Universal Digital Twin – Land Use

Jethro Akroyd^{1,2}, Zachary Harper¹, David Soutar¹, Feroz Farazi^{1,2}, Amit Bhave³, Sebastian Mosbach^{1,2}, Markus Kraft^{1,2,4}

released: July 29, 2021

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

 ⁴ School of Chemical and Biomedical Engineering Nanyang Technological University 62 Nanyang Drive Singapore, 637459 ² CARES Cambridge Centre for Advanced Research and Education in Singapore 1 Create Way CREATE Tower, #05-05 Singapore, 138602

³ CMCL Innovations Sheraton House Cambridge CB3 0AX United Kingdom

Preprint No. 276



Keywords: Land use, crop map, ontology, knowledge graph, digitalisation, digital twin, geospatial search

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 develops an ontological description of land use and applies it to incorporate geospatial information describing land coverage into a knowledge-graph-based Universal Digital Twin. Sources of data relating to land use in the UK have been surveyed. The Crop Map of England (CROME) is produced annually by the UK Government and was identified as a valuable source of open data. Formal ontologies to represent land use and the geospatial data arising from such surveys have been developed. The ontologies have been deployed using a high-performance graph database. A customised vocabulary was developed to extend the geospatial capabilities of the graph database to support the CROME data. The integration of the CROME data into the Universal Digital Twin is demonstrated in a cross-domain use case that combines data about land use with a geospatial analysis of scenarios for energy provision. Opportunities for the extension and enrichment of the ontologies, and further development of the Universal Digital Twin are discussed.



Highlights

- Developed formal ontologies to represent geospatial land use data.
- Applied to Crop Map of England data published by UK Government.
- Developed customised vocabulary to enable geospatial queries.
- Land use data incorporated into knowledge-graph-based Universal Digital Twin.
- Cross-domain use case that performs geospatial analysis of energy provision.

Contents

1	Intr	Introduction		
2	2 Background			
	2.1	The W	Vorld Avatar	6
		2.1.1	Ontologies	7
		2.1.2	Linked Data and Knowledge Graphs	7
		2.1.3	Data Storage, Queries and Updates	8
		2.1.4	Geospatial Data	8
		2.1.5	A Universal Digital Twin	9
	2.2	Land U	Use and Biomass	9
		2.2.1	Biomass Availability Data	10
		2.2.2	Land Cover and Land Use Data	11
		2.2.3	The Crop Map of England	11
		2.2.4	Energy & Yield Data	14
		2.2.5	Existing Ontologies	15
3	Met	hodolog	<u>zy</u>	17
	3.1	Ontolo	bgy Development	17
		3.1.1	OntoLandUse	17
		3.1.2	OntoCropMapGML	19
		3.1.3	OntoCropEnergy	20
		3.1.4	Interconnection Between Ontologies	21
4	Use	Case		22
	4.1	Know	ledge Graph Deployment	22
	4.2	Examp	ple Geospatial Queries	23
	4.3	Elean	Power Station	25
5	Con	clusion		30

Nomenclature

Appendix		
A.1	Description Logic Representation of Ontologies	33
	A.1.1 OntoLandUse	33
	A.1.2 OntoCropMapGML	33
	A.1.3 OntoCropEnergy	34
A.2	Namespaces	36
A.3	Example Geospatial Queries	37
A.4	Queries Performed by the Elean Power Station Use Case	38
D.C		41
Kete	erences	41

31

1 Introduction

The population of the world is projected to increase to 9.7 billion by 2050, rising to a peak of 11 billion in 2100 [124]. This will lead to land use change as more people migrate to cities in search of a better quality of life [123]. Moreover, it is predicted that there will be an increase in energy demand and therefore increased potential for greenhouse gas emissions [134]. Increasing population is known to cause stress on the environment, including resource depletion, biodiversity loss, and deforestation [85].

The global average temperature has risen by over 1°C since 1880 [111]. In December 2015 in Paris, 197 countries pledged to aggressively curb their greenhouse gas emissions and work together to limit the increase in global temperature to 2°C by the end of the 21^{st} century [125]. In 2019, the UK became the first major economy to pass a law to cut its emissions to net zero by the year 2050 [58]. More recently, the UK Government has pledged to 'build back greener' from the Covid-19 pandemic [33], and to cut emissions by 78% by 2035 relative to 1990 levels [122].

The increasing global demand for energy coupled with the drive of society to cut emissions is known as the dual challenge [16]. This presents a formidable problem because the two goals are at odds with one another. It is well understood that solving the dual challenge will involve the widespread deployment of renewable technologies including, solar photovoltaics, wind turbines and bioenergy [23]. The deployment of these technologies has significant implications for land use and cannot be considered in isolation from the impact on food production and biodiversity [10, 75].

Over the last decade the cost of solar photovoltaic systems has fallen by more than 74%, owing primarily to government policies supporting research and development [67]. At present, the efficiency of single-junction solar cells in converting solar radiation to electricity is confirmed to be as high as 29.1% [59]. Recent advances in the fabrication of multi-junction photocells has led to efficiencies as high as 39.2% being observed [59]. With sustained research into novel materials and fabrication methods, the efficiency associated with solar photovoltaic systems is projected to increase further [86]. Despite the significant cost reductions and efficiency increases, solar photovoltaics only generated 4% of the UK's electricity in 2019 [28]. Like many renewables [18], solar photovoltaics suffer from intermittency and rarely operate at full capacity [107]. This means that solar photovoltaic electricity generation is variable such that solar photovoltaic systems on their own are not suitable to provide baseload electricity requirements.

The UK is regarded to have the best wind resources in Europe [34]. In 2019 wind power was responsible for generating almost 20% of the UK's electricity, with approximately equal contributions from onshore and offshore wind [28]. Costs associated with onshore and offshore wind have fallen by 40% and 29% respectively over the last decade [68]. These cost reductions have resulted in onshore wind becoming the most cost-effective technology amongst all electricity generation methods in the UK [96]. However, until recently, the growth of onshore wind power has been stifled by restrictions on new projects [8]. This resistance to onshore developments coupled with the favourable wind conditions at sea has led to the UK becoming the world leader in offshore wind capac-

ity [108]. Dogger Bank Wind Farm [38] is being developed in the North Sea by SSE Renewables [106], Equinor [48] and ENI [47]. On completion, it will be the largest off-shore wind farm in the world, capable of producing 3.6 GWe.

The use of biomass to generate electricity is proposed by the Committee on Climate Change as being crucial in helping the UK to achieve net zero [23]. In fact, the use of biomass to generate electricity is one of the few renewable energy sources which could be used to supply the UK's baseload electricity requirements [81]. Biomass generated 11.5% of the UK's electricity in 2019 [28]. The largest single user of biomass in the UK is Drax power station [40, 63], which has an installed capacity of 2.6 GWe for biomass and 1.3 GWe for coal [27] and which supplies 5% of the UK's electricity [39]. The coupling of Bioenergy with Carbon Capture and Storage, known as BECCS, is a net negative technology that results in the removal of carbon dioxide from the atmosphere. The retrofitting of CCS technologies on bioenergy plants comes at a cost, reducing the plant efficiency between 6 and 15% [13] as energy is required to capture the carbon dioxide. The Committee on Climate Change [24] recommended that UK Government policies support the deployment of BECCS technologies to help realise their climate goals [22]. However, bioenergy is not a universal solution and is not without its controversies. One common example is the 'Food vs. Fuel' debate relating to the diversion of land used for the cultivation of food crops to bioenergy crops [110]. Photosynthesis in plants has a maximum efficiency of 2% [75]. Consequently, large areas of land are required to grow biomass to generate sufficient electricity and this can lead to a loss in biodiversity [62].

Care must be taken when committing land to deploy these technologies to ensure that they complement one another and contribute to solving the problem holistically. This is a complex challenge and the ability to explore and assess different options can be aided through digitilisation [64]. Dynamic knowledge graph technology developed as part of the World Avatar project [44, 117] has been shown to provide an architecture that is suitable for implementing a Universal Digital Twin that can address this type of challenge [1, 25]. In this approach, autonomous computational agents interact to perform tasks including updating the knowledge graph to ensure that it remains current in time, simulating systems and sending responses back to the physical world. It has been shown how this architecture can be used to implement digital twins that provide information about the state of the world, provide intelligent control using computational agents to model the behaviour of complex systems and provide support for intelligent design via what-if scenario analysis [1, 45]. The World Avatar approach takes advantage of Semantic Web technologies [12, 132] to support cross-domain interoperability of models and data. These technologies ensure that data are connected, portable, discoverable, and queryable via a uniform interface.

The **purpose of this paper** is to create an ontological description of land use, and to apply the ontologies to provide a geospatial description of the land use in England as part of a knowledge-graph-based Universal Digital Twin. The paper is structured as follows. Section 2 provides an overview of the relevant technologies and surveys ontologies and data sets relating to land use. Section 3 describes the development of ontologies to represent land use within a knowledge graph. Section 4 presents a cross-domain use case that combines data about land use in the digital twin with a geospatial analysis of scenarios for energy provision. Conclusions are drawn in Section 5.

2 Background

2.1 The World Avatar

The World Avatar project aims to create a dynamic world model that is generic and allencompassing, with a focus so far geared towards the decarbonisation of energy and the process industries [6, 36, 37, 72, 93, 94, 138], city planning [21, 126] and chemistry [50, 73, 84, 141].

The world model is implemented as a dynamic knowledge graph that is built using Semantic Web technologies. The dynamic knowledge graph combines an ontological description of the concepts and instances (*i.e.* data) that are known to the world model with automated computational agents that operate on the knowledge graph. The computational agents are described as part of the knowledge graph and can perform actions on both concepts and instances. The design concept is illustrated in **fig. 1**.



Figure 1: The design of the World Avatar dynamic knowledge graph. Image reproduced from [1] under a CC BY 4.0 licence.

The design of the World Avatar is intended to confer versatility by allowing the agents acting on the knowledge graph to perform a number of different types of task, including:

- Implementing data pipelines to ensure that the knowledge graph remains current in time, for example updating instances in the knowledge graph with data feeds from sensors or smart infrastructure.
- Sending signals back to the real world, for example to control an actuator.
- Providing an interface to computational models to calculate quantities, for example estimates of air quality [52], and updating the instances in the knowledge graph.
- Restructuring the knowledge graph by adding instances, for example to explore the consequences of design choices [35, 45], or by adding concepts and relationships between concepts and instances, for example using ontology matching to improve the coverage of the knowledge graph.

• Discover and compose new agents simply by reading from and writing to the knowledge graph so as to combine existing agents to form composite agents to perform more complex tasks [139].

The Semantic Web technologies that are fundamental to the World Avatar are summarised below. The application of these technologies to facilitate creating a Universal Digital Twin, as part of the World Avatar project, is also discussed.

2.1.1 Ontologies

Ontologies are fundamental to the Semantic Web. An ontology, in this context, is a semantic model created using classes (also known as concepts), object properties and data properties to represent information about a domain of interest. Object properties link an instance of a class (the domain of the property) to an instance of a class (the range of the property); data properties link an instance of a class (the domain of the property) to a data element (the range of the property).

The classes in an ontology may be arranged to form a hierarchy, where concepts belonging to a class can be specified as sub-classes of a common parent class. Similarly, object properties and data properties may form a hierarchy of properties and sub-properties. An example could be '*CoalPowerStation*', '*GasFiredPowerStation*' and '*BiomassPowerStation*', which could all be considered as sub-classes of '*PowerStation*'. The properties of the parent class will also apply to all sub-classes.

When discussing ontologies, it is common to refer to a Terminological Component (TBox) and an Assertion Component (ABox). The TBox specifies the classes, object and data properties that can exist in the ontology. The ABox defines instances of classes, relationships with other instances (through object properties) and data associated with instances (through data properties). In the above example of power plants, the TBox would define the existence of concepts and the associated properties for different types of power plant, whereas the ABox would define the instances of the power plants, for example all of the power plants in the UK.

A number of formats exist to represent ontologies including the Resource Description Framework (RDF) [2], Turtle [131] and Web Ontology Language (OWL) [128]. OWL was developed to enable the representation of ontologies with complex logical formulae to meet the data publishing requirements of different domains [9]. It is possible to verify the consistency of an ontology represented in OWL (or other formats) and infer indirect subclass-of relations using reasoners like HermiT [26]. The interested reader is referred to Allemang and Hendler [3] for more detail.

2.1.2 Linked Data and Knowledge Graphs

Linked Data [11] refers to the idea of linking Semantic Web data. Linked Data uses the logical and semantic capability of RDF to represent instances, classes and links. The

links take the form of internationalised resource identifiers (IRIs) and play a pivotal role in enabling the discovery of Linked Data.

