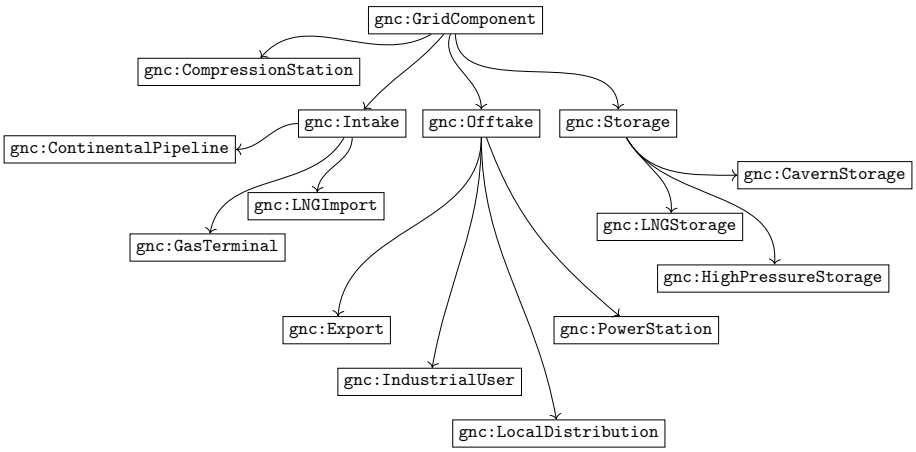



**Figure 4:** *Hierarchy of grid infrastructure in OntoGasGrid where all arrows represent the property* `SubClassOf`.

The ontology contains the property `isConnectedToPipeline`, with domain `GridComponent` and range `GasPipeConnection`. This property links connected infrastructure (*i.e.* instances of the class `GridComponent`) and gas pipelines. By providing this relation we are able to link the previously separate gas grid site map and gas grid route map presented in Table 1. Section 4.4 outlines how geospatial calculations are performed to identify connections between infrastructure and pipelines.

The complete description logic (DL) representation of OntoGasGrid is provided within the appendix. At the time of writing OntoGasGrid contains 79 classes, 18 data properties which are associated to specific classes, and 841 axioms.

### 3.1.2 OntoClimateObservations

OntoClimateObservations is a small ontology created to describe geospatial climate observations semantically. The ontology defines the minimum terminology to provide a link between previously described statistical regions and the concept of a climate measurement. Therefore, the ontology itself makes no effort to semantically describe the generating system (*i.e.* climate itself) and instead focuses on the concept of a measurement.

The complete ontology is shown graphically in **Fig.** 5

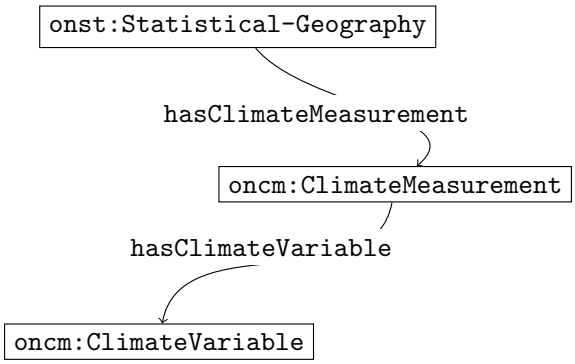



**Figure 5:** *Ontology, $\mathcal{G}_T$ to describe climate measurements associated to statistical regions. An example of assertional triples $\mathcal{G}_A$ using this ontology is shown in Fig 8.*

By providing a link to statistical output areas we enable the potential unification of statistics published throughout these areas, such as sub-national gas consumption, with gridded climate data sets such as HadUK-Grid. The complete description logic (DL) representation of OntoClimateObservations is provided within the appendix.

## 3.2 Computational agents

Computational agents are described in the knowledge graph using an agent ontology [56]. When activated, the agents interact with the knowledge graph to facilitate knowledge population, maintenance, information processing and retrieval. In this paper, agents are created to instantiate domain specific knowledge using vocabularies from OntoGasGrid, OntoClimateMeasurements and other existing ontologies. The agents exhibit three types of behaviour.

- **Input.** Agents convert information and sources of data into new triples that extend the knowledge graph. The input to the knowledge graph can be either static one-off information such as the location of physical infrastructure or dynamic information that is updated dynamically such as real-time flow rates.
- **Output.** Agents parse data from the knowledge graph to interact with the physical world, for example by controlling actuators or displaying data in convenient human-readable forms.
- **Update.** Agents query the knoweldge graph, calculate new information, for example optimised model parameters [4], and update the knowledge graph with the results, either through the modification of existing triples or by the creation of new triples. Such agents may also perform maintenance tasks such as the detection and deletion of invalid triples.

The agents developed in this work are described in detail in the following section.

# 4 Use case

In this section we outline the agents responsible for creating instances of classes previously outlined, such as gas grid infrastructure and climate values. Geospatial visualisations are enabled by a series of output agents.

## 4.1 Instantiation of HadUK-Grid climate observations

When considering the addition of the HadUK-Grid data set into the knowledge graph there are two potential approaches.

1. Insert the HadUK-Grid data set directly within the knowlege graph by representing individual grid points and respective climate variable values as triples.
2. Link the HadUK-Grid data set to ONS statistical regions as opposed to representing individual grid points.

In this work we take the second approach, first aggregating values within statistical regions and subsequently assigning values such as mean temperature to instances of these regions as opposed to representing grid points themselves.

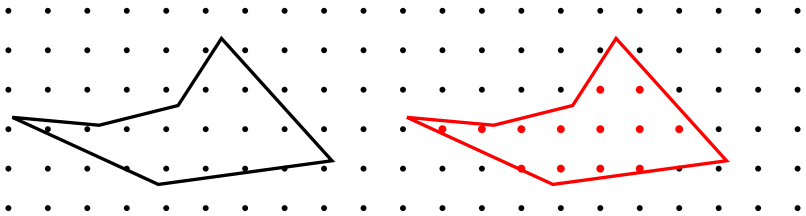
When considering the compatibility between grid points and statistical regions it is noted that a statistical region may contain multiple grid points or alternatively a single point. In the case of small regions within urban areas a grid point may not even be enclosed within a region. This provides an additional challenge in unifying both climate variables across the UK and the set of statistical resources made available by the ONS, ensuring that regions are all assigned appropriate climate values.

