

Lappeenranta University of Technology

School of Engineering Science

Erasmus Mundus Master's Program in Pervasive Computing and Communications for  
Sustainable Developments (PERCCOM)

**Md Anisul Islam**

**LEVERAGING KNOWLEDGE GRAPH FOR UK ENVIRONMENTAL  
LEGISLATION RELATED TO BUILDING CONSTRUCTION AND  
MAINTENANCE**

**Examiners:** Prof. Jari Porras (Lappeenranta University of Technology)  
Prof. Eric Rondeau (University of Lorraine)  
Assoc. Prof Karl Andersson (Lulea University of Technology)

**Supervisors:** Dr. Ah-Lian Kor (Leeds Beckett University)  
Dr. Brandon Bennett (University of Leeds)

**This thesis is prepared as part of an European Erasmus Mundus Programme  
PERCCOM – PERvasive Computing & COMmunication for sustainable  
development**



Co-funded by the  
Erasmus+ Programme  
of the European Union

Successful defense of this thesis is obligatory for graduation with the following national diplomas:

- Master in Complex Systems Engineering (University of Lorraine)
- Master of Science in Technology (Lappeenranta University of Technology)
- Master of Science in Computer Science and Engineering, specialiaztion in Pervasive Computing and Communications for Sustainable Development (Luleå University of Technology)

## **ABSTRACT**

Lappeenranta University of Technology

LUT School of Engineering Science

Erasmus Mundus Master's Programme in Pervasive Computing & Communications for Sustainable Development (PERCCOM)

Md Anisul Islam

## **LEVERAGING KNOWLEDGE GRAPH FOR UK ENVIRONMENTAL LEGISLATION RELATED TO BUILDING CONSTRUCTION AND MAINTENANCE**

2019

45 pages, 19 figures, 2 tables

Keywords: Knowledge Graph, Linked Data, RDF, Legislation

Carbon footprint generated by built environment is one of the biggest concerns in UK. Relevant legislations, act and guidelines for building design, construction and maintenance that interests environmental aspects are created and curated by many authorities from Government in UK. Stakeholders often find it difficult to access these legal and regulatory information across different jurisdiction since there is no common platform. This research proposes a framework for developing a domain-neutral platform where heterogeneous data will be highly interlinked in machine readable fashion paving further opportunities such as making general queries on, training statistical models for decision making and so on. Our proposed methodology uses domain ontologies and further applied in the development of a proof of concept system, called Knowledge Graph for UK Environmental Legislations related to Building Construction and Maintenance (UKENV-KG). UKENV-KG enables legislations and guidelines to be properly annotated through use of W3C-based semantic technologies. The report presents the methodology, the system architecture of proposed framework with implementation. A discussion on further possibilities is made at very end of the thesis.

## ACKNOWLEDGEMENTS

This Master's thesis report is supported by PERCCOM Erasmus Mundus Program [1] and funded by European Commission.

My heartiest gratitude to everyone who supported me accomplishing this Master Thesis work. Especially I would like to thank Dr. Ah-Lian Kor, Dr. Brandon Bennett and Prof. Colin Pattinson for patiently guiding me.

Leeds, 2019

*Md Anisul Islam*

# TABLE OF CONTENTS

<b>1</b>	<b>INTRODUCTION .....</b>	<b>10</b>
1.1	MOTIVATION .....	11
1.2	PROBLEM DEFINITION .....	12
1.3	AIM, RESEARCH QUESTION AND OBJECTIVES .....	12
1.4	DELIMITATION .....	13
1.5	CONTRIBUTION .....	14
1.6	ORGANIZATION OF THE THESIS .....	14
<b>2</b>	<b>REVIEW OF RELATED WORK.....</b>	<b>15</b>
2.1	KNOWLEDGE REPRESENTATION & ONTOLOGY .....	15
2.2	KNOWLEDGE GRAPH .....	15
2.3	APPLICATION OF KNOWLEDGE GRAPH FOR LEGISLATION DOMAIN .....	18
2.4	KNOWLEDGE CONSTRUCTION .....	19
2.5	COMPETENCY QUESTION .....	22
<b>3</b>	<b>METHODOLOGY .....</b>	<b>23</b>
3.1	SYSTEM DEVELOPMENT LIFECYCLE .....	23
3.2	SYSTEM ARCHITECTURE .....	24
<b>4</b>	<b>REQUIREMENTS AND DESIGN.....</b>	<b>26</b>
4.1	REQUIREMENTS .....	26
4.2	DESIGN .....	28
<b>5</b>	<b>UKENV-KG SYSTEM IMPLEMENTATION AND EVALUATION .....</b>	<b>31</b>
5.1	EXTENDING DCAT VOCABULARY.....	31
5.2	DATA ANNOTATION, STORAGE AND QUERYING .....	33
5.3	EVALUATION .....	34
5.3.1	<i>Use Cases.....</i>	<i>34</i>
5.3.2	<i>Performance Analysis.....</i>	<i>35</i>
5.4	SUSTAINABILITY ANALYSIS .....	35
<b>6</b>	<b>CONCLUSION AND FUTURE WORK .....</b>	<b>39</b>