Knowledge graphs express Linked Data as a directed graph, where the nodes of the graph are the concepts and instances, and the edges of the graph are the links between related concepts and instances. Typically, the number instances would far exceed the number of concepts. The ontological basis of knowledge graphs is such that reasoners, for example HermiT [26], can be used to infer insights from existing facts. Linked Data and knowledge graphs offer a useful approach to storing information because they can be navigated to find related data and can be created with an open license to provide a collective, readily accessible knowledge base.

2.1.3 Data Storage, Queries and Updates

RDF data (and therefore knowledge graphs) can be hosted in graph databases, often referred to as *RDF stores* or *triple stores*, that store RDF statements in 'subject', 'predicate' and 'object' columns. (In fact, many triple stores are actually *quad stores* and store an additional 'context' column that can be used to provide additional information about a statement). The data hosted in a store can be queried and updated via endpoints identified by IRIs. Operations can be executed over multiple endpoints by employing federated queries. SPARQL updates and queries can be carried out for individual triple stores through their own Application Programming Interfaces (APIs).

SPARQL Protocol and RDF Query Language (SPARQL) [129] is a query language designed to query semantic information, for example querying instances of an ontological class, querying instances that are connected via links of interest or querying data linked to an instance. SPARQL Update [130] is an update language designed to insert and delete statements from a triple store. It allows statements to be added to a triple store, or copied or moved from one triple store to another. Though SPARQL query and update operations can be performed on individual triple stores through their own APIs, the Jena-JDBC API [70] provides a scalable high-performance triple-store-agnostic means to perform SPARQL queries and updates.

2.1.4 Geospatial Data

Several best practices for the RDF encoding of geospatial data have been published, including the GeoSPARQL [89] standards developed by the Open Geospatial Consortium (OGC) and guidelines from the Infrastructure for Spatial Information in Europe (INSPIRE) [5]. Ontologies that provide definitions for Geography Markup Language (GML) [90] properties and objects also exist, for example the Ontology for Geography Markup Language (GML3.0) [41].

GeoSPARQL extends the SPARQL query language to support the querying of and reasoning about geospatial information. However, the GeoSPARQL support offered by different triple stores remains limited and inconsistent [21, 71], for example RDF4J [42] offers '*partial GeoSPARQL support*' [43] whilst Blazegraph [15], which is used in this work, does not support GeoSPARQL, but offers simple geospatial reasoning via a custom interface.

2.1.5 A Universal Digital Twin

The World Avatar uses a dynamic knowledge graph to instantiate semantic models of the domains of interest and of computational agents. By doing so, it aims to link information to create knowledge repositories on the World Wide Web through the standards laid out by the World Wide Web Consortium (W3C), resulting in a delocalised system of organised knowledge that is machine-readable and where entities are identified by unique IRIs [127]. Unlike typical databases, the World Avatar contains an ecosystem of autonomous computational agents that continuously update it, where the semantic annotation of the data and agents seeks to enable cross-domain interoperability [44].

The design of the World Avatar has been demonstrated as one way to implement a comprehensive Universal Digital Twin, for example a digital twin of the UK [1]. The design confers a number of advantages which include the adoption of a distributed architecture that supports a uniform interface to query multiple data sources, the ability to use logical reasoning to verify the consistency of semantic models [37], and a mechanism that allows computational agents to interact to answer questions [139, 140]. The use of Linked Data helps address ambiguity, facilitates the discovery and reuse of information and enables the linking of related information, such as provenance.

The World Avatar, and the digital twins based on it, include the idea of a 'base world' that describes the real world and that is kept current in time by computational agents that input data from the real world into the knowledge graph and that simulate the behaviour of the world, and of 'parallel worlds' that support what-if scenario analysis [45]. The parallel worlds capability allows exploring consequences of alternative design and policy choices to support decision making in complex environments.

Work to develop a dynamic-knowledge-graph-based Universal Digital Twin of the UK [1, 25] is ongoing. The digital twin currently includes a description of all the power plants in the UK [6]. Work to establish data pipelines to describe buildings [21], the gas transmission network, geospatial climate data and the potential for solar and wind power in the UK is underway. The representation of biomass availability and land use within this digital twin will enrich the description of the UK provided by the base world. This will support parallel world analyses that consider the role of land use in supporting decarbonisation.

2.2 Land Use and Biomass

The following sections review data and ontologies that describe biomass availability, land cover and land use, and assesses the potential for their application in a digital twin of the UK. Data for the energy content and yield of crops are also surveyed.

2.2.1 Biomass Availability Data

Biomass is material derived from living, or recently living organisms [54]. The Forest Research agency defines five categories of biomass: virgin wood, energy crops, agricultural residues, food waste, and industrial waste and co-products [54]. A number of reports attempt to quantify the biomass resources available to the UK.

The Forestry Commission (FC) produces a National Inventory of Woodland and Trees [53] every 15 to 20 years, providing statistics by forest type, species and ownership. A mapped distribution of woodland that is over two hectares in area is included, however quantitative data are only reported regionally ('South West', 'South East' and so on). In 2003, the Forestry Contracting Association (FCA) and FC [14] provided statistics about woody biomass production with greater resolution, but it was restricted to Forest Districts that were predefined by the FC. Despite the quantity of data offered, the data are only updated infrequently, which does not align with the ambition of a dynamic digital twin.

In 2013 a model was published that sought to predict the future availability of lignocellulosic biomass in the UK to 2050 [61]. The model is based on variables that could in the future be provided via a digital twin, including soil composition and weather conditions as well as constraints on where biomass can be grown. The types of biomass considered in the model are miscanthus, short rotation coppice willow, short rotation coppice poplar, and short rotation forest poplar. Calculations were performed on land areas of 1 km² and the results reported by region. A subsequent report by the NNFCC consultancy also assessed the availability of lignocellulosic materials in the UK [77]. Data was presented for the availability of forest harvest residues, energy crops, agricultural straw residues and green wastes including paper. Although information was provided for the whole UK, the geographical resolution of data was again only by region.

In 2014 the National Forest Inventory produced a 50-year forecast for the availability of softwood timber in the UK [56]. The forecast considered the area of woodland, wood characteristics, growth rate and when trees are harvested. The data include the volume of wood available by region and whether the wood is located on public or private land. However, the report necessarily caveats that the forecast is subject to unpredictable external factors that have the potential to cause significant disparity from the forecast results.

Contemporary data about the quantities of different wastes and methods of disposal are available through annual government reports in the UK [30, 104, 137]. However, the data are only resolved to a regional level, for example 'South West', 'North West' and so on, so are not able to provide a geographically precise description of biomass availability. The Digest of Waste and Resource Statistics [29] also publishes data about waste streams, including some biomass, and waste treatment facilities in the UK. The Waste and Resources Action Programme (WRAP) charity reported bulk figures of household food and drink waste for the UK, separated into different categories and by possible end uses [133].

The data surveyed in this section are limited by their geospatial resolution and accuracy, and are mainly provided via text-based documents which reduces the accessibility of the data. They will not be pursued further in this iteration of the digital twin.

2.2.2 Land Cover and Land Use Data

Land Cover is the observed cover of the surface of the earth [60], whereas Land Use is the socio-economic function of the land [49]. The UK Government publishes annual data about land use in England [82]. Developed land use categories include residential, transport and industry, while non-developed land uses include agriculture, forestry and undeveloped land. Data for the percentage of developed and non-developed land is provided by region, and higher resolution data is available on request for smaller land areas. The land use is classified using products from Ordnance Survey [92], and the resulting data are subject to quality assurance tests so are of known quality. However, the data are published in the form of infographics which are not readily machine-readable.

The UK Centre for Ecology & Hydrology produces a map of land cover across the UK [121]. Crops classified in this map include, field beans, grass, maize, oilseed rape, potatoes, barley (spring & winter) and wheat (spring & winter). The data is produced annually using satellite data from Sentinel-2 [113] and is verified using land declarations produced by farmers. The data has high geospatial resolution and is reported in the form of the land cover associated with individual fields. The geospatial boundaries of the fields across the UK are obtained from the 2007 Ordnance Survey Mastermap[®] data [92], and although this is historic data, the likelihood of field boundaries changing is considered unlikely [121]. However, the boundaries of the fields are highly irregular, such that this is a detailed but complex data set. In addition, the only data that is available for free for educational use dates from 2015. This data is a valuable resource, but the restrictions on accessing it limit our ability to use it in a digital twin.

The UK Government also publishes an annual Crop Map of England [101]. This turns out to be a very useful source of data and is discussed in detail in the next section.

2.2.3 The Crop Map of England

The Crop Map of England (CROME) [101] describes land use in England. The data are published annually on behalf of the UK Government by the Rural Payments Agency (RPA). The land use is classified using a combination of data from the Sentinel-1 (radar images) [112] and Sentinel-2 (optical images) [113] satellites, and ground truth data from land declarations submitted by farmers, woodland owners, foresters and land managers when applying for Basic Payments Scheme [99] and Countryside Stewardship [100] grants. Automatic image classification is performed using a supervised Random Forest [17] machine learning algorithm. The algorithm learns by associating satellite images with the ground truth data. The accuracy of the classification technique has been estimated as 95.4% [101] based on a comparison of ground truth data versus the corresponding Random Forest classification (sample size, n = 4883). The resulting data are published on a regular grid consisting of approximately 32 million hexagonal cells, each with an area of 4156 m². Figure 2 shows a sample of the 2019 data.



Figure 2: A sample of the Crop Map of England (CROME) 2019 data [101]. The data are superimposed on an OpenStreetMap [91] monochrome layer using QGIS [95].

The land use for each hexagonal cell is described by means of a Land Use Code (LUCode). A LUCode is an alphanumeric code up to 5 characters long. There are 81 LUCodes used in the CROME data set [101], however, more LUCodes exist and the RPA publishes a definitive list [102]. The LUCodes belong to land cover categories of *Cereal Crops*, *Leguminous Crops, Energy Crops, Grassland, Non-Agricultural Land, Water, Trees*, and *Unknown Vegetation or Mixed Vegetation*. Example LUCodes and their associated land cover categories and descriptions are given in **Table 1**.

The CROME data is available for download [101] in the form of 46 files covering different regions of England. The files are available in Geography Markup Language (GML) [90] and Geospatial JavaScript Object Notation (GeoJSON) [69] format. The GML format data is 30.1 GB in size, whilst while the GeoJSON format data set is smaller at 16.0 GB, where the difference is a result of the format as opposed to the content. A description of the schema used in the CROME 2019 data set is given in **Table 2**.

Table 1: Example Land Use Codes (LUCodes), their respective land cover and land use descriptions used by the CROME 2019 data set.

Land Cover Description	LUCode	Land Use Description
Cereal Crop	AC100	Italian Ryegrass
Leguminous Crop	LG14	Clover
Energy Crop	SR01	Short Rotation Coppice
Grassland	PG01	Grass
Non-Agricultural Land	NA01	Non-vegetated or sparsely-vegetated Land
Water	WA01	Water
Trees	TC01	Perennial Crops and Isolated Trees
Unknown Vegetation Or Mixed Vegetation	AC00	Unknown or Mixed Vegetation

Table 2: Description of the schema used by the CROME 2019 data set.

Envelope Type			
Property	Description		
srsName	EPSG:27700 (coordinate system).		
srsDimension	2 (2D coordinate system).		
lowerCorner	Extreme south-westerly coordinate of the envelope.		
upperCorner	Extreme north-easterly coordinate of the envelope.		
Ce	ll Features		
Property	Description		
Crop_Map_of_England_2019_EnvelopeName	<i>EnvelopeName</i> is replaced by the name of the region to which the cell belongs, for example <i>Cambridgeshire</i> .		
OBJECTID	Cells are numbered consecutively for a given envelope.		
CROMEID	An identifier of the form: RPAxxxxxyyyyyyy. This is a unique key associated with a cell across all survey years, where xxxxyyyyyy identifies the centre point of the cell using the EPSG:27700 coordinate system.		
LUCode	Identifies the land use associated with the cell.		
RefDate	The date the land use classification was performed (yyyymmdd format).		
Shape_Length	The perimeter of the cell (m).		
Shape_Area	The area of the cell (m^2) .		
surfaceProperty	Geospatial description of the cell.		
Cell Geo	ospatial Features		
Property	Description		
srsName	EPSG:27700 (coordinate system).		
srsDimension	2 (2D coordinate system).		
posList	Series of pairs of coordinates to define a closed perime- ter (the first and last pair in the series are the same). In the CROME data, each list describes a hexagonal cell so contains 7 pairs of coordinates.		