A computational agent was created to interpret HadUK-Grid Network Common Data Form (netCDF) files (these are commonly used in climate research and are designed to be an appendable, portable, and self describing method of sharing array-orientated scientific data), in order to populate the knowledge graph with climate data from throughout the UK as well as link to existing concepts

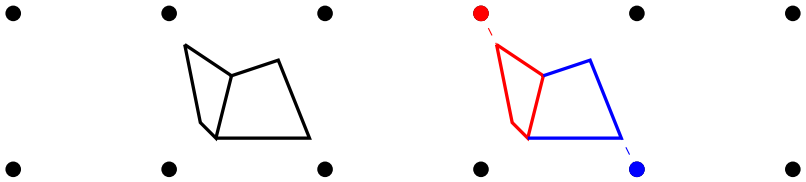
of statistical regions. Subsequently, the agent represents this information as linked data using the OntoClimateObservations and Ontology of Units of Measure ontologies and uploads these triples to the knowledge graph. Grid points are assigned to statistical regions as follows:

- If a region contains multiple grid points, take the mean of the climate variable values of respective contained points. In the case of minimum or maximum values (such as that of minimum air temperature or mean air temperature) take the minimum or maximum value respectively of the set of contained points.
- If a region contains no grid points, identify the closest grid point to the region and return associated values for climate variables of interest.

A visual demonstration of this procedure is outlined in **Fig.** 6.



(a) Area containing multiple grid points



(b) Area not containing any grid points

**Figure 6:** *Demonstration of grid points associated to an example output area in the case that a) the area contains multiple grid points, and b) the area does not contain a single grid point.*

In this work we use mean, minimum and maximum temperature variables however other variables may be appended in the future using the same procedure.

A flowchart detailing the HadUK-Grid climate input agent is described in **Fig.** 7.

The reason we take this approach as opposed to representing grid points themselves is as follows: geospatial reasoning within knowledge graphs (for example queries such as 'which grid points lie within this region?') is not fully implemented across triple-stores. Standards such as geoSPARQL [46] in theory allow for geospatial queries to be performed however currently the standard is not fully adopted. Jovanovik et al. [25] performs a GeoSPARQL benchmark across the most commonly used triple-stores, concluding that the GeoSPARQL standard, almost nine years after its initial release, is often only partially supported by major triple-stores [25]. For this reason, inferring links between the discrete climate grid points and statistical regions becomes temperamental and dependant on the specific choice of triple-store. There is also the argument that specialised databases such as triple-stores are not in-fact the appropriate location to perform geospatial reasoning such as standard set operations, for example finding the union of two regions, the closest discrete point to a region *etc.* [22].

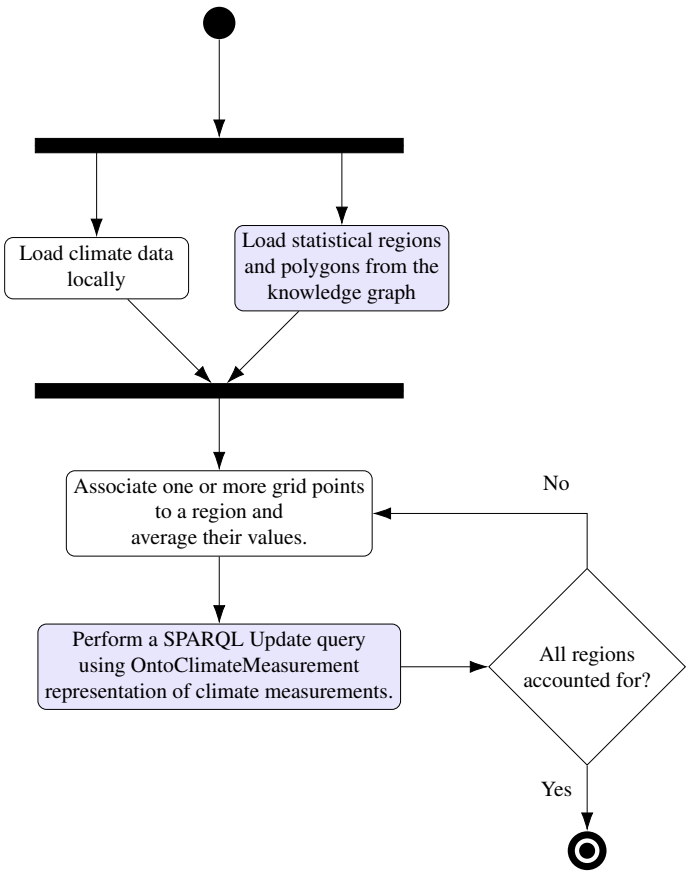



**Figure 7:** *UML diagram describing how information from the HadUK-Grid climate data set [38] is instantiated within the knowledge graph using a computational agent that associates discrete grid points with statistical regions. Purple shading indicates actions that interact with the knowledge graph.*

Geospatial calculations within the agent such as the identification of discrete points within regional polygons were performed by loading WKT representations of regions stored directly within the knowledge graph into the Shapely Python library [22]. The agent operates using Python 3.7.9.

An example upload query produced by the agent is demonstrated graphically in **Fig. 8**.

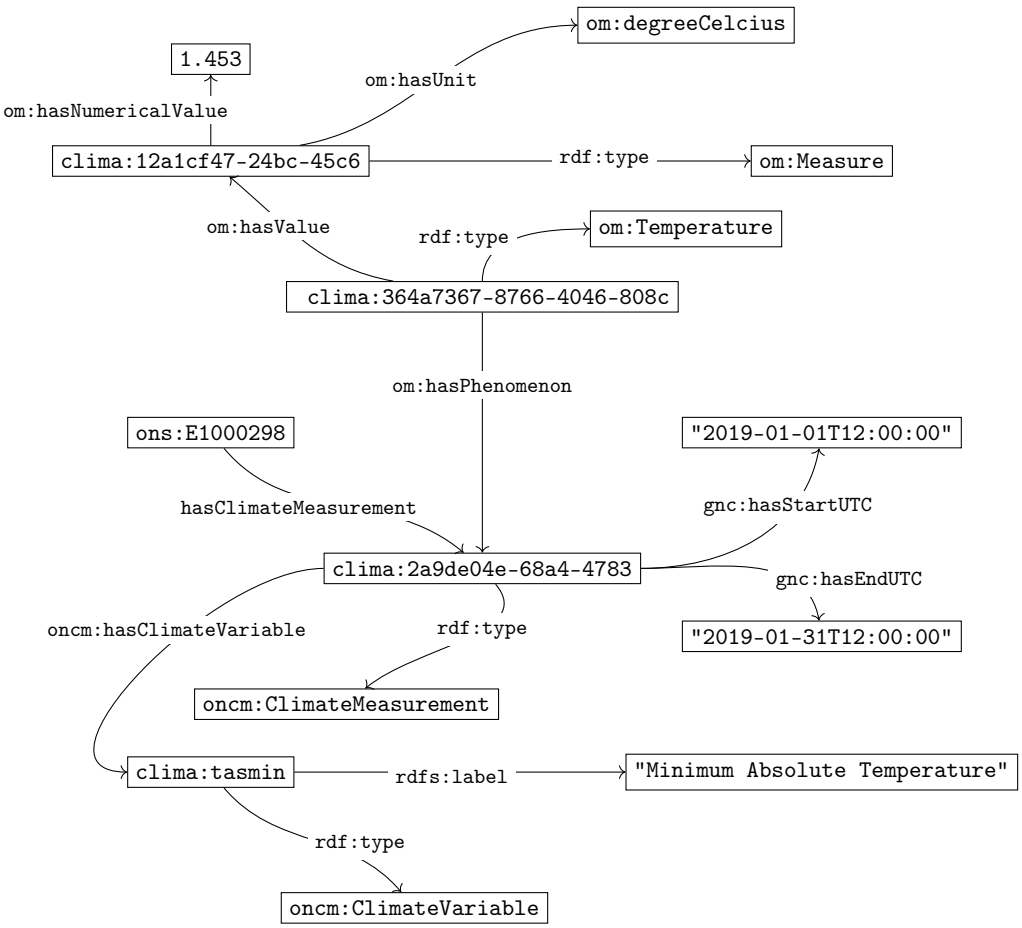



**Figure 8:** *An example set of triples produced by the agent responsible for the addition of HadUK Grid climate measurements to the knowledge graph. Specifically the set of triples describes a single climate variable, minimum absolute temperature, for a single statistical region, E1000298, within the month of January 2019.*

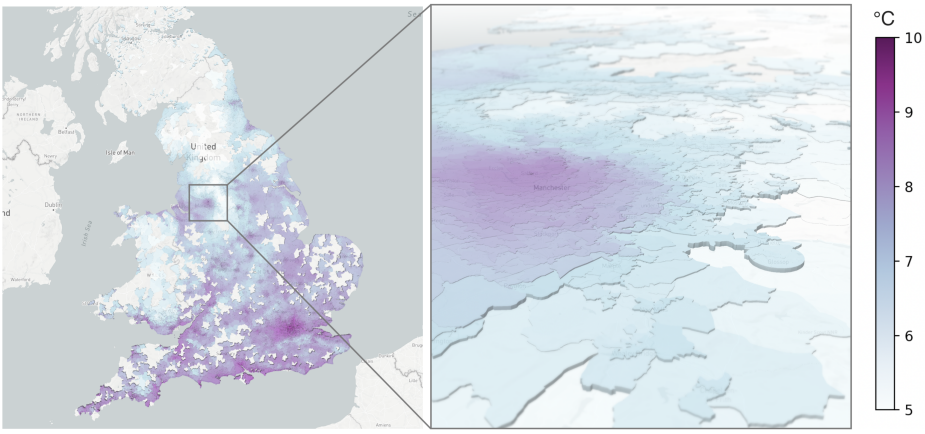
## 4.2 Instantiation of gas consumption statistics

An input agent was created to represent sub-national gas consumption statistics within the knowledge graph based on the respective annual spreadsheet file provided by the Department for Business, Energy & Industrial Strategy [13]. In an ideal world this data would be originally published as linked data, subverting many of the issues discussed in the introduction. However, here we make the conversion to linked data to facilitate addition to the knowledge graph. The agent links to existing instances of output areas and associates respective gas consumption values using vocabulary from OntoGasGrid and the Ontology of Units of Measure [49].

The agent constructs an additional 7 triples for each of the LSOA annual gas usage measurements, and an additional 5 triples to represent the number of consuming and non-consuming gas meters per LSOA region (not presented above). The agent is therefore responsible for the addition of 482,544 triples to the knowledge graph per annual data set.

## 4.3 Climate and gas consumption visualisation agent

An output agent was created to query information from the knowledge graph and render into a human-usable form *i.e.* a visualisation. **Fig.** 9 shows output from this agent.




(a) Mean temperature.


(b) Total monthly gas usage per LSOA.

**Figure 9:** *Geospatial data from the knowledge graph showing mean temperature and gas consumption in 2019, both displayed in the statistical regions defined by the Office of National Statistics [40]. The data is queried by an output agent. The resulting geoJSON is displayed in Mapbox.*

The agent interacts via a series of SPARQL queries, with WKT representations of output areas returned from the knowledge graph and subsequently parsed into geoJSON files containing values for mean temperature and gas consumption. These geoJSON files are in turn visualised using Mapbox [28] enabling interactive output directly from the knowledge graph.

## 4.4 Instantiation of UK gas transmission system

An input agent was created to instantiate the UK gas transmission system within the knowledge graph. The agent parses the grid pipeline shapefile published by the National Grid [36], containing

information such as pipe locations and respective diameters. Instances of pipe segments belonging to individual pipelines were created using the vocabulary defined in OntoGasGrid (Section 3.1.1).

A separate input agent was created to generate instances of `GridComponent` to describe infrastructure such as local distribution offtakes, power stations, and industrial users from information in [33]. The agent updates the knowledge graph by linking each new `GridComponent` to the closet instance of `GasPipeConnection` to express how these are connected, based on the assumption that this sufficiently approximates the physical connection. The `isConnectedToPipeline` property then provides a semantic link between these previously disjoint information. Below are example triples from the knowledge graph containing geospatial information, along with the connecting triple derived from these locations.

```
<gnsa:AberdeenToKirriemuirone855Connection, bd:lat-lon, "56.7349838#-2.726316081">
<oga:Careston, bd:lat-lon, "56.73503023#-2.726519436">
<oga:Careston, gnc:isConnectedToPipeline, gnsa:AberdeenToKirriemuirone855Connection>
```

## 4.5 Dynamic addition of live data feeds

An input agent was created in order to include dynamic data within the knowledge graph. The agent acts autonomously. The agent receives public information regarding instantaneous flow rates into the national transmission system and which is parsed into triples associated to instances of each gas terminal. A graphical example of triples generated by the agents is seen in **Fig.** 10.