The CROME data offers several advantages. The data are available under an open government license [115] and are published annually in a machine-readable format. It is further assumed that the data are likely to continue to be available in future years. However, there are also a few issues. Firstly, there are instances in which land use is misclassified when a hexagonal cell overlaps two different land use types. Secondly, although CROME is produced by the RPA, the LUCodes that appear in the CROME data [101] and the LUCodes published by the RPA [102] exhibit some minor differences. For example, CROME defines TC01 as 'perennial crops and isolated trees' [101], whereas the RPA defines TC01 as 'permanent crops other than nursery crops and short rotation coppice' [102]. Thirdly, it was found that there were no instances of LUCode SR01. This corresponds to 'short rotation coppice' which is an energy crop. This is surprising because short rotation coppice certainly exists, and the reason for the absence of SR01 remains unexplained. Finally, the CROME data only cover England, rather than the whole of the UK.

A further challenge relates to coordinate systems. CROME uses EPSG:27700 [79] (also known as OSGB36 or British National Grid), an easting-northing system commonly used in topographic mapping of the UK. However, the geospatial capability offered by Blaze-graph [15] requires EPSG:4326 [80] (latitude-longitude, also known as WSG84), a standard system for satellite navigation and GPS. This means that the process of ontologising the CROME data must also include a coordinate transformation if the resulting digital twin is to use the geospatial capability of Blazegraph.

2.2.4 Energy & Yield Data

This section surveys data that can be used to estimate the power associated with different types of biomass to support a use case relating to electricity generation. Data relating to the power available from crops are typically expressed in the form of the power per unit area of land (W/m^2) [75]. However, energy content and yield data that can be used to derive the power per unit area are desired because this will provide a broader scope for the ontology. In practice, the energy content and yield of a crop will of course vary spatially and temporally due to different environmental conditions and farming techniques. It is not attempted to account for these variations at this iteration of the digital twin. **Table 3** summarises the surveyed data and evaluates its suitability for inclusion in an ontology that can be used to support estimates of the power available from biomass.

Database	Information Provided	Evaluation of Use
Phyllis2 [119]	Online database describing physico- chemical properties of biomass with over 3000 data entries.	Useful resource, multiple references for each biomass type and both gross- and net- calorific value data available.
Farming Statistics (UK) [31]	Annual yield data for 5 classes of crop in the UK.	Government resource updated annually, limited number of crop yields available.
FAOSTAT [114]	Online database for annual crop yields in many countries.	United Nations data for many countries (including the UK) over multiple decades.

Table 3: Comparison of surveyed data resources for energy content and yield of crops.

Phyllis2 [119] is an online database for the physico-composition of biomass. It contains over 3000 data entries and is updated and extended regularly. The entries include information about the gross- and net-calorific value associated with each biomass type, including distinguishing between different forms of biomass type, including, 'As Received', 'Dry' and 'Dry and Ash Free'. The database contains information relating to 17 crops that appear in the CROME data set. The data can be downloaded in CSV format.

The UK Government publishes farming statistics that include yields of common crops [31], including wheat, barley, oilseed, oats, and minor cereals (rye, maize, triticale). There is overlap with 11 crops in the CROME data set. The data is updated annually and is specifically concerned with the UK. However, the data is published in PDF format so additional work would be required to make it machine-readable. Two further sources were found to extend the yield data to include miscanthus [55] and sunflower [136].

The Food and Agriculture Organization of the United Nations (FAO) has created an online FAOSTAT database [114] that includes data about the yields of crops in many countries. There is overlap with 30 crops in the CROME data set. The database is updated annually and includes historic data going back to 1961. The data can be downloaded in CSV format. The availability of data for many countries is likely to be of further value in the future to describe the yield of crops grown outside of the UK.

2.2.5 Existing Ontologies

Ontologies that describe biomass availability and land use have been surveyed. Consideration was given to the coverage of the ontologies and their applicability to the data surveyed in the previous sections. A summary of the survey is given in **Table 4**.

Ontology	Subject Material Covered	Evaluation of Use
BiOnto [120]	Different types of biomass, biomass composition & biorefining processes.	Full ontology is yet to be published so a full assessment is not possible.
ENVO [19, 20]	Range of concepts relating to how organisms interact with their environments.	Broad scope and definitions mean the ontology is not able to describe land use in the required detail.
Bioenergy Ontology [103]	Different types of biomass and their properties. Provides information on the which bioenergy pathway to pursue for a given biomass.	Lack of quantification of variables means the ontology does not meet the needs of this iteration of the digital twin.
Crop Ontology [65, 66]	Many crops and their traits relevant to breeding.	The purpose of this ontology means that although its scope is broad, it is not relevant to this iteration of the digital twin.
ONS Geography Linked Data [88]	Geospatial entities in the UK with a hierarchical approach.	Ontology properties allow geospatial queries within regions to be carried out but not within arbitrary areas.

Table 4: Comparison of ontologies related to biomass availability and land use.

BiOnto was developed in 2015 to describe biomass types, biomass composition and biorefining process technologies [120]. An ontology describing biomass, the type and composition of biomass would be valuable information. It appears that BiOnto may provide only qualitative descriptions of some things, for example, heating values are described as 'low', 'medium' or 'high'. However, the full ontology is yet to be published, so it is not possible to make a full assessment.

The Environmental Ontology (ENVO) was created in 2013 to ontologise a range of concepts relating to environmental classification [19]. Relevant concepts in the ontology include land descriptions such as grassland and woodland. As of 2016, the objective of ENVO was to describe the interactions of organisms with their environments [20]. Whilst this focus is clearly of considerable value, ENVO is not able to describe land use in the level of detail required for the current iteration of the digital twin.

The Bioenergy Ontology [103] was published in 2015 to describe different types of biomass and its properties. The ontology aims to help farmers and other stakeholders make decisions about the best bioenergy pathway to pursue for a given biomass. The ontology describes the major biomass types: agricultural residues, municipal waste, wood and crops, which are further divided into subcategories. This hierarchical classification is useful. However, the individual biomass types are described qualitatively in terms of properties such as consistency, phase, and moisture content, whereas a quantitative description is required to estimate the power available from the biomass.

Development of the Crop Ontology [66] began in 2008 and is ongoing. The ontology describes many crops, perhaps in part owing its longevity and its open-source nature. It is part of the Integrated Breeding Platform [65] and focuses on traits relevant to breeding. Again, whilst this is clearly valuable, it does not meet the needs of the current work.

The Office for National Statistics (ONS) publishes 'Geography Linked Data' [88] which describes various geographical entities in the UK, for example town, district and regional boundaries. Each entity has a geometry defined by a *hasGeometry* property. A hierarchy is created using the *within* property, which states that one geographic entity is within another. This approach is extremely useful. However, it cannot support arbitrary geospatial queries so it will still be necessary to encode additional geospatial data if it desired to answer such queries, for example to identify all resources within a particular distance of some location. The ONS data, along with data from several branches of UK government [32, 83, 87, 116], is published using the PublishMyData platform from Swirrl [109].

The broad scope of the ontologies surveyed here mean they may be useful in the future. However, they do not provide the classification of land use nor the support for geospatial queries required by the current use case. This means that it will be necessary to create new ontologies.

3 Methodology

3.1 Ontology Development

Ontologies have been created for the following domains: land use classification; representation of geospatial land use data; relation of land use to its constituent biomass and associated energy content. The ontologies provide a geospatial description of land use and biomass that the existing ontologies surveyed in Section 2.2.5 could not.

Ontologies can be developed using either a 'top-down' or 'bottom-up' approach [74]. The top-down approach involves defining high-level concepts that can be widely applied before defining increasingly specific terminologies, often involving manual work to define non-trivial relationships and concepts. The bottom-up approach is application focused, resulting in ontologies that cover the necessary concepts rather than having broad scope. In this work we adopt a bottom-up methodology.

The following sections describe the development of the ontologies and their interconnections. Whilst some aspects of the ontologies mirror the structure of the resources identified in Section 2.2.3 and 2.2.4, the ontologies were constructed to maximise their generality wherever practicable. The names of ontological classes and properties are written in italics, for example the *LandUseCode* and *LandUseCodeType* classes that are related through the *hasLandUseType* object property. A formal description logic [7] representation of the ontologies is provided in Appendix A.1. An archived version of the ontologies are provided as part of the Research Data associated with this paper.

3.1.1 OntoLandUse

OntoLandUse was developed to provide an ontology to describe land use classification based on the approach taken by the RPA and CROME. The ontology uses LUCodes and their associated definitions to distinguish land uses. OntoLandUse was instantiated to represent the union of the 217 LUcodes defined by the RPA and 81 LUCodes used by CROME (77 of which overlap with the RPA definitions) using OWL. The design of the ontology is intended to generalise to allow land use codes defined by other jurisdictions.

Figure 3(a) shows the structure of the OntoLandUse TBox. The *LandCover* categories follow from the CROME data. Examples include 'Cereal Crops', 'Trees', 'Grassland' and 'Leguminous Crops'. These are general and could be used to describe land use anywhere. The *LandUseCode* follows from the definitions provided by the RPA and CROME. The *LandUseCodeType* allows for the grouping of LUCodes. The groups that contain the most LUCodes and with the most relevance to land use classification are 'Agricultural Land' and 'Non-Agricultural Land' [102]. Sub-classes of *LandUseCodeType* are created for these categories. The ontology also allows the specification of the *AdministrativeDivision* in which the *LandUseCode* and *LandUseCodeType* are used. This is intended to allow the extension of OntoLandUse to describe other countries and regions.

Figure 3(b) shows an excerpt from the ABox for spring barley (LUCode AC01). The

rdfs:label data property is used for the primary description of the LUCode provided by CROME. The 'Land Use' name and 'Description' of the LUCode provided by the RPA [102] are captured using *skos:altLabel* and *rdfs:comment* properties respectively. Although not shown in fig. 3, the TBox imposes cardinality restrictions that limit the number of range instances (of classes) that can be related to a domain instance by a given object property. For example, an instance of *LandUseCode* can only be linked to one instance of *LandCover* by the *isConnectedTo* object property.



(a) OntoLandUse TBox. OntoCropMapGML defines classes and their relations that conceptualise the domain of Land Use classification.



(b) Excerpt from the OntoLandUse ABox for AC01, showing the instances, classes and the relations which populate the Land Use classification ontology.

Figure 3: Structure of the OntoLandUse ontology.

3.1.2 OntoCropMapGML

OntoCropMapGML establishes a vocabulary for the geospatial terminology that appears in the (Geography Markup Language format) CROME data to enable geospatial queries.

Figure 4 shows the TBox of OntoCropMapGML. It incorporates properties and ranges described by a number of general-purpose ontologies, including data and object properties defined by OntoCityGML [21], which defines terms used to describe built environments that can appropriately describe certain geospatial elements of the CROME data. The classes and relationships that have been defined elsewhere are prefixed with the namespace of their native ontologies. For example, the *OntoCityGML:boundedBy* object property is defined by OntoCityGML. This is consistent with best-practice because the sharing of common terminologies from high-level top-down ontologies that define abstract concepts promotes interoperability with other ontologies. A full list of namespaces is provided in Appendix A.2.



Figure 4: OntoCropMapGML TBox. OntoCropMapGML links the description of land use classification provided by OntoLandUse with a geospatial description of land use.

Each instance of the *CropMap* class, also referred to as a feature member, derives from one entry (*i.e.* one hexagonal cell) in the CROME data set. The data properties associated with an instance of *CropMap* describe its centrepoint location, geometry and associated metadata. The centre point is encoded via a *datex:centrePoint* object property that links to an instance of the *BigData:lat-lon* class, whilst the geometry of the (hexagonal) cell boundary in encoded via a *hasGeometry* object property that links to an instance of a *WA:POLYGON-2-14* class. Both enable geospatial queries via Blaze-graph [15]. This choice of how to encode the geospatial data and its implications are discussed later in Section 4. The *hasLucode* object property links to an instance of the *OntoLandUse:LandUseCode* class, allowing geospatial queries levied via OntoCropGML

to retrieve data about land use. The *OntoCityGML:boundedBy* object property links to an instance of the *OntoCityGML:EnvelopeType* class that describes the bounding region to which the instance of *CropMap* belongs. The bounding coordinates and other metadata about the *OntoCityGML:EnvelopeType* are described by further data properties.

OntoCropMapGML was instantiated to represent the full CROME data set using OWL. The instantiation was complicated by the need to convert the geospatial data elements from the EPSG:27700 [79] to the EPSG:4326 [80] coordinate system. EPSG:27700 uses easting and northing references to a two-dimensional projection of Great Britain, whereas EPSG:4326 includes a more advanced geodesy with angles of latitude and longitude specifying a location on the surface of an ellipsoidal model of Earth. The conversion is nonlinear and the conversion procedure inherently iterative and non-exact, with some procedures resulting in significant errors that vary with geographical position. Permitting such conversion errors would run counter to the objective to develop a high-quality digital twin and would negate the accuracy of the raw CROME data. The conversion was performed using pyproj [135], which provides accurate conversion (error $\ll 1$ m) at acceptable computational cost.

3.1.3 OntoCropEnergy

OntoCropEnergy has been created to provide an ontology to define the minimum terminology required to allow land use to be related to the biomass made available by the land, and to estimate the energy content and rate of production of the biomass. This enables quantitative calculations surrounding land use. Although analyses of energy provision frequently focus on power per unit area [4, 76], it was decided to distinguish between crop yield (mass productivity per unit area of biomass) and calorific value (energy content per unit mass) to broaden the scope of OntoCropEnergy. For example, food production applications can use OntoCropEnergy in a way that would not be possible it solely defined concepts related to power generation from biomass.

Figure 5 shows an excerpt from the OntoCropEnergy TBox. The *Crop* class is so named to maintain a consistent nomenclature with OntoLandUse and OntoCropMapGML. The full TBox defines 23 sub-classes of *Crop*, however for the sake of clarity, fig. 5 shows only *Barley*. The *OntoCropMapGML:hasLucode* object property allows an instance of *Crop* (or a sub-class of *Crop*) to be linked to an instance of *OntoLandUse:LandUseCode* in order to enable queries relating to land use to resolve information about the crop. The *Crop* class has object properties that allow links to *GrossCalorificValue*, *NetCalorificValue* and *CropYield* classes, each of which have data and object properties to allow the specification of a numerical value and associated units, and a URL and access date to provide information about the provenance of the data. The OM namespace refers to a fork of the Ontology of units of Measure 2.0 [97, 98] (where the fork was necessary to define new units). The properties and classes used to encode web links are imported from OntoSpecies [51, 118], which was developed as part of the World Avatar. A full list of namespaces is provided in Appendix A.2.

OntoCropEnergy was instantiated to represent data for 33 crops using OWL. The yield



Figure 5: Excerpt from the OntoCropEnergy TBox. OntoCropEnergy links the description of land use classification provided by OntoLandUse with data describing the energy content and yield of different crops.

data were sourced from FAOStat [114], except for miscanthus [55], sunflower [136] and maize [31]. The calorific value data were sourced from Phyllis2 [119], wherever possible using the 'As Received' value given that the yield data does not account for further treatment of the crop. It was necessary to choose between different data sources for the calorific value and use a country-averaged mass productivity per unit area for each crop. This is not the most accurate approach. Ideally, the factors affecting the crops would be described via links to more detailed ontologies that included things such as climate (including temperature, rainfall and solar intensity variation), soil condition, nutrient availability as well as agronomical effects. This is beyond the scope of this work, although it represents an opportunity for valuable future work. Nevertheless, OntoCropEnergy offers quantitative information that was lacking from the ontologies surveyed in Section 2.2.5. These data extend the scope of possible queries of OntoLandUse and OntoCropMapGML to address quantitative questions relating to land use.

3.1.4 Interconnection Between Ontologies

OntoLandUse, OntoCropMapGML and OntoCropEnergy are interconnected in order to enable geospatial queries of land use, and to enable the results of such queries to be related information about the biomass available on that land.

Figure 6 shows the interconnection between the ontologies. The *LandUseCode* concept is central to the ability to relate information provided by one ontology to that provided by the others. OntoLandUse uses *LandUseCode* to classify land (based on the LUCodes defined by the RPA and CROME). OntoCropMapGML encodes (a feature member from the CROME data representing) a parcel of land as a *CropMap* that is related to a *Lan*-

dUseCode by a *hasLucode* property. OntoCropEnergy defines the relationship between a *Crop* and a *LandUseCode* using the same *hasLucode* property. In this manner, the interconnection enables the land represented by a *CropMap* to be associated with a *Crop*.



Figure 6: Interconnection between the OntoLandUse, OntoCropMapGML and OntoCropEnergy ontologies.

The structure of the ontologies is such that they can incorporate land use and biomass data from other sources, including for other countries and regions of the world. The inclusion of terminologies specifying the energy content and mass productivity per unit area (and thus the available power per unit area) of crops enables the ontologies to be use to support calculations regarding the use of biomass for energy (and food).

4 Use Case

4.1 Knowledge Graph Deployment

The OntoLandUse, OntoCropMapGML and OntoCropEnergy ontologies described in Section 3 have been deployed in a knowledge graph hosted using an instance of Blaze-graph (https://kg.cmclinnovations.com/blazegraph_geo). The deployed data describe the land use in the counties of Cambridgeshire, Norfolk and Suffolk in South East England. The data consist of approximately 33 million RDF triples and are 4.6 GB in size.

The native geospatial capability of Blazegraph is limited to queries of 2D or 3D points. In order to semantically represent the hexagonal cells in the CROME data, Blazegraph was extended by integrating a custom vocabulary to define a *POLYGON-2-14* data type that can be linked by data properties to classes. The name *POLYGON-2-14* was an arbitrary choice, but was chosen to indicate that the data type represents a 2D object described by 14 data values (seven pairs of latitude and longitude coordinates to represent a hexagon,

where the first and last pair of coordinates are the same for a closed shape). This method of defining and naming custom data types is extensible and mirrors the approach developed by Chadzynski et al. [21], where a family of custom data types were used for the purpose of describing 3D city data. An archived version of the custom vocabulary and data type is provided as part of the Research Data associated with this paper.

4.2 Example Geospatial Queries

Blazegraph provides native support for geospatial queries via *inRectangle* and *inCircle* search methods. The following queries to illustrate the native and extended geospatial capability of the knowledge graph. The limitations of the capability is discussed.

Query 1 shows an example that uses the native *inRectangle* search method to retrieve the location, geometry and land use code of land features. Blazegraph resolves the query by using (non-GeoSPARQL) geospatial reasoning to find features with centre points (described by the *BigData:lat-lon* type) located inside the search area. The south-west and north-east points that define the search area must be specified as *BigData:lat-lon* points. **Queries A.1 and A.2** (in the Appendix) show similar queries using the *inCircle* method.

Query 1: Geospatial SPARQL query to retrieve the location (lat#lon), geometry (POLYGON-2-14) and LUCode of land features located in a region defined by the south-west and north-east corners of a rectangle. Blazegraph resolves the query by performing geospatial reasoning against the centre points of the land features.

```
PREFIX geo: <http://www.bigdata.com/rdf/geospatial#>
PREFIX datex: <http://vocab.datex.org/terms#>
PREFIX BigData: <http://www.bigdata.com/rdf/geospatial/literals/v1#>
PREFIX OntoCropMapGML:
  <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#>
SELECT ?location ?geometry ?LUCode
WHERE
  SERVICE geo:search
  {
    ?cropMap geo:search "inRectangle" .
    ?cropMap geo:predicate datex:centrePoint .
    ?cropMap geo:searchDatatype BigData:lat-lon
    ?cropMap geo:spatialRectangleSouthWest "52.35#0.07"
    ?cropMap geo:spatialRectangleNorthEast "52.44#0.21" .
  }
  ?cropMap datex:centrePoint ?location .
  ?cropMap OntoCropMapGML:hasGeometry ?geometry .
  ?cropMap OntoCropMapGML:hasLucode ?LUCode .
} LIMIT 10 \# limit number of results to keep response time reasonable
```

Query 2 shows an example that uses the extended geospatial capability of Blazegraph to retrieve the location, geometry and land use code of land features. The query is resolved by using geospatial reasoning to find features described by the custom *POLYGON-2-14*

type. Searching against a custom geospatial data type that consists of more than one geospatial point requires the specification of a search area using the *geo:customFields*, *geo:customFieldsLowerBounds* and *geo:customFieldsUpperBounds* predicates. The number of items specified for each predicate must match the number of items specified in the custom type linked to the vocabulary (so 14 items to conform with *POLYGON-2-14* in this case). Likewise, the names (*e.g.* LATO, LONO) specified for *geo:customFields* must also match the names specified in the custom type. Despite the more complex syntax, the coordinates used to specify the search area in Query 2 consist of 7 repeats of the coordinates used in Query 1, so both examples actually search the same area.

Query 2: Geospatial SPARQL query to retrieve the location (lat#lon), geometry (POLYGON-2-14) and LUCode of land features located in a region defined by the south-west and north-east corners of a bounding box. Blazegraph resolves the query by performing geospatial reasoning against instances of the custom POLYGON-2-14 data type. The corners of the bounding box are specified using the geo:customFields, geo:customFieldsLowerBounds and geo:customFieldsUpperBounds predicates. The number of items specified in the predicates must conform with the number of items in the specification of the custom type, so each must have 14 items to conform with POLYGON-2-14.

```
PREFIX geo: <http://www.bigdata.com/rdf/geospatial#>
PREFIX datex: <http://vocab.datex.org/terms#>
PREFIX OntoCropMapGML:
    <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#>
SELECT ?location ?geometry ?LUCode
WHERE
  SERVICE geo:search
    ?cropMap geo:predicate OntoCropMapGML:hasGeometry .
    ?cropMap geo:searchDatatype
     <http://www.theworldavatar.com/ontology/datatype/POLYGON-2-14> .
    ?cropMap geo:customFields
      "LAT0#LON0#LAT1#LON1#LAT2#LON2#LAT3#LON3#LAT4#LON4#LAT5#LON5#LAT6#LON6" .
    ?cropMap geo:customFieldsLowerBounds
      "52.35#0.07#52.35#0.07#52.35#0.07#52.35#0.07#52.35#0.07#52.35#0.07#52.35#0.07" .
    ?cropMap geo:customFieldsUpperBounds
      "52.44#0.21#52.44#0.21#52.44#0.21#52.44#0.21#52.44#0.21#52.44#0.21#52.44#0.21
    ?cropMap geo:customFieldsValues ?customFields .
  ?cropMap datex:centrePoint ?location .
  ?cropMap OntoCropMapGML:hasGeometry ?geometry .
  ?cropMap OntoCropMapGML:hasLucode ?LUCode .
} LIMIT 10 # limit number of results to keep response time reasonable
```

Query 3 shows an example that uses a standard SPARQL query to retrieve data from an irregular area. The query uses the instance of *OntoCityGML:EnvelopeType* for Cambridgeshire that is linked to *OntoCityGML:CropMap* by *OntoCityGML:boundedBy* to define the scope of the query. (See fig. 4 for a reminder of the structure of OntoCropMapGML).