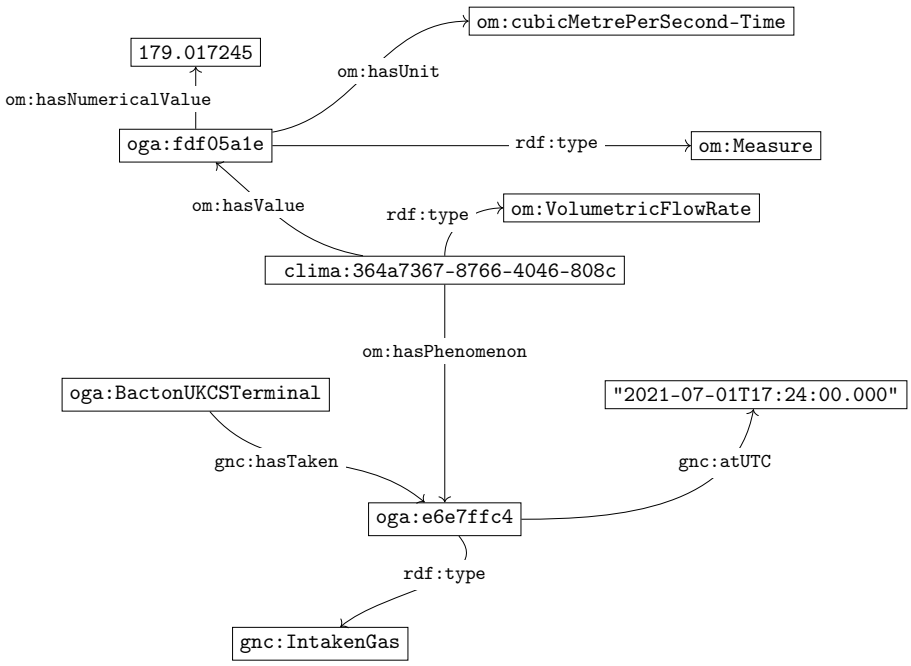



**Figure 10:** *Representation of instantaneous flow rates as linked data applying the ontology of units of measure. Here the instance of Bacton UKCS gas terminal is instantiated with the triples describing an instantaneous flow rate of* $179\ m^3 s^{-1}$ *at 2021-07-01T17:24:00 UTC, a value taken from the National Grid website by an input agent.*

The values of the intake gas associated with a single instance of time and gas terminal are expressed in cubic meters per second using the Ontology of Units of Measure [49] having been converted from published units of million-cubic meters per day. The agent is responsible for the addition of 9 additional triples per 2 minute flow rate measurement, per gas terminal, resulting in an additional 1620 triples per hour being instantiated within the knowledge graph.

## 4.6 UK gas grid visualisation

Three output agents were created in order to visualise the UK gas transmission system and connected infrastructure, as previously instantiated within the knowledge graph. These three agents perform SPARQL queries for the location and property relations of pipelines, offtakes, and gas terminals respectively.

**Query 1** demonstrates the SPARQL query used to return information about all local distribution offtakes. A subset of the results of this query are presented in **Table 2**.

**Query 1:** *SPARQL query to obtain local distribution offtakes and associated information.*

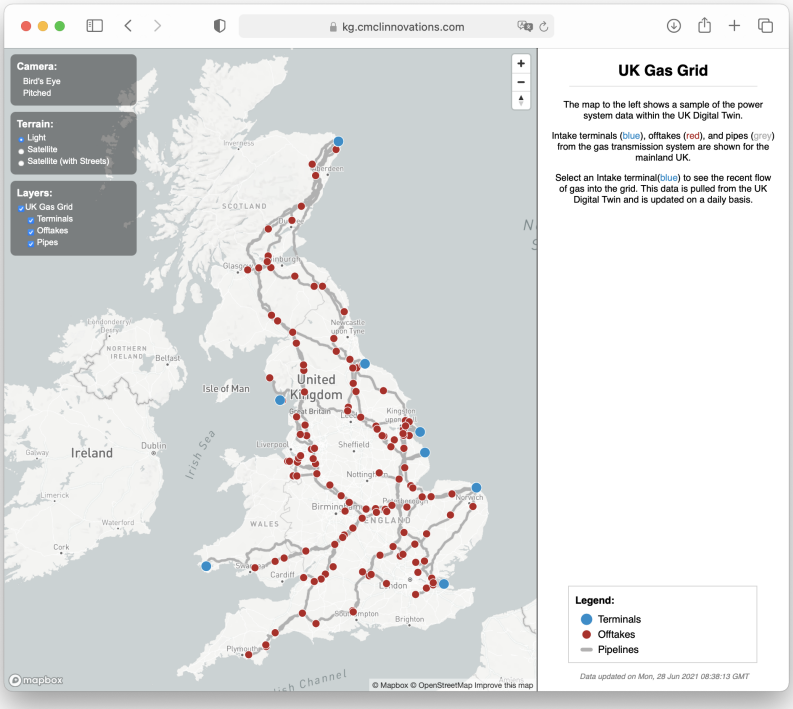
```
PREFIX rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>
PREFIX loc:   <http://www.bigdata.com/rdf/geospatial/literals/v1#>
PREFIX comp:  <http://www.theworldavatar.com/ontology/
               ontogasgrid/gas_network_components.owl#>

SELECT *
WHERE
{
    ?Offtake rdf:type comp:LocalDistribution.
    ?Offtake rdfs:label ?Label.
    ?Offtake loc:lat-lon ?Location.
    ?Offtake comp:hasLinepackZone ?LDZone.
    ?Offtake comp:hasNTSExitArea ?NTSArea.
    ?Offtake comp:hasNTSExitZone ?NTSZone.
}
```

**Table 2:** *Output from Query 1. This is subsequently encoded within a geoJSON file for interactive visualisation.*

| Offtake | Label | Location (Lat#Lon) | LDZone | NTSArea | NTSZone |
|---|---|---|---|---|---|
| http://www.theworldavatar.com/kb/ ontogasgrid/offtakes_abox/ThorntonCurtis | Thornton Curtis | 53.692#-0.282 | 3.0 | North | E11 |
| http://www.theworldavatar.com/kb/ ontogasgrid/offtakes_abox/Thrintoft | Thrintoft | 54.338#-1.484 | 3.0 | North | E03 |
| http://www.theworldavatar.com/kb/ ontogasgrid/offtakes_abox/Towlaw | Towlaw | 54.783#-1.890 | 1.0 | North | E01 |
| http://www.theworldavatar.com/kb/ ontogasgrid/offtakes_abox/Towton | Towton | 53.863#-1.305 | 3.0 | North | E03 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

The information from these queries is then reconstructed as a geoJSON [6] file that is visualised using Mapbox [28] as shown in **Fig. 11**. The creation of geoJSON files based on the knowledge graph representation of the gas grid can be performed periodically by these agents, should any information be updated within the knowledge graph (for example decommissioned pipelines or the publication of a new shapefile).



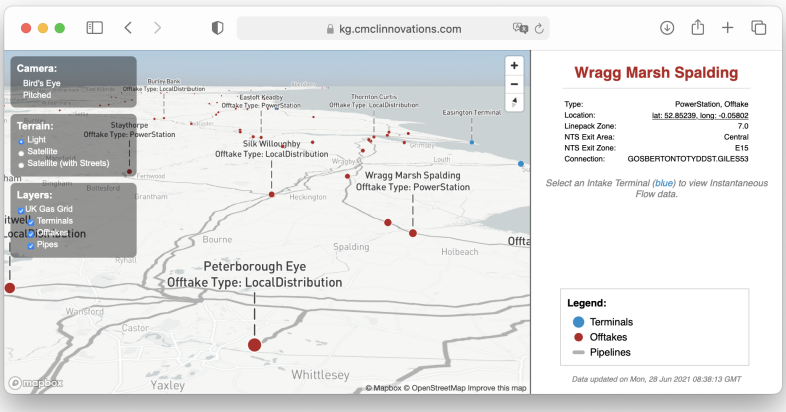




**Figure 11:** *Web-based interactive visualisation of the UK gas transmission system produced by agents operating on the knowledge graph. The panel on the right displays information about selected instances of physical infrastructure.*

An additional agent was created to query the last 24 hours of instantaneous flow data associated with each gas terminal, which in turn is provided to the visualisation as demonstrated in **Fig. 12**.

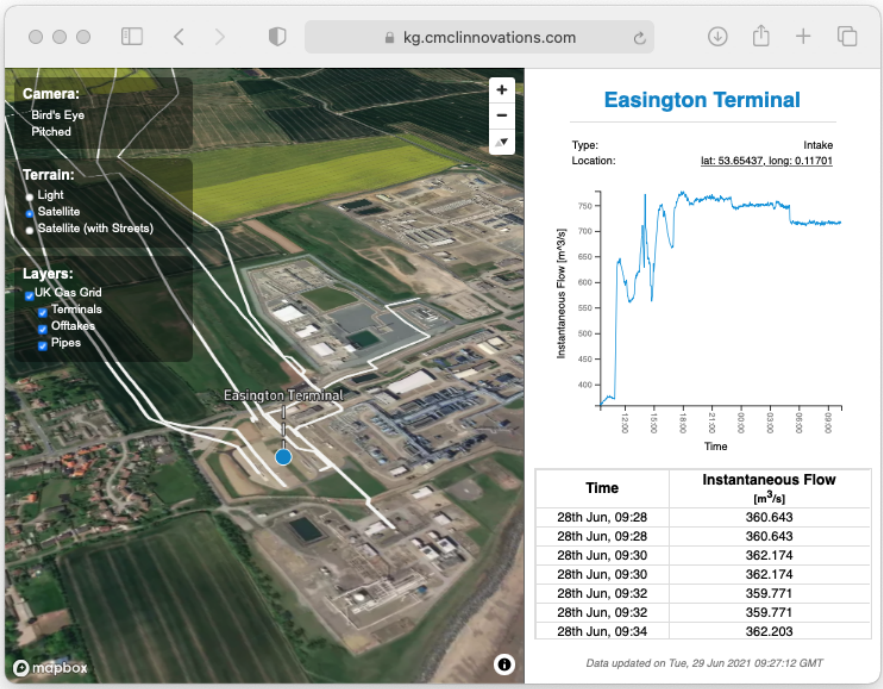



**Figure 12:** *Instantaneous gas flow rates are added to the knowledge graph by an input agent. The data are assigned to the corresponding instances of physical gas terminals and queried from the knowledge graph by output agents. Located at https://kg.cmclinnovations.com/explore/digital-twin/gas-grid.*

# 5 Conclusions

In this work we have extended the World Avatar to describe gas supply systems and climate data as part of an effort to create a national-scale digital twin of the UK.

Two new ontologies were created in order to represent these systems semantically. *OntoGasGrid* defines the vocabulary and respective relations to represent gas transmission systems and associated infrastructure. *OntoClimateMeasurements* allows for the representation of links between the existing concept of the output areas specified by the ONS, with new concepts to represent climate values.

The ontologies were used to extend the World Avatar dynamic knowledge graph to include data describing the UK gas transmission system, gas consumption statistics, real-time instantaneous intake of gas, in addition to data derived from the HadUK-Grid climate data set [24]. The extended knowledge graph includes links between the data and the geospatial output areas used by the ONS to report governmental data throughout the UK, for the first time formally linking these regions to the HadUK-Grid climate data.

A series of input agents were developed to incorporate data into the dynamic knowledge graph such that it remains current in time. The agents demonstrate both the addition of static data describing the physical infrastructure of the gas transmissions system, and the addition of live feeds of real-time data describing the intake of gas into the transmission system. Output agents were created to allow visualisation of geospatial and temporal information queried from the knowledge graph.

The architecture of the World Avatar has been suggested to provide a suitable architecture for implementing a Universal Digital Twin [2]. This paper demonstrates the universality of the approach both in terms of the range of geospatial and temporal data that can be semantically represented and linked in the knowledge graph, and the ability of agents to incorporate new data, process the data and interact with the real-world. The ability of such a Universal Digital Twin to link previously disjoint geospatial and temporal data sets enables increased interpretability across domains, offering a means to simplify analyses that previously would have required a bespoke and time-consuming solution that may be prone to errors. Future work will demonstrate this in analyses of future energy scenarios that combine the HadUK-Grid climate data with administrative data including energy consumption and social indicators such as fuel poverty.

# Nomenclature

**DL** Description Logic

**FAIR** Findable, Accessible, Interoperable, Reusable

**IRI** Internationalized Resource Identifier

**JSON** Javascript Object Notation

**LSOA** Local Super Output Area

**netCDF** Network Common Data Form

**NTS** National Transmission System

**ONS** Office for National Statistics

**WKT** Well-Known Text

**gnc** http://www.theworldavatar.com/ontology/ontogasgrid/gas_network_components.owl#

**gnsa** http://www.theworldavatar.com/kb/ontogasgrid/gas_network_system/

**gns** http://www.theworldavatar.com/ontology/ontogasgrid/gas_network_system.owl#

**gsp** http://www.opengis.net/ont/geosparql#

**oga** http://www.theworldavatar.com/kb/ontogasgrid/offtakes_abox/

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

**oncm** http://www.theworldavatar.com/ontology/ontogasgrid/ontoclimate.owl#

**onsa** http://statistics.data.gov.uk/id/statistical-geography/

**onst** http://statistics.data.gov.uk/def/statistical-geography#

**pmd** http://publishmydata.com/def/ontology/foi/

# Research data

Research data supporting this publication is available in the University of Cambridge data repository (doi:10.17863/CAM.72550).

# 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. Thomas Savage gratefully acknowledges the support of the John Dudding Studentship awarded by Fitzwilliam College. Markus Kraft gratefully acknowledges the support of the Alexander von Humboldt foundation.