Query 3: SPARQL query to retrieve the location (lat#lon) and LUCode for all land features located in Cambridgeshire.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX datex: <http://vocab.datex.org/terms#>
PREFIX OntoCropMapGML:
   <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#>
PREFIX OntoCropMapGMLKB:
   <http://www.theworldavatar.com/kb/ontocropmapgml/>
PREFIX OntoCityGML:
   <http://www.theworldavatar.com/ontology/ontocitygml/citieskg/OntoCityGML.owl#>
SELECT ?location ?LUCode
WHERE
  ?cropMap rdf:type OntoCropMapGML:CropMap .
  ?cropMap OntoCityGML:boundedBy
   OntoCropMapGMLKB:Envelope_of_Crop_Map_of_England_2019_Cambridgeshire .
  ?cropMap datex:centrePoint ?location
 ?cropMap OntoCropMapGML:hasLucode ?LUCode .
} LIMIT 10 # limit number of results to keep response time reasonable
```

Figure 7 shows the land use in Cambridgeshire, Norfolk and Suffolk. The data were retrieved using Query 3 for Cambridgeshire, and analogous queries for Norfolk and Suffolk. The query output was written to GeoJSON files and visualised using Mapbox [78].

The limitation of the approach in Query 3 is that it is necessary to know that the instance of *OntoCityGML:EnvelopeType* is named 'Envelope_of_Crop_Map_of_England _2019_Cambridgeshire' to formulate the query. In the future, it is recommended to use *within* properties, analogous to the approach used by the ONS. In this case, linking to the Geography Linked Data [88] regions published by the ONS would enable simple queries to retrieve data from OntoCropMapGML for any region of the UK.

4.3 Elean Power Station

This section presents an example use case that uses the knowledge graph to identify the minimum radius from Elean Power Station that would be required to source enough biomass for it to operate at it maximum generation capacity. This is a cross-domain use case. It requires knowledge of the electrical power system and land use, and the properties of the biomass grown on he land. The ability to support such a cross-domain case highlights the benefit of the knowledge graph approach to digital twins, in this case for solving problems relating to the decarbonisation of the energy system.

Elean Power Station is a bioenergy plant located in Ely, Cambridgeshire. It was commissioned in the year 2000 and is fuelled using wheat, oilseed, and miscanthus [46]. It has a maximum generation capacity of 38 MWe [46] and an efficiency of 32.5% [105]. The queries and calculations performed by the use case were implemented in Python. An example calculation is given below and all relevant SPARQL queries are provided in Section A.4 of the Appendix.

1. Query data for biomass-fired power stations in the UK. Table 5 shows the result for



(a) Cambridgeshire.



(b) Norfolk.





Figure 7: The complete land use data set for Cambridgeshire, Norfolk and Suffolk.

Elean Power Station.

- 2. Query the land use codes of the crops (wheat, miscanthus and oilseed) that are able to be used by Elean Power Station [46]. **Table 6** shows the results of the query.
- 3. Query the yield and net calorific value of wheat, miscanthus and oilseed crops. **Table 7** shows the results of the query.
- 4. Perform a geospatial *inCircle* query centred on Elean Power Station to count the number of occurrences of the land use codes for wheat, miscanthus and oilseed as a function of radius around the power station. **Table 8** shows the results of the query.
- 5. Estimate the total power that could be generated using all the wheat, miscanthus and oilseed crops grown within a given radius of Elean Power Station.

$$P = NAYE, \tag{1}$$

where *P* is the power available for a crop (W), *N* is the number of occurrences of the land use code for the crop within the search radius (-), *A* is the area of corresponding to each occurrence of a land use code, *Y* is the yield (kg m⁻² s⁻¹) and *E* is the net calorific value of the crop (J kg⁻¹).

Table 9 shows the results of the calculation. The area *A* could have been retrieved via another query. However, in this case it was known *a priori* that $A \approx 4156 \text{ m}^2$, corresponding to the area of the hexagonal cells used by the CROME data [101].

6. Estimate the electricity that could be generated from crops in the search radius.

$$G = P_{\text{total}} \eta, \qquad (2)$$

where G is the power generated by Elean Power Station, P_{total} is the total power available and $\eta = 32.5\%$ is the assumed efficiency of Elean Power Station [105].

Table 5: Results of a SPARQL query to retrieve data for Elean Power Station.¹

Name	Ely
Capacity	40 MW
Built	2001
Fuel	Biomass
Technology	Conventional Steam
Latitude	52.3955987
Longitude	0.1640088

¹ The data in the knowledge graph returned by this query originate from the 2020 version of the Digest of UK Energy Statistics (DUKES) [27]. The discrepancy in the value of the capacity, which is reported elsewhere as 38 MW is noted.

Crop Name	LUCode
Winter Wheat	AC66
Winter Oilseed	AC67
Miscanthus	TC01
Spring Wheat	AC32
Spring Oilseed	AC36

Table 6: Results of a SPARQL query to retrieve the LUCodes of crops consumed by Elean

 Power Station.

Table 7: Results of SPARQL queries to retrieve the yield and net calorific value of crops consumed by Elean Power Station.

Crop Name	Yield / te ha ⁻¹ yr ⁻¹	Net Calorific Value / MJ kg ⁻¹
Winter Wheat	8.9345	14.70
Winter Oilseed	3.3057	15.73
Miscanthus	14.0000	10.12
Spring Wheat	8.9345	14.70
Spring Oilseed	3.3057	15.73

Table 8: Results of a geospatial query to count the LUCode occurrences within 17.1 km of Elean Power Station.

Crop Name	LUCode	Occurrences / -
Winter Wheat	AC66	58,510
Winter Oilseed	AC67	7,656
Miscanthus	TC01	3,186
Spring Wheat	AC32	2,511
Spring Oilseed	AC36	0

Table 9: Power associated with crops within 17.1 km of Elean Power Station.

	101.0
Winter Wheat	101.2
Winter Oilseed	5.2
Miscanthus	5.9
Spring Wheat	4.3
Spring Oilseed	0.0
Total	116.7

Figure 8 shows the geospatial distribution of crops that can be used by Elean Power Station as a function of radius around the power station. **Figure 9** shows the corresponding electrical power estimated using eq. (2). The total wheat, miscanthus and oilseed grown within a 17.1 km radius of Elean Power Station would be sufficient for it to operate at its maximum generation capacity of 38 MWe. Clearly this is significantly less than the actual 100 km radius used to source biomass for Elean [105] because biomass supply chains are

not determined on geometric grounds alone! Nevertheless, it provides a useful illustration of the type of problem that can be addressed using the knowledge graph.

Careful consideration must be employed when making decisions concerning land use and biomass resources. The total land area associated with the crops required for Elean Power Station to generate 38 MWe is approximately 300 km². If this land was instead used to produce food or used for other types of renewables, it is estimated that it would be sufficient to grow food for approximately 60 thousand people, generate approximately 300 MWe using wind power [10] or 1020 MWe using solar photovoltaics [10, 57, 107]. This glosses over many other important considerations including the fact that solar and wind power are intermittent, and that bioenergy with carbon capture and storage can produce negative emissions. Nevertheless, these numbers highlight the 'premium' associated with using land to grow crops for bioenergy.

Future developments of digital twin will extend the knowledge graph to include solar, wind, and population data, and will seek to automate this type of analysis using computational agents. It will also seek to link to ontologies that enable biodiversity to be considered in the assessments and to generalise the natural language capability of the knowledge graph [see 141] to make it easier to search for and retrieve data.



Figure 8: The results from a geospatial query of crops grown in the vicinity of Elean Power Station. The query was restricted to the types of crops (wheat, miscanthus and oilseed) that can be used by the power station.



Figure 9: Estimated electrical power that could be generated using all the wheat, miscanthus and oilseed crops grown within a given radius of Elean Power Station. The points show calculated values, the line serves as a guide to the eye.

5 Conclusion

A set of ontologies has been developed to allow a geospatial description of land use to be incorporated into a dynamic-knowledge-graph-based Universal Digital Twin. The benefit of this strategy has been demonstrated through a cross-domain use case that shows an example of how such a knowledge graph could be used to support decision making about how to balance the use of land resources to meet increasing demand for energy whilst cutting emissions.

Resources that provide data about land coverage and biomass, and existing ontologies that describe these subject domains were critically examined. The Crop Map of England (CROME) published by the UK Government was found to provide a detailed geospatial description of land use in England. The data are updated annually and are available under an Open Government Licence in a choice of machine-readable formats. However, no ontologies capable of describing the data were found.

Three new ontologies were developed to support a geospatial description of land use.

- i OntoLandUse provides an ontology to describe land use classification based on the use of alphanumeric land use codes alongside natural language descriptions of the land use (the socio-economic function of the land) and land cover (the observed to coverage of the land). The ontology is structured to allow the description of land use codes from multiple countries and regions, and was instantiated to represent all the land use codes used by the UK Government.
- ii OntoCropMapGML provides a definition of the terminology required to provide a geospatial description of land use. It was instantiated to represent the full CROME data set, providing a geospatial description of land use across the whole of England.

iii OntoCropEnergy was developed to define the minimum terminology required to allow land use to be related to the yield (mass productivity per unit area) and calorific value (energy content per unit mass) of the biomass made available by the land. It was instantiated to provide data for 33 crops.

Ontologies for a subset of the instantiated data were deployed in a knowledge graph that is hosted using an instance of the Blazegraph graph database (https://kg.cmclinnovatio ns.com/blazegraph_geo). The deployed data consist of approximately 33 million RDF triples and describe the land use in the counties of Cambridgeshire, Norfolk and Suffolk in South East England. The native geospatial capability of Blazegraph is limited to pointwise data. The geospatial capability of Blazegraph was extended by integrating a custom vocabulary to allow the semantic representation of the boundaries of the hexagonal cells used to discretise the geospatial description of land use in the CROME data. The extended geospatial capability of Blazegraph was critically assessed. The custom geospatial queries could only be formulated in terms of a single custom data type. This was not an issue for the CROME data (because the features of interest were uniformly hexagonal), but will clearly be limiting in other cases.

The capability of knowledge-graph-based digital twins has been demonstrated in an illustrative cross-domain use case concerning a bioenergy plant in Cambridgeshire. The use case combined information in the knowledge graph about UK power plants with a geospatial analysis of land use to estimate the minimum land required to provide enough biomass to operate the bioenergy plant. It is trivial for this type of analysis to be performed for any region described in the knowledge graph (in this case anywhere in England). Comparisons with alternative uses for the land highlighted the complexities and trade-offs that will be required when making decisions about the best way to use land to meet our future energy needs whilst achieving net zero.

Opportunities for future work to expand the data coverage and capabilities of the digital twin have been identified. Potential improvements to OntoCropEnergy have been high-lighted, with a view to enabling the digital twin to take into account more detailed data about the factors influencing the yield and energy content of biomass. Potential improvements to how to encode geospatial data have been discussed.

Nomenclature

ABox Assertional Component (of an ontology)
API Application Programming Interface
BECCS Bioenergy with Carbon Capture and Storage
CCS Carbon Capture and Storage
CROME Crop Map of England
CSV Comma Separated Variable
DL Description Logic
DUKES Digest of UK Energy Statistics

ENVO Environmental Ontology FAO Food and Agriculture Organization of the United Nations FCA Forestry Contracting Association FC Forestry Commission GeoJSON Geospatial JavaScript Object Notation GeoSPARQL Geographic Query Language for RDF Data **GML** Geography Markup Language Internationalised Resource Identifier IRI LUCode Land Use Code **OGC** Open Geospatial Consortium **ONS** Office for National Statistics **OWL** and Web Ontology Language **RDF** Resource Description Framework **RPA** Rural Payments Agency SPARQL SPARQL Protocol and RDF Query Language **TBox** Terminological Component (of an ontology) W3C World Wide Web Consortium

Research data

Research data supporting this publication is available in the University of Cambridge data repository (doi:10.17863/CAM.68278).

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. Part of the research was also funded by the European Commission, Horizon 2020 Programme, DOME 4.0 Project, GA 953163. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Markus Kraft gratefully acknowledges the support of the Alexander von Humboldt foundation.

A Appendix

A.1 Description Logic Representation of Ontologies

A.1.1 OntoLandUse

Classes

LandUseCode $\sqsubseteq \top$ LandCover $\sqsubseteq \top$ LandUseCodeType $\sqsubseteq \top$ AgriculturalLand \sqsubseteq LandUseCodeType Non-AgriculturalLand \sqsubseteq LandUseCodeType AdministrativeDivision $\sqsubseteq \top$

Object Properties

LandUseCode $\sqsubseteq \leq 1$ isConnectedTo.LandCover \sqcap
\geq 1 isConnectedTo.LandCover
LandUseCode $\sqsubseteq \leq 1$ hasLandUseType.LandUseCodeType \sqcap
\geq 1 hasLandUseType.LandUseCodeType
(LandUseCode \sqcup LandUseCodeType) $\sqsubseteq \forall$ usedIn.AdministrativeDivision

A.1.2 OntoCropMapGML

Classes

OntoCityGML:EnvelopeType $\sqsubseteq \top$ CropMap $\sqsubseteq \top$

Object Properties

 $\begin{array}{ll} CropMap & \sqsubseteq \leq 1 \mbox{ OntoCityGML:boundedBy.EnvelopeType} \ \sqcap \\ & \geq 1 \mbox{ OntoCityGML:boundedBy.EnvelopeType} \\ CropMap & \sqsubseteq \leq 1 \mbox{ hasLandUseCode.OntoLandUse:LandUseCode} \ \sqcap \\ & \geq 1 \mbox{ hasLandUseCode.OntoLandUse:LandUseCode} \end{array}$

Data Properties

 $\exists \ OntoCityGML: lowerCornerPoint. \top \sqsubseteq OntoCityGML: EnvelopeType$

 $\top \sqsubseteq \forall$ OntoCityGML:lowerCornerPoint.BigData:lat-lon \exists OntoCityGML:upperCornerPoint. $\top \sqsubseteq$ OntoCityGML:EnvelopeType $\top \sqsubseteq \forall$ OntoCityGML:upperCornerPoint.BigData:lat-lon \exists OntoCityGML:srsname. $\top \sqsubseteq$ OntoCityGML:EnvelopeType $\top \sqsubseteq \forall$ OntoCityGML:srsname.String \exists srsDimension. $\top \sqsubseteq$ OntoCityGML:EnvelopeType $\top \sqsubseteq \forall$ srsDimension.Integer \exists sql:ObjectIdentifier. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ sql:ObjectIdentifier.Integer \exists hasCromeID. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ hasCromeID.String \exists dbo:area. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ dbo:area.Double \exists vocab:length. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ vocab:length.Double \exists hasRefDate. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ hasRefDate.Integer \exists hasGeometry. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ hasGeometry.WA:POLYGON-2-14 \exists datex:centrePoint. $\top \sqsubseteq$ CropMap $\top \sqsubseteq \forall$ datex:centrePoint.BigData:lat-lon

A.1.3 OntoCropEnergy

Classes

 $\begin{array}{c} \text{Crop} \ \sqsubseteq \ \top \\ \text{Barley} \ \sqsubseteq \ \text{Crop} \\ \text{Beet} \ \sqsubseteq \ \text{Crop} \\ \text{Carrot} \ \sqsubseteq \ \text{Crop} \\ \text{Chicory} \ \sqsubseteq \ \text{Crop} \\ \text{Chicory} \ \sqsubseteq \ \text{Crop} \\ \text{Lettuce} \ \sqsubseteq \ \text{Crop} \\ \text{Linseed} \ \sqsubseteq \ \text{Crop} \\ \text{Maize} \ \sqsubseteq \ \text{Crop} \\ \text{Oats} \ \sqsubseteq \ \text{Crop} \\ \text{Onions} \ \sqsubseteq \ \text{Crop} \\ \text{Rye} \ \sqsubseteq \ \text{Crop} \end{array}$

Strawberry	\sqsubseteq Crop
Triticale	\sqsubseteq Crop
Wheat	\sqsubseteq Crop
Cabbage	\sqsubseteq Crop
Turnip	\sqsubseteq Crop
Oilseed	\sqsubseteq Crop
Potato	\sqsubseteq Crop
Tomato	\sqsubseteq Crop
Sunflower	\sqsubseteq Crop
Field Beans	\sqsubseteq Crop
Green Beans	\sqsubseteq Crop
Peas	\sqsubseteq Crop
Miscanthus	\sqsubseteq Crop
GrossCalorificValue	ΞT
NetCalorificValue	ΞT
CropYield	ΞT
SurfacePowerDensity	ΞT
CropSurfacePowerDensity	$\sqsubseteq SurfacePowerDensity$

Object Properties

$Crop \ \sqsubseteq \forall \ OntoCropMapGML: hasLandUseCode. OntoLandUse: LandUseCode$	
Crop $\sqsubseteq \leq 1$ hasGrossCalorificValue.GrossCalorificValue \sqcap	
\geq 1 hasGrossCalorificValue.GrossCalorificValue	
$Crop \ \sqsubseteq \le 1 \ has NetCalorific Value. NetCalorific Value \ \sqcap$	
\geq 1 hasNetCalorificValue.NetCalorificValue	
$Crop \ \sqsubseteq \le 1 \ hasCropYield.CropYield \ \sqcap$	
\geq 1 hasCropYield.CropYield	
Crop $\sqsubseteq \leq 1$ hasCropSurfacePowerDensity.CropSurfacePowerDensity \sqcap	
\geq 1 hasCropSurfacePowerDensity.CropSurfacePowerDensity	
\exists OM:hasValue. $\top \sqsubseteq$ GrossCalorificValue \sqcup NetCalorificValue \sqcup	
CropYield CropSurfacePowerDensity	
$\top \sqsubseteq \forall$ OM:hasValue.OM:Measure	
\exists OM:hasUnit. $\top \sqsubseteq$ OM:Measure	
$\top \sqsubseteq \forall$ OM:hasUnit.OM:UnitDivision	
(CropYield ⊔	
NetCalorificValue \sqcup	
GrossCalorificValue) $\sqsubseteq \leq 1$ OntoSpecies:hasWeblink.OntoSpecies:Weblink \sqcap	
\geq 1 OntoSpecies:hasWeblink.OntoSpecies:Weblink	

Data Properties

 \exists OM:hasNumericalValue. $\top \sqsubseteq$ OM:Measure

 $\top \sqsubseteq \forall \mathsf{OM:hasNumericalValue.Double}$

 $\exists \ dateOfAccess. \top \sqsubseteq Weblink$

 $\top \sqsubseteq \forall \text{ dateOfAccess.String}$

 \exists hasURL. $\top \sqsubseteq$ Weblink

 $\top \sqsubseteq \forall$ hasURL.String

A.2 Namespaces

BigData: <http://www.bigdata.com/rdf/geospatial/literals/v1#> **Datex**: <http://vocab.datex.org/terms#> **Dbo**: <http://dbpedia.org/ontology/> OM: <http://www.ontology-of-units-of-measure.org/resource/om-2/> Geo: <http://www.bigdata.com/rdf/geospatial#> **OntoCityGML**: <http://www.theworldavatar.com/ontology/ontocitygml/OntoCityGML.owl#> **OntoCropEnergy**: <http://www.theworldavatar.com/ontology/ontocropenergy/OntoCropEnergy.owl#> **OntoCropEnergyKB**: <http://www.theworldavatar.com/kb/ontocropenergy/> **OntoCropMapGML**: <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#> **OntoCropMapGMLKB:** <http://www.theworldavatar.com/kb/ontocropmapgml/> **OntoLandUse**: <http://www.theworldavatar.com/ontology/ontolanduse/OntoLandUse.owl#> **OntoLandUseKB**: <http://www.theworldavatar.com/kb/ontolanduse/> **OntoSpecies:** <http://www.theworldavatar.com/ontology/ontospecies/OntoSpecies.owl#> Rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> Sql: <http://ns.inria.fr/ast/sql#> Vocab: <http://open.vocab.org/terms/> WA: <http://www.theworldavatar.com/ontology/datatype/>

A.3 Example Geospatial Queries

Query A.1: Geospatial SPARQL query to count the number of land features located in a region defined by the centre point and radius of a circle. Blazegraph resolves the query by performing geospatial reasoning against the centre points of the land features.