# A Appendix

## A.1 Description logic (DL) representation of ontologies

### A.1.1 OntoGasGrid

Bold text names denote concepts that build on existing concepts from other ontologies. The full ontology, including definition of all namespaces and references to other ontologies, is provided as part of the research data supporting this publication. See the University of Cambridge data repository (doi:10.17863/CAM.72550).

$$
\begin{aligned}
\textbf{GridPipelineSegment} &\sqsubseteq \text{NetworkSystem} \\
\text{NetworkSystem} &\sqsubseteq \text{CompositeSystem} \\
\text{CompositeSystem} &\sqsubseteq \top \\
\textbf{Gas} &\sqsubseteq \top \\
\textbf{IntakenGas} &\sqsubseteq \textbf{Gas} \\
\textbf{OfftakenGas} &\sqsubseteq \textbf{Gas} \\
\textbf{StoredGas} &\sqsubseteq \textbf{Gas} \\
\textbf{StoredGas} &\sqsubseteq \textbf{Gas} \\
\textbf{LongTermStoredGas} &\sqsubseteq \textbf{StoredGas} \\
\textbf{MediumTermStoredGas} &\sqsubseteq \textbf{StoredGas} \\
\textbf{ShortTermStoredGas} &\sqsubseteq \textbf{StoredGas} \\
\textbf{GridPipeline} &\sqsubseteq \text{DirectedConnection} \\
\text{DirectedConnection} &\sqsubseteq \text{Connection} \\
\text{Connection} &\sqsubseteq \text{System} \\
\textbf{GridComponent} &\sqsubseteq \text{Device} \\
\text{Device} &\sqsubseteq \text{System} \\
\text{System} &\sqsubseteq \top \\
\textbf{GasPipelineStart} &\sqsubseteq \text{DirectedConnection} \\
\textbf{GasPipelineEnd} &\sqsubseteq \text{DirectedConnection} \\
\textbf{GasPipelineTube} &\sqsubseteq \text{Device} \\
\textbf{GasPipeConnection} &\sqsubseteq \text{Device} \\
\textbf{GridComponent} &\sqsubseteq \text{Device} \\
\textbf{CompressionStation} &\sqsubseteq \textbf{GridComponent} \\
\textbf{Intake} &\sqsubseteq \textbf{GridComponent} \\
\textbf{Offtake} &\sqsubseteq \textbf{GridComponent} \\
\textbf{Storage} &\sqsubseteq \textbf{GridComponent} \\
\textbf{ContinentalPipeline} &\sqsubseteq \textbf{Intake} \\
\textbf{GasTerminal} &\sqsubseteq \textbf{Intake} \\
\textbf{LiquefiedNaturalGasImport} &\sqsubseteq \textbf{Intake} \\
\textbf{Export} &\sqsubseteq \textbf{Offtake}
\end{aligned}
$$

$$\mathbf{IndustrialUser} \sqsubseteq \mathbf{Offtake}$$
$$\mathbf{LocalDistribution} \sqsubseteq \mathbf{Offtake}$$
$$\mathbf{PowerStation} \sqsubseteq \mathbf{Offtake}$$
$$\mathbf{CavernStorage} \sqsubseteq \mathbf{Storage}$$
$$\mathbf{HighPressureStorage} \sqsubseteq \mathbf{Storage}$$
$$\mathbf{LiquiefiedNaturalGasStorage} \sqsubseteq \mathbf{Storage}$$
$$\exists \mathbf{atUTC}.\top \sqsubseteq \top$$
$$\top \sqsubseteq \forall \mathbf{atUTC}.\mathrm{Datetime}$$
$$\exists \mathbf{hasDiameter}.\top \sqsubseteq \mathbf{GasPipelineTube}$$
$$\top \sqsubseteq \forall \mathbf{hasDiameter}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasEndUTC}.\top \sqsubseteq \top$$
$$\top \sqsubseteq \forall \mathbf{hasEndUTC}.\mathrm{Datetime}$$
$$\exists \mathbf{hasName}.\top \sqsubseteq \top$$
$$\top \sqsubseteq \forall \mathbf{hasName}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasObjectId}.\top \sqsubseteq \mathbf{GridPipeline}$$
$$\top \sqsubseteq \forall \mathbf{hasObjectId}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasOrder}.\top \sqsubseteq \mathbf{GasPipeConnection}$$
$$\top \sqsubseteq \forall \mathbf{hasOrder}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasLatitude}.\top \sqsubseteq \top$$
$$\exists \mathbf{hasLongitude}.\top \sqsubseteq \top$$
$$\top \sqsubseteq \forall \mathbf{hasLatitude}.\mathrm{DatatypeString}$$
$$\top \sqsubseteq \forall \mathbf{hasLongitude}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasLinepackZone}.\top \sqsubseteq \mathbf{Offtake}$$
$$\top \sqsubseteq \forall \mathbf{hasLinepackZone}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasLocalDistributionZone}.\top \sqsubseteq \mathbf{LocalDistribution}$$
$$\top \sqsubseteq \forall \mathbf{hasLocalDistributionZone}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasNTSExitArea}.\top \sqsubseteq \mathbf{Offtake}$$
$$\top \sqsubseteq \forall \mathbf{hasNTSExitArea}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasNTSExitZone}.\top \sqsubseteq \mathbf{Offtake}$$
$$\top \sqsubseteq \forall \mathbf{hasNTSExitZone}.\mathrm{DatatypeString}$$
$$\exists \mathbf{hasStartPart}.\top \sqsubseteq \mathbf{GridPipelineSegment}$$
$$\top \sqsubseteq \forall \mathbf{hasStartPart}.\mathbf{GasPipelineStart}$$
$$\exists \mathbf{hasTubePart}.\top \sqsubseteq \mathbf{GridPipelineSegment}$$
$$\top \sqsubseteq \forall \mathbf{hasTubePart}.\mathbf{GasPipelineTube}$$
$$\exists \mathbf{hasEndPart}.\top \sqsubseteq \mathbf{GridPipelineSegment}$$
$$\top \sqsubseteq \forall \mathbf{hasEndPart}.\mathbf{GasPipelineEnd}$$
$$\exists \mathbf{entersPipeConnection}.\top \sqsubseteq \mathrm{DirectedConnection}$$
$$\exists \mathbf{entersSegmentPart}.\top \sqsubseteq \mathrm{DirectedConnection}$$
$$\exists \mathbf{entersSegmentPart}.\top \sqsubseteq \mathrm{DirectedArc}$$