```
PREFIX BigData: <http://www.bigdata.com/rdf/geospatial/literals/v1#>
PREFIX geo: <http://www.bigdata.com/rdf/geospatial#>
PREFIX datex: <http://vocab.datex.org/terms#>
PREFIX OntoCropMapGML:
    <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#>
SELECT (COUNT (?Feature) as ?FeaturesTotal)
WHERE
{
  SERVICE geo:search
    ?cropMap geo:search "inCircle" .
    ?cropMap geo:predicate datex:centrePoint .
    ?cropMap geo:searchDatatype BigData:lat-lon .
    ?cropMap geo:spatialCircleCenter "52.40#0.13"
    ?cropMap geo:spatialCircleRadius "5" . # default unit: km
  }
  ?cropMap OntoCropMapGML:hasGeometry ?Feature
}
```

Query A.2: Geospatial SPARQL query to retrieve the location (lat#lon) and LUcode of land features located in a region defined by the centre point and radius of a circle. Blazegraph resolves the query by performing geospatial reasoning against the centre points of the land features.

```
PREFIX BigData: <http://www.bigdata.com/rdf/geospatial/literals/v1#>
PREFIX geo: <http://www.bigdata.com/rdf/geospatial#>
PREFIX datex: <http://vocab.datex.org/terms#>
PREFIX OntoCropMapGML:
    <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#>
SELECT ?location ?geometry ?LUCode
WHERE
  SERVICE geo:search
    ?cropMap geo:search "inCircle" .
    ?cropMap geo:predicate datex:centrePoint .
    ?cropMap geo:searchDatatype BigData:lat-lon
    ?cropMap geo:spatialCircleCenter "52.40#0.13"
    ?cropMap geo:spatialCircleRadius "5" . # default unit: km
  }
  ?cropMap datex:centrePoint ?location .
  ?cropMap OntoCropMapGML:hasGeometry ?geometry .
  ?cropMap OntoCropMapGML:hasLucode ?LUCode .
}
```

A.4 Queries Performed by the Elean Power Station Use Case

Query A.3 is levied against a part of the knowledge graph that is not currently public. However, a previous iteration of this part of the knowledge graph can be viewed online. See https://kg.cmclinnovations.com/explore/digital-twin/power-system.

Query A.3: *Query power station coordinates.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ontopowsys_PowSysRealization:
  <http://www.theworldavatar.com/ontology/ontopowsys/PowSysRealization.owl#>
PREFIX ontopowsys_PowSysPerformance:
  <http://www.theworldavatar.com/ontology/ontopowsys/PowSysPerformance.owl#>
PREFIX upper_level_system:
  <http://www.theworldavatar.com/ontology/ontocape/upper_level/system.owl#>
PREFIX ontoeip_powerplant:
 <http://www.theworldavatar.com/ontology/ontoeip/powerplants/PowerPlant.owl#>
PREFIX technical_system:
 <http://www.theworldavatar.com/ontology/ontocape/upper_level/technical_system.owl#>
PREFIX meta_model_topology:
  <http://www.theworldavatar.com/ontology/meta_model/topology/topology.owl#>
PREFIX space_and_time_extended:
  <http://www.theworldavatar.com/ontology/ontocape/supporting_concepts/space_and_time/</pre>
      space_and_time_extended.owl#>
PREFIX power_plant:
  <http://www.theworldavatar.com/ontology/ontoeip/powerplants/PowerPlant.owl#>
SELECT ?PowerPlant ?Built ?Capacity ?Unit ?Fuel ?Technology ?Latitude ?Longitude
WHERE
  ?PowerPlant ontoeip_powerplant:hasYearOfBuilt ?v_built.
  ?v_built upper_level_system:hasValue ?vv_built .
  ?vv_built upper_level_system:numericalValue ?Built .
  ?PowerPlant technical_system:hasRequirementsAspect/upper_level_system:hasValue ?v_capa
  ?v_capa upper_level_system:numericalValue ?Capacity .
  ?v_capa upper_level_system:hasUnitOfMeasure ?Unit .
  ?PowerPlant technical_system:hasRealizationAspect ?PowerGenerator .
  ?PowerGenerator a ontoeip_powerplant:PowerGenerator .
  ?PowerGenerator technical_system:realizes/ontoeip_powerplant:consumesPrimaryFuel ?Fuel
  ?PowerGenerator technical_system:realizes/ontoeip_powerplant:usesGenerationTechnology
      ?Technology .
  ?PowerPlant space_and_time_extended:hasGISCoordinateSystem ?CoordinateSystem .
  ?CoordinateSystem space_and_time_extended:hasProjectedCoordinate_x ?x_coordinate .
  ?CoordinateSystem space_and_time_extended:hasProjectedCoordinate_y ?y_coordinate .
  ?x_coordinate upper_level_system:hasValue ?GPS_x_coordinate .
  ?y_coordinate upper_level_system:hasValue ?GPS_y_coordinate .
  ?GPS_x_coordinate upper_level_system:numericalValue ?Longitude. # east/west
  ?GPS_y_coordinate upper_level_system:numericalValue ?Latitude . # north/south
  Filter(?Fuel= power_plant:Biomass) # only return data for biomass plants
}
```

The following queries are levied against https://kg.cmclinnovations.com/blazegraph_geo.

Query A.4: Query crops and LUCodes.

Query A.5: Query yield.

Query A.6: Query net calorific value.

Query A.7: Count LUCode occurrences within 17.1 km of Elean Power Station.

```
PREFIX BigData: <http://www.bigdata.com/rdf/geospatial/literals/v1#>
PREFIX geo: <http://www.bigdata.com/rdf/geospatial#>
PREFIX datex: <http://vocab.datex.org/terms#>
PREFIX OntoCropMapGML:
 <http://www.theworldavatar.com/ontology/ontocropmapgml/OntoCropMapGML.owl#>
PREFIX OntoLandUseKB:
 <http://www.theworldavatar.com/kb/ontolanduse/>
SELECT ?LUCode (COUNT(?LUCode) AS ?Occurrences)
WHERE
  SERVICE geo:search
  {
    ?cropMap geo:search "inCircle" .
    ?cropMap geo:predicate datex:centrePoint .
   ?cropMap geo:searchDatatype BigData:lat-lon .
    ?cropMap geo:spatialCircleCenter "52.3955987# 0.1640088" .
    ?cropMap geo:spatialCircleRadius "17.1" .
  3
  ?cropMap OntoCropMapGML:hasLucode ?LUCode
  Filter(?LUCode= OntoLandUseKB:AC32||
    ?LUCode= OntoLandUseKB:AC66
     ||?LUCode= OntoLandUseKB:AC36
        ||?LUCode= OntoLandUseKB:AC67
          ||?LUCode= OntoLandUseKB:TC01)
}
GROUP BY ?LUCode
ORDER BY DESC(?Occurrences)
```

References

- [1] J. Akroyd, S. Mosbach, A. Bhave, and M. Kraft. Universal Digital Twin A Dynamic Knowledge Graph. *Data-Centric Engineering*, 2021. In press.
- [2] D. Allemang and J. Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kauffman, 2nd edition, 2011. doi:10.1016/B978-0-12-385965-5.10003-2.
- [3] 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.
- [4] E. G. A. Antonini and K. Caldeira. Atmospheric pressure gradients and Coriolis forces provide geophysical limits to power density of large wind farms. *Applied Energy*, 281, 2021.
- [5] ARE3NA project "INSPIRE Re3ference Platform Phase 2". Guidelines for the RDF encoding of spatial data, 2017. Available at https://inspire-eu-rdf.github.io/in spire-rdf-guidelines/#ref_cr_prop (accessed Apr 2021).
- [6] 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? *Applied Energy*, 298:117117, 2021. doi:10.1016/j.apenergy.2021.117117.
- [7] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge University Press, 2nd edition, 2007. ISBN 978-0-521-87625-4. doi:10.1017/CBO9780511711787.
- [8] BBC. Onshore wind 'to make comeback in UK', 2020. Available at https://www. bbc.co.uk/news/science-environment-51708817 (accessed Apr 2021).
- [9] S. Bechhofer, F. Harmelen, J. Hendler, I. Horrocks, D. McGuinness, P. Patel-Schneijder, and L. Stein. OWL Web Ontology Language Reference. Recommendation, World Wide Web Consortium (W3C), February 10 2004. Available at http://www.w3.org/TR/owl-ref (accessed Apr 2021).
- [10] M. Berners-Lee. *There is no Planet B: a handbook for the make or break years*. Cambridge University Press, 2019. ISBN 9781108424240.
- [11] T. Berners-Lee. Linked data design issues, 2006. Available at https://www.w3.o rg/DesignIssues/LinkedData.html (accessed Apr 2021).
- [12] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284:34–43, 2001.