$$\top \sqsubseteq \forall \textbf{entersSegmentPart}.\text{Device}$$
$$\top \sqsubseteq \forall \textbf{entersSegmentPart}.\text{Node}$$
$$\top \sqsubseteq \forall \textbf{entersPipeConnection}.\text{Device}$$
$$\top \sqsubseteq \forall \textbf{entersPipeConnection}.\text{Node}$$
$$\exists \textbf{hasConnectedComponent}.\top \sqsubseteq \textbf{GasPipeConnection}$$
$$\top \sqsubseteq \forall \textbf{hasConnectedComponent}.\textbf{GridComponent}$$
$$\exists \textbf{hasPipeConnectionOutput}.\top \sqsubseteq \text{Node}$$
$$\exists \textbf{hasPipeConnectionOutput}.\top \sqsubseteq \text{Device}$$
$$\top \sqsubseteq \forall \textbf{hasPipeConnectionOutput}.\text{DirectedConnection}$$
$$\exists \textbf{hasSegmentPartOutput}.\top \sqsubseteq \text{Device}$$
$$\top \sqsubseteq \forall \textbf{hasSegmentPartOutput}.\text{DirectedConnection}$$
$$\top \sqsubseteq \forall \textbf{hasSegmentPartOutput}.\text{DirectedArc}$$
$$\exists \textbf{hasStored}.\top \sqsubseteq \textbf{Storage}$$
$$\top \sqsubseteq \forall \textbf{hasStored}.\textbf{StoredGas}$$
$$\exists \textbf{hasTaken}.\top \sqsubseteq \textbf{Intake}$$
$$\top \sqsubseteq \forall \textbf{hasTaken}.\textbf{IntakenGas}$$
$$\exists \textbf{hasUsed}.\top \sqsubseteq \textbf{Offtake}$$
$$\top \sqsubseteq \forall \textbf{hasUsed}.\textbf{OfftakenGas}$$
$$\exists \textbf{isConnectedToPipeline}.\top \sqsubseteq \textbf{GridComponent}$$
$$\top \sqsubseteq \forall \textbf{isConnectedToPipeline}.\textbf{GasPipeConnection}$$
$$\textbf{GasMeters} \sqsubseteq \top$$
$$\exists \textbf{hasConsumingGasMeters}.\top \sqsubseteq \textbf{GasMeters}$$
$$\top \sqsubseteq \forall \textbf{hasConsumingGasMeters}.\text{DatatypeString}$$
$$\exists \textbf{hasNonConsumingGasMeters}.\top \sqsubseteq \textbf{GasMeters}$$
$$\top \sqsubseteq \forall \textbf{hasConsumingGasMeters}.\text{DatatypeString}$$

### A.1.2 OntoClimateObservations

Bold text names denote concepts that build on existing concepts from other ontologies. The full ontology, including definition of all namespaces and references to other ontologies, is provided as part of the research data supporting this publication. See the University of Cambridge data repository (doi:10.17863/CAM.72550).

$$\textbf{ClimateMeasurement} \sqsubseteq \top$$
$$\textbf{ClimateVariable} \sqsubseteq \top$$
$$\text{Statistical-Geography} \sqsubseteq \top$$
$$\exists \textbf{hasClimateMeasurement}.\top \sqsubseteq \text{Statistical-Geography}$$
$$\top \sqsubseteq \forall \textbf{hasClimateMeasurement}.\textbf{ClimateMeasurement}$$
$$\exists \textbf{hasClimateVariable}.\top \sqsubseteq \textbf{ClimateMeasurement}$$
$$\top \sqsubseteq \forall\ \textbf{hasClimateVariable}.\textbf{ClimateVariable}$$

# References


[1] J. Akroyd, Z. Harper, D. Soutar, F. Farazi, A. Bhave, S. Mosbach, and M. Kraft. Universal digital twin – land use, 2021. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/276/.

[2] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal Digital Twin – A Dynamic Knowledge Graph. *Data-Centric Engineering*, 2021. In press.

[3] 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? *Appl. Energy*, 298:117117, 2021. doi:10.1016/j.apenergy.2021.117117.

[4] J. Bai, R. Geeson, F. Farazi, S. Mosbach, J. Akroyd, E. J. Bringley, and M. Kraft. Automated calibration of a poly(oxymethylene) dimethyl ether oxidation mechanism using the knowledge graph technology. *J. Chem. Inf. Model*, 61(4):1701–1717, Apr. 2021. doi:10.1021/acs.jcim.0c01322.

[5] G. Bristowe and A. Smallbone. The key techno-economic and manufacturing drivers for reducing the cost of power-to-gas and a hydrogen-enabled energy system. *Hydrogen*, 2(3): 273–300, 2021. doi:10.3390/hydrogen2030015.

[6] H. Butler, M. Daly, A. Doyle, S. Gillies, T. Schaub, and T. Schaub. The GeoJSON Format. *ISOC*, Aug. 2016. doi:10.17487/RFC7946.

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

[8] S. Cockings, A. Harfoot, D. Martin, and D. Hornby. Maintaining existing zoning systems using automated zone-design techniques: Methods for creating the 2011 census output geographies for england and wales. *Environ. Plan A.*, 43(10):2399–2418, Oct. 2011. doi:10.1068/a43601.

[9] Committee on Climate Change. Hydrogen in a low-carbon economy. *Climate Change Committee Publication*, Nov. 2018. Available at https://www.theccc.org.uk/publication/hydrogen-in-a-low-carbon-economy/ (accessed Feb 2021).

[10] J. F. DeCarolis, K. Hunter, and S. Sreepathi. The case for repeatable analysis with energy economy optimization models. *Energy Econ.*, 34(6):1845–1853, Nov. 2012. doi:10.1016/j.eneco.2012.07.004.

[11] J. F. DeCarolis, P. Jaramillo, J. X. Johnson, D. L. McCollum, E. Trutnevyte, D. C. Daniels, G. Akın-Olçum, 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. Schivley, 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(12):2523–2526, Dec. 2020. doi:10.1016/j.joule.2020.11.002.

[12] E. C. Delmelle. Toward a more socially impactful geographical analysis. *Geogr Anal.*, 53 (1):148–156, June 2019. doi:10.1111/gean.12213.

[13] Department for Business, Energy & Industrial Strategy. Sub-national consumption statistics: methodology and guidance booklet, 2020. Available at https://www.gov.uk/government/publications/regional-energy-data-guidance-note (accessed May 2021).

[14] Department for Business, Energy & Industrial Strategy. Provisional UK greenhouse gas emissions national statistics 2020, 2021. Available at https://www.gov.uk/government/statistics/provisional-uk-greenhouse-gas-emissions-national-statistics-2020 (accessed Jul 2021).

[15] A. Devanand, G. Karmakar, N. Krdzavac, R. Rigo-Mariani, Y. F. Eddy, 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, Aug. 2020. doi:10.1016/j.egyai.2020.100008.

[16] A. Devanand, G. Karmakar, F. Farazi, N. Krdzavac, M. Q. Lim, E. Y. Foo, I. A. Karimi, and M. Kraft. ElChemo: A cross-domain interoperability in a chemical plant, 2021. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/265/.

[17] P. E. Dodds and W. McDowall. The future of the UK gas network. *Energy Policy*, 60: 305–316, Sept. 2013. doi:10.1016/j.enpol.2013.05.030.

[18] 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, Dec. 2019. doi:10.1016/j.compchemeng.2019.106586.