- [13] A. Bhave, R. H. S. Taylor, P. Fennell, W. R. Livingston, N. Shah, N. M. Dowell, J. Dennis, M. Kraft, M. Pourkashanian, M. Insa, J. Jones, N. Burdett, A. Bauen, C. Beal, A. Smallbone, and J. Akroyd. Screening and techno-economic assessment of biomass-based power generation with CCS technologies to meet 2050 CO₂ targets. *Applied Energy*, 190:481–489, 2017. doi:10.1016/j.apenergy.2016.12.120.
- [14] A. Bijlsma, G. Bull, R. Coppock, R. Duckworth, L. Halsall, J. B. Hudson, R. J. Hudson, D. Johnson, B. Jones, M. Lightfoot, E. Mackie, A. Mason, R. Matthews, H. McKay, N. Purdy, M. Sendros, and S. Smith. Woodfuel Resource in Britain. Forest Research, Forestry Commission, UK, 2003. Available at https://www.forest research.gov.uk/documents/2766/Woodfuel_In_Britain_2003.pdf (accessed Apr 2021).
- [15] Blazegraph, 2021. Available at https://blazegraph.com (accessed Mar 2021).
- [16] BP. In three charts the dual challenge, 2019. Available at https://www.bp.com /en/global/corporate/news-and-insights/reimagining-energy/dual-challenge-inthree-charts.html (accessed Mar 2021).
- [17] L. Breiman. Random Forests. *Machine learning*, 45(1):5–32, 2001. doi:10.1023/A:1010933404324.
- [18] A. S. Brouwer, M. van den Broek, A. Seebregts, and A. Faaij. Impacts of largescale intermittent renewable energy sources on electricity systems, and how these can be modeled. *Renewable and Sustainable Energy Reviews*, 33:443–466, 2014. doi:10.1016/j.rser.2014.01.076.
- [19] P. L. Buttigieg, N. Morrison, B. Smith, C. Mungall, and S. Lewis. The Environment Ontology: contextualising biological and biomedical entities. *Journal of biomedical semantics*, 4:43, 12 2013. doi:10.1186/2041-1480-4-43.
- [20] P. L. Buttigieg, E. Pafilis, S. E. Lewis, M. P. Schildhauer, R. L. Walls, and C. J. Mungall. The Environment Ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of biomedical semantics*, 7 (1):57–57, 2016. doi:10.1186/s13326-016-0097-6.
- [21] 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, 2021. ISSN 1473–4273. Submitted for publication. Preprint available at https://como.ceb.cam.ac.uk/preprints/273/.
- [22] Climate Change Committee. Biomass in a low-carbon economy, 2018. Available at https://www.theccc.org.uk/publication/biomass-in-a-low-carbon-economy/ (accessed Apr 2021).
- [23] Climate Change Committee. Net zero the UK's contribution to stopping global warming, 2019. Available at https://www.theccc.org.uk/wp-content/uploads/2019 /05/Net-Zero-The-UKs-contribution-to-stopping-global-warming.pdf (accessed Apr 2021).

- [24] Climate Change Committee, 2021. Available at https://www.theccc.org.uk (accessed Apr 2021).
- [25] CMCL Innovations. UK Digital Twin, 2021. Available at https://kg.cmclinnovatio ns.com/explore/digital-twin (accessed Jun 2021).
- [26] Data & Knowledge Group. HermiT OWL Reasoner, 2019. Available at http: //www.hermit-reasoner.com (accessed Dec 2020).
- [27] Department for Business, Energy & Industrial Strategy. Digest of UK Energy Statistics (DUKES): Electricity, 2020. Available at https://www.gov.uk/governmen t/statistics/electricity-chapter-5-digest-of-united-kingdom-energy-statistics-dukes (accessed Dec 2020).
- [28] Department for Business, Energy & Industrial Strategy. UK Energy in Brief, 2020. Available at https://assets.publishing.service.gov.uk/government/uploads/system /uploads/attachment_data/file/904503/UK_Energy_in_Brief_2020.pdf (accessed Apr 2021).
- [29] Department for Environment, Food & Rural Affairs (United Kingdom). Digest of waste and resource statistics, 2015 edition, 2015. Available at https://www.gov. uk/government/statistics/digest-of-waste-and-resource-statistics-2015-edition (accessed Apr 2021).
- [30] Department for Environment, Food & Rural Affairs (United Kingdom). Statistics on waste managed by local authorities in England in 2018/19, 2019. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attac hment_data/file/918853/201819_Stats_Notice_FINAL_accessible.pdf (accessed Jan 2021).
- [31] Department for Environment, Food & Rural Affairs (United Kingdom). Farming Statistics – final crop areas, yields, livestock populations and agricultural workforce at 1 June 2020 United Kingdom, 2020. Available at https://www.gov.uk/governm ent/statistics/farming-statistics-final-crop-areas-yields-livestock-populations-andagricultural-workforce-at-1-june-2020-uk (accessed Apr 2021).
- [32] Department for Environment, Food & Rural Affairs (United Kingdom), 2021. Available at https://www.gov.uk/government/organisations/department-forenvironment-food-rural-affairs (accessed Apr 2021).
- [33] Department for Environment, Food & Rural Affairs (United Kingdom) and The Rt Hon George Eustice MP. Legally binding targets to help 'build back greener', 2020. Available at https://www.gov.uk/government/news/legally-binding-targetsto-help-build-back-greener (accessed Jan 2021).
- [34] Department of Energy & Climate Change. UK renewable energy road map, 2011. Available at https://assets.publishing.service.gov.uk/government/uploads/system /uploads/attachment_data/file/48128/2167-uk-renewable-energy-roadmap.pdf (accessed Apr 2021).

- [35] 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.
- [36] 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.
- [37] 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/.
- [38] Dogger Bank Wind Farm, 2021. Available at https://doggerbank.com (accessed Apr 2021).
- [39] Drax. Drax's largest biomass shipment arrives at the uk's biggest biomass handling facility, 2020. Available at https://www.drax.com/press_release/draxs-largest-bi omass-shipment-arrives-at-the-uks-biggest-biomass-handling-facility/ (accessed Apr 2021).
- [40] Drax, 2021. Available at https://www.drax.com (accessed Apr 2021).
- [41] K. Durante. Ontology for Geography Markup Language (GML3.0), 2017. Available at https://bioportal.bioontology.org/ontologies/GML (accessed Feb 2021).
- [42] Eclipse Foundation. RDF4J, 2021. Available at https://rdf4j.org/ (accessed Mar 2021).
- [43] Eclipse Foundation. RDF4J Documentation GeoSPARQL, 2021. Available at https://rdf4j.org/documentation/programming/geosparql/ (accessed Mar 2021).
- [44] A. Eibeck, M. Q. Lim, and M. Kraft. J-Park Simulator: An ontology-based platform for cross-domain scenarios in process industry. *Computers & Chemical Engineering*, 131:106586, 2019. doi:10.1016/j.compchemeng.2019.106586.
- [45] 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.
- [46] Engineering Timelines. Elean biomass power station, 2000. Available at https: //web.archive.org/web/20141213020510/http://www.engineering-timelines.com/ scripts/engineeringItem.asp?id=978 (accessed Apr 2021).
- [47] ENI, 2021. Available at https://www.eni.com/en-IT/home.html (accessed Apr 2021).

- [48] Equinor, 2021. Available at https://www.equinor.com (accessed Apr 2021).
- [49] EuroStat, Office for Official Publications of the European Communities. Manual of concepts on land cover and land use information systems (2000 edition), 2001. Available at https://ec.europa.eu/eurostat/documents/3859598/5854849/KS-34-00-407-EN.PDF.pdf/29c7f777-fdbe-405e-a29d-7dde0536a68c?t=14147803900 00 (accessed Jan 2021).
- [50] 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.
- [51] F. Farazi, N. B. Krdzavac, J. A. S. Mosbach, A. Menon, D. Nurkowski, and M. Kraft. Linking reaction mechanisms and quantum chemistry: An ontological approach. *Computers & Chemical Engineering*, 137, 2020. doi:10.1016/j.compchemeng.2020.106813.
- [52] 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.
- [53] Forest Research, Forestry Commission, UK. National Inventory of Woodland and Trees – England, 2003. Available at https://www.forestresearch.gov.uk/toolsand-resources/national-forest-inventory/national-inventory-of-woodland-andtrees/national-inventory-of-woodland-and-trees-england (accessed Apr 2021).
- [54] Forest Research, Forestry Commission, UK. What is biomass?, 2021. Available at https://www.forestresearch.gov.uk/tools-and-resources/fthr/biomass-energyresources/general-biomass-information/what-is-biomass/ (accessed Apr 2021).
- [55] Forest Research, Forestry Commission, UK. Grasses, 2021. Available at https: //www.forestresearch.gov.uk/tools-and-resources/fthr/biomass-energy-resources/ fuel/energy-crops/grasses/ (accessed Apr 2021).
- [56] Forestry Commission. National Forest Inventory 50-year forecast of softwood availability GB, 2014. Available at https://www.forestresearch.gov.uk/documents/ 2744/50_YEAR_FORECAST_OF_SOFTWOOD_AVAILABILITY.pdf (accessed Jan 2021).
- [57] Global Solar Atlas, 2021. Available at https://globalsolaratlas.info/map (accessed Apr 2021).
- [58] Great Britain. The Climate Change Act 2008 (2050 target amendment) order 2019, 2019. Available at https://www.legislation.gov.uk/uksi/2019/1056/contents/made (accessed Jan 2021).

- [59] M. A. Green, E. D. Dunlop, J. Hohl-Ebinger, M. Yoshita, N. Kopidakis, and A. W. Ho-Baillie. Solar cell efficiency tables (version 55). *Progress in photovoltaics*, 28 (1):3–15, 2020. doi:10.1002/pip.3228.
- [60] A. Gregorio and L. Jansen. A new concept for a land-cover classification system. *The Land*, 2:55–65, 01 1998.
- [61] A. Hastings, M. J. Tallis, E. Casella, R. W. Matthews, P. A. Henshall, S. Milner, P. Smith, and G. Taylor. The technical potential of Great Britain to produce lignocellulosic biomass for bioenergy in current and future climates. *Global change biology. Bioenergy*, 6(2):108–122, 2014. doi:10.1111/gcbb.12103.
- [62] D. J. Immerzeel, P. A. Verweij, F. Hilst, and A. P. C. Faaij. Biodiversity impacts of bioenergy crop production: a state-of-the-art review. *Global change biology*. *Bioenergy*, 6(3):183–209, 2014. doi:10.1111/gcbb.12067.
- [63] Independent. Biomass power: Is the UK's second-largest source of renewable energy sustainable?, 2021. Available at https://www.independent.co.uk/climatechange/news/biomass-power-renewable-energy-sustainable-b1805168.html (accessed Apr 2021).
- [64] O. Inderwildi, C. Zhang, X. Wang, and M. Kraft. The impact of intelligent cyberphysical systems on the decarbonization of energy. *Energy & Environmental Science*, 13, 02 2020. doi:10.1039/C9EE01919G.
- [65] Integrated Breeding Platform, 2020. Available at https://www.integratedbreeding .net (accessed Jan 2021).
- [66] Integrated Breeding Platform. Crop Ontology, 2021. Available at https://www.cr opontology.org/ (accessed Jan 2021).
- [67] International Renewable Energy Agency (IRENA). Future of solar pholtovoltaic, 2019. Available at https://irena.org/-/media/Files/IRENA/Agency/Publication/20 19/Nov/IRENA_Future_of_Solar_PV_2019.pdf (accessed Apr 2021).
- [68] International Renewable Energy Agency (IRENA). Renewable power generation costs in 2019, 2020. Available at https://www.irena.org/-/media/Files/IRENA/Age ncy/Publication/2020/Jun/IRENA_Power_Generation_Costs_2019.pdf (accessed Apr 2021).
- [69] Internet Engineering Task Force (IETF). The GeoJSON format, 2016. Available at https://tools.ietf.org/html/rfc7946 (accessed Mar 2021).
- [70] Jena-JDBC. A SPARQL over JDBC driver framework, 2011. Available at https: //jena.apache.org/documentation/jdbc (accessed Jun 2021).
- [71] 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).