[19] A. Eibeck, A. Chadzynski, M. Q. Lim, K. Aditya, L. Ong, A. Devanand, G. Karmakar, S. Mosbach, R. Lau, I. A. Karimi, and et al. A parallel world framework for scenario analysis in knowledge graphs. *Data-Centric Engineering*, 1:e6, 2020. doi:10.1017/dce.2020.6.

[20] 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, Dec. 2019. doi:10.1021/acs.jcim.9b00960.

[21] F. Farazi, N. B. 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, June 2020. doi:10.1016/j.compchemeng.2020.106813.

[22] S. Gillies. Shapely: manipulation and analysis of geometric objects, 2007–. Available at https://github.com/Toblerity/Shapely (accessed Jun 2021).

[23] L. M. H. Hall and A. R. Buckley. A review of energy systems models in the UK: Prevalent usage and categorisation. *Appl. Energy*, 169:607–628, 2016. doi:10.1016/j.apenergy.2016.02.044.

[24] D. Hollis, M. McCarthy, M. Kendon, T. Legg, and I. Simpson. HadUK-grid—a new UK dataset of gridded climate observations. *Geosci. Data J.*, 6(2):151–159, Sept. 2019. doi:10.1002/gdj3.78.

[25] M. Jovanovik, T. Homburg, and M. Spasić. A GeoSPARQL Compliance Benchmark. *arXiv:2102.06139*, 2021. Available at https://arxiv.org/abs/2102.06139 (accessed Jul 2021).

[26] M. J. Kleinelanghorst, 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.

[27] 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.

[28] Mapbox. Mapbox GL JS, 2021. Available at https://github.com/mapbox/mapbox-gl-js (accessed Jul 2021).

[29] W. Marquardt, J. Morbach, A. Wiesner, and A. Yang. *OntoCAPE*. Springer Berlin Heidelberg, 2010. doi:10.1007/978-3-642-04655-1.

[30] Met Office. HadUK-Grid Frequently Asked Questions, 2021. Available at https://www.metoffice.gov.uk/research/climate/maps-and-data/data/haduk-grid/faq (accessed May 2021).

[31] J. Morbach, A. Yang, and W. Marquardt. OntoCAPE—a large-scale ontology for chemical process engineering. *Eng. Appl. Artif. Intell.*, 20(2):147–161, Mar. 2007. doi:10.1016/j.engappai.2006.06.010.

[32] S. Mosbach, A. Menon, F. Farazi, N. Krdzavac, X. Zhou, J. Akroyd, and M. Kraft. Multiscale cross-domain thermochemical knowledge-graph. *J. Chem. Inf. Model*, 60(12):6155–6166, Nov. 2020. doi:10.1021/acs.jcim.0c01145.

[33] National Grid. NTS Exit Capacity Release Methodology Statement, 2014. Available at https://www.nationalgrid.com/uk/gas-transmission/document/66126/download (accessed Jun 2021).

[34] National Grid. History of gas transmission in Britain, 2021. Available at https://www.nationalgrid.com/group/about-us/our-history/history-gas-transmission-britain (accessed Jan 2021).

[35] National Grid. Instantaneous Flows into the NTS, 2021. Available at https://mip-prd-web.azurewebsites.net/InstantaneousView/Index (accessed Jun 2021).

[36] National Grid. Network route maps, 2021. Available at https://www.nationalgrid.com/uk/gas-transmission/land-and-assets/network-route-maps (accessed Jun 2021).

[37] N. F. Noy and D. L. Mcguinness. Ontology development 101: A guide to creating your first ontology. Technical report, Knowledge Systems Laboratory Stanford University, 2001. Available at https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (accessed Jul 2021).

[38] M. Office, D. Hollis, M. McCarthy, M. Kendon, T. Legg, and I. Simpson. HadUK-Grid gridded and regional average climate observations for the UK. *Centre for Environmental Data Analysis*, 2018.

[39] Office for National Statistics. Census geography , 2011. Available at https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography (accessed Jul 2021).

[40] Office for National Statistics. Geography Linked Data, 2019. Available at https://www.ons.gov.uk/methodology/geography/geographicalproducts/geographylinkeddata (accessed May 2021).

[41] Office for National Statistics. ONS Geography Linked Data Portal - Geography, 2019. Available at http://statistics.data.gov.uk/data/geography (accessed Jun 2021).

[42] 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. *Sustain. Cities Soc.*, 62:102412, 2020. doi:10.1016/j.scs.2020.102412.

[43] 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.

[44] 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. *Appl. Energy*, 175:305–323, 2016. doi:10.1016/j.apenergy.2016.05.019.

[45] M. Perry and D. Hollis. The generation of monthly gridded datasets for a range of climatic variables over the UK. *Int J Climatol*, 25(8):1041–1054, 2005. doi:10.1002/joc.1161.

[46] M. Perry, J. Herring, N. J. Car, T. Homburg, and S. J.D. Cox. OGC GeoSPARQL - A geographic query language for RDF data: GeoSPARQL 1.1 draft. *OGC implementation standard draft*, 2021. Available at https://opengeospatial.github.io/ogc-geosparql/ (accessed May 2021).

[47] 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.

[48] 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 Rev.*, 19:63–71, Jan. 2018. doi:10.1016/j.esr.2017.12.002.

[49] H. Rijgersberg, M. van Assem, and J. Top. Ontology of units of measure and related concepts. *Semantic Web*, 4(1):3–13, 2013. doi:10.3233/SW-2012-0069.

[50] J. Speirs, P. Balcombe, E. Johnson, J. Martin, N. Brandon, and A. Hawkes. A greener gas grid: What are the options? *Sustainable Gas Institute*, July 2017. Available at https://www.imperial.ac.uk/sustainable-gas-institute/research-themes/white-paper-series/white-paper-3-a-greener-gas-grid-what-are-the-options/ (accessed Feb 2021).

[51] A. von Richthofen, P. Herthogs, M. Kraft, and S. Cairns. Semantic city planning systems (SCPS): A literature review, 2021. ISSN 1473–4273. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/270/.

[52] W3C. Semantic Web, 2015. Available at https://www.w3.org/standards/semanticweb/ (accessed May 2021).

[53] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3(1): 160018, 2016. doi:10.1038/sdata.2016.18.

[54] S. G. Yalew, M. T. H. van Vliet, D. E. H. J. Gernaat, F. Ludwig, A. Miara, C. Park, E. Byers, E. D. 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(10):794–802, Aug. 2020. doi:10.1038/s41560-020-0664-z.

[55] C. Zhang, A. Romagnoli, L. Zhou, and M. Kraft. Knowledge management of eco-industrial park for efficient energy utilization through ontology-based approach. *Appl. Energy*, 204: 1412–1421, 2017. doi:10.1016/j.apenergy.2017.03.130.

[56] 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.