- [72] 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 ecoindustrial parks. In *Proceedings of the Second International Conference on Internet* of Things, Data and Cloud Computing, 2017. doi:10.1145/3018896.3025155.
- [73] 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.
- [74] S. C. J. Lim, Y. Liu, and Y. Chen. Ontology in Design Engineering: Status and Challenges. *Proceedings of the 20th International Conference on Engineering Design (ICED15)*, 2015. Available at https://www.designsociety.org/publication/37 688/ontology+in+design+engineering%3a+status+and+challenges (accessed Apr 2021).
- [75] D. MacKay. Sustainable Energy without the hot air. UIT Cambridge Ltd, Cambridge, 2009. ISBN 1906860017.
- [76] D. MacKay. Solar energy in the context of energy use, energy transportation and energy storage. *Phil. Trans. R. Soc. A.*, 371, 2013. doi:10.1098/rsta.2011.0431.
- [77] D. D. Maio and D. Turley. Lignocellulosic feedstock in the UK. NNFCC, 2014. Available at https://www.nnfcc.co.uk/files/mydocs/LBNet%20Lignocellulosic% 20feedstockin%20the%20UK_Nov%202014.pdf (accessed Apr 2021).
- [78] Mapbox, 2021. Available at https://www.mapbox.com (accessed Apr 2021).
- [79] Maptiler. OSGB 1936 / British National Grid United Kingdom Ordnance Survey epsg:27700, 2019. Available at https://epsg.io/27700 (accessed Apr 2021).
- [80] Maptiler. WGS 84 WGS84 World Geodetic System 1984, used in GPS EPSG:4326, 2019. Available at https://epsg.io/4326 (accessed Apr 2021).
- [81] B. Matek and K. Gawell. The benefits of baseload renewables: A misunderstood energy technology. *The Electricity Journal*, 28(2):101–112, 2015. doi:10.1016/j.tej.2015.02.001.
- [82] Ministry of Housing, Communities & Local Government (United Kingdom). Land use in England, 2018. Available at https://www.gov.uk/government/statistics/landuse-in-england-2018 (accessed Feb 2021).
- [83] Ministry of Housing, Communities & Local Government (United Kingdom), 2021. Available at https://www.gov.uk/government/organisations/ministry-of-housingcommunities-and-local-government (accessed Apr 2021).
- [84] 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., 60(12):6155–6166, 2020. doi:10.1021/acs.jcim.0c01145.

- [85] National Academy of Sciences, National Academy of Engineering and Institute of Medicine. *Population Summit of the World's Scientific Academies*. National Academies Press, Washington, 1986. ISBN 0309582628.
- [86] P. Nayak, S. Mahesh, H. Snaith, and D. Cahen. Photovoltaic solar cell technologies: analysing the state of the art. *Nature Reviews Materials*, 4:1, 03 2019. doi:10.1038/s41578-019-0097-0.
- [87] Office for National Statistics, 2021. Available at https://www.ons.gov.uk (accessed Apr 2021).
- [88] Office for National Statistics. Geography linked data, 2021. Available at http: //statistics.data.gov.uk/home (accessed Apr 2021).
- [89] Open Geospatial Consortium. GeoSPARQL a geographic query language for RDF data, 2012. Available at https://www.ogc.org/standards/geosparql (accessed Mar 2021).
- [90] Open Geospatial Consortium. Geography Markup Language (GML), 2021. Available at https://www.ogc.org/standards/gml (accessed Mar 2021).
- [91] Open Street Map, 2021. Available at https://www.openstreetmap.org/#map=7/52. 366/-1.620 (accessed Apr 2021).
- [92] Ordnance Survey. OS MasterMap topography layer, 2021. Available at https://ww w.ordnancesurvey.co.uk/business-government/products/mastermap-topography (accessed Feb 2021).
- [93] 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.
- [94] 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.
- [95] QGIS, 2021. Available at https://www.qgis.org/en/site/ (accessed Apr 2021).
- [96] Renewable UK. Wind energy, 2018. Available at https://www.renewableuk.com/ page/WindEnergy (accessed Apr 2021).
- [97] H. Rijgersberg. OM Ontology of units of Measure, 2021. Available at https: //github.com/HajoRijgersberg/OM (accessed Apr 2021).
- [98] H. Rijgersberg. OM Ontology of units of Measure (forked), 2021. Available at https://github.com/cambridge-cares/OM (accessed Jul 2021).

- [99] Rural Grants & Payments. Basic Payments Scheme, 2020. Available at https: //www.gov.uk/government/collections/basic-payment-scheme (accessed Jan 2021).
- [100] Rural Grants & Payments. Countryside Stewardship, 2021. Available at https: //www.gov.uk/government/collections/countryside-stewardship (accessed Jan 2021).
- [101] Rural Payments Agency. Crop Map of England, 2019. Available at https://data.g ov.uk/dataset/8c5b635f-9b23-4f32-b12a-c080e3f455d0/crop-map-of-englandcrome-2019 (accessed Jan 2021).
- [102] Rural Payments Agency. Rural Payments land use codes, 2019. Available at https: //www.gov.uk/guidance/rural-payments-land-use-codes-2019 (accessed Jan 2021).
- [103] K. Sapkota, P. Raju, C. Chapman, W. Byrne, and L. Melville. Bioenergy ontology for automatic pathway generation. *International Journal of Knowledge Engineering-IACSIT*, 1:1–8, 01 2015. doi:10.7763/IJKE.2015.V1.1.
- [104] Scottish Environment Protection Agency. Waste from all sources summary document and commentary text, 2018. Available at https://www.sepa.org.uk/medi a/500273/waste-from-all-sources-summary-document-and-commentary-text-2018.pdf (accessed Jan 2021).
- [105] Select Committee on Science & Technology, House of Lords. Fourth report, 2004. Available at https://publications.parliament.uk/pa/ld200304/ldselect/ldsctech/126 /12618.htm (accessed Apr 2021).
- [106] SSE Renewables, 2021. Available at https://www.sserenewables.com (accessed Apr 2021).
- [107] Statista. Load factor of electricity from solar PV in the United Kingdom (UK) 2010 to 2019, 2021. Available at https://www.statista.com/statistics/555697/solarelectricity-load-factor-uk/ (accessed Apr 2021).
- [108] Statista. Offshore wind capacity H1 2020, by country, 2021. Available at https: //www.statista.com/statistics/264264/capacity-of-offshore-wind-power-plantsworldwide-2010/ (accessed Apr 2021).
- [109] Swirrl. Publish My Data, 2021. Available at https://www.swirrl.com (accessed Apr 2021).
- [110] D. J. Tenenbaum. Food vs. fuel: Diversion of crops could cause more hunger. Environmental health perspectives, 116(6):A254–A257, 2008. doi:10.1289/ehp.116-a254.
- [111] The Earth Observatory. World of change: Global temperatures, 2020. Available at https://earthobservatory.nasa.gov/world-of-change/global-temperatures (accessed Jan 2021).

- [112] The European Space Agency. Sentinel-1 radar vision for Copernicus, 2021. Available at http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentin el-1 (accessed Apr 2021).
- [113] The European Space Agency. Sentinel-2 colour vision for Copernicus, 2021. Available at http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentin el-2 (accessed Apr 2021).
- [114] The Food and Agriculture Organization of the United Nations (FAO). FAOSTAT, 2021. Available at http://www.fao.org/faostat/en/#data/QC (accessed Apr 2021).
- [115] The National Archives. Open government licence for public sector information, 2021. Available at http://www.nationalarchives.gov.uk/doc/open-governmentlicence/version/3/ (accessed Apr 2021).
- [116] The Scottish Government, 2021. Available at https://www.gov.scot (accessed Apr 2021).
- [117] The World Avatar. J-Park Simulator, 2021. Available at http://www.jparksimulat or.com (accessed Dec 2020).
- [118] The World Avatar. OntoSpecies, 2021. Available at http://www.theworldavatar.c om/ontology/ontospecies/OntoSpecies.owl (accessed Feb 2021).
- [119] TNO Biomass and Circular Technologies. Database for the physico-chemical composition of (treated) lignocellulosic biomass, micro- and macroalgae, various feedstocks for biogas production and biochar, 2021. Available at https://phyllis.nl (accessed Apr 2021).
- [120] N. Trokanas, M. Bussemaker, E. Velliou, H. Tokos, and F. Cecelja. Bionto: An ontology for biomass and biorefining technologies. In K. V. Gernaey, J. K. Huusom, and R. Gani, editors, 12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering, volume 37 of Computer Aided Chemical Engineering, pages 959–964. Elsevier, 2015. doi:10.1016/B978-0-444-63577-8.50005-X.
- [121] UK Centre for Ecology & Hydrology. Land Cover plus: Crops, 2021. Available at https://www.ceh.ac.uk/crops2015 (accessed Feb 2021).
- [122] UK Government. UK enshrines new target in law to slash emissions by 78% by 2035, 2021. Available at https://www.gov.uk/government/news/uk-enshrinesnew-target-in-law-to-slash-emissions-by-78-by-2035 (accessed Apr 2021).
- [123] United Nations. 68% of the world population projected to live in urban areas by 2050, says UN, 2018. Available at https://www.un.org/development/desa/en/new s/population/2018-revision-of-world-urbanization-prospects.html (accessed Mar 2021).

- [124] United Nations. Global issues population, 2019. Available at https://www.un.org/en/global-issues/population (accessed Mar 2021).
- [125] United Nations. The Paris Agreement, 2021. Available at https://www.un.org/en/ climatechange/paris-agreement (accessed Jan 2021).
- [126] 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/.
- [127] W3C. An introduction to multilingual web addresses, 2008. Available at https: //www.w3.org/International/articles/idn-and-iri/Overview.en (accessed Apr 2021).
- [128] W3C. OWL Web Ontology Language Overview, 2012. Available at https://www. w3.org/TR/owl-features/ (accessed Jul 2021).
- [129] W3C. SPARQL 1.1 Overview, 2013. Available at https://www.w3.org/TR/sparql 11-overview/ (accessed Mar 2021).
- [130] W3C. SPARQL 1.1 Update, 2013. Available at https://www.w3.org/TR/sparql11update/ (accessed Jun 2021).
- [131] W3C. RDF 1.1 Turtle, Terse RDF Triple Language, 2014. Available at https: //www.w3.org/TR/turtle (accessed Jul 2021).
- [132] W3C. Semantic Web, 2015. Available at https://www.w3.org/standards/semanti cweb/ (accessed Mar 2021).
- [133] Waste and Resources Action Programme (WRAP). Household Food and Drink Waste in the United Kingdom, 2012. Available at https://wrap.org.uk/resources/re port/household-food-and-drink-waste-united-kingdom-2012 (accessed Apr 2021).
- [134] H. Weber and J. D. Sciubba. The effect of population growth on the environment: Evidence from European regions. *European journal of population*, 35(2):379–402, 2019. doi:10.1007/s10680-018-9486-0.
- [135] J. Whitaker. pyproj 3.0.1, 2021. Available at https://pypi.org/project/pyproj/ (accessed Mar 2021).
- [136] YARA International. Sunflower crop nutrition, 2018. Available at https://www.ya ra.co.uk/crop-nutrition/novel-crops/sunflower/ (accessed Apr 2021).
- [137] Ystadegau ar gyfer Cymru Statistics for Wales. Local authority municipal waste management, 2018-19, 2019. Available at https://gov.wales/sites/default/files/stati stics-and-research/2019-10/local-authority-municipal-waste-management-april-2018-march-2019-253.pdf (accessed Jan 2021).
- [138] C. Zhang, A. Romagnoli, L. Zhou, and M. Kraft. Knowledge management of ecoindustrial park for efficient energy utilization through ontology-based approach. *Applied Energy*, 204:1412–1421, 2017. doi:10.1016/j.apenergy.2017.03.130.

- [139] X. Zhou, A. Eibeck, M. Q. Lim, N. Krdzavac, and M. Kraft. An agent composition framework for the J-Park Simulator – A knowledge graph for the process industry. *Computers & Chemical Engineering*, 130:108–122, 2019. doi:10.1016/j.compchemeng.2019.106577.
- [140] 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.
- [141] X. Zhou, D. Nurkowski, S. Mosbach, J. Akroyd, and M. Kraft. A question answering system for chemistry. *Journal of Chemical Information and Modeling*, 2021. In press. Preprint available at https://como.ceb.cam.ac.uk/preprints/266/.