6.1	CONCLUSIONS.....	39
6.2	FUTURE WORK .....	39
<b>7</b>	<b>REFERENCES .....</b>	<b>41</b>

## LIST OF FIGURES

<b>Fig. 1.</b> Carbon impact of the built environment in UK, 2014.....	10
<b>Fig. 2.</b> A RDF statement .....	15
<b>Fig. 3.</b> Knowledge construction and maintenance lifecycle [45] .....	20
<b>Fig. 4.</b> Architecture of Knowledge Graph Construction (bottom up approach) [55] .....	21
<b>Fig. 5.</b> CQ feature hierarchy. ....	22
<b>Fig. 6.</b> Iterative Lifecycle Model. ....	23
<b>Fig. 7.</b> System Architecture. ....	24
<b>Fig. 8.</b> Environmental guidelines, legislations and policies related to building construction and maintenance. ....	26
<b>Fig. 9.</b> A part of domestic building service compliance guideline. ....	27
<b>Fig. 10.</b> Two layer of Knowledge Graph. ....	28
<b>Fig. 11.</b> Overview of DCAT (Data Catalogue Vocabulary) .....	29
<b>Fig. 12.</b> Onto Visualization of DCAT vocabulary.....	30
<b>Fig. 13.</b> Ontology classes in Protégé. ....	31
<b>Fig. 14.</b> Enhancement of DCAT vocabulary to meet requirements. ....	32
<b>Fig. 15.</b> Segment of RDF Turtle based on modified vocabulary.....	33
<b>Fig. 16.</b> GraphDB as RDF storage and SPARQL query engine. ....	34
<b>Fig. 17.</b> A sample SPARQL query made on Knowledge Graph using GraphDB query engine.....	<b>Error! Bookmark not defined.</b>
<b>Fig. 18.</b> Three pillars of sustainability. ....	36
<b>Fig. 19.</b> Sustainability analysis pentagon. ....	37

## LIST OF TABLES

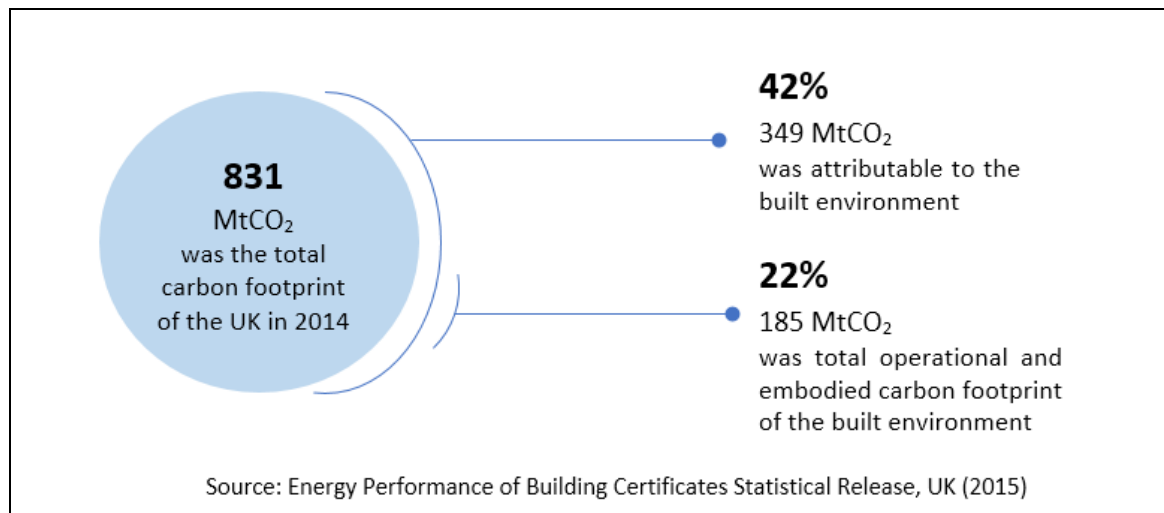
Table 1. RDF Datasets vs. Linked Data vs. Knowledge Graph [33].....	17
Table 2. Knowledge Graph for Web search vs. for enterprise [33].....	18

## LIST OF SYMBOLS AND ABBREVIATIONS

<b>KG</b>	Knowledge Graph
<b>KB</b>	Knowledge Base
<b>CO<sub>2</sub></b>	Carbon Dioxide
<b>WWW</b>	World Wide Web
<b>AI</b>	Artificial Intelligence
<b>OWL</b>	Web Ontology Language
<b>RDF</b>	Resource Descriptive Framework
<b>W3C</b>	World Wide Web Consortium
<b>LOD</b>	Linked Open Data
<b>PERCCOM</b>	Pervasive Computing and Communications for Sustainable Development
<b>SPARQL</b>	Protocol and RDF Query Language
<b>URI</b>	Universal Resource Identifier
<b>API</b>	Application Program Interface
<b>MVC</b>	Model, View, Controller
<b>CQOA</b>	Competency Question-driven Ontology Authoring
<b>REST</b>	Representational State Transfer
<b>IRI</b>	Internationalized Resource Identifier
<b>XML</b>	Extensible Markup Language

# 1 INTRODUCTION

Building construction and maintenance put a significant impact on total Carbon Footprint of UK. Around 40% of total carbon emission comes from built environment [2]. Government aims to reduce the cost of construction by one-third while halving carbon footprint in construction industry by 2025 [3]. It's a big challenge for stakeholders in this industry to meet the new requirements and guidelines. According to UK Low Carbon Transition Plan [4], national goal is to reduce carbon emission by 80% compared to carbon emission level of year 1990 within year 2050. Also aim is to have a reduction of at least 34% by 2020.



**Fig. 1.** Carbon impact of the built environment in UK, 2014

To meet obligations under the 2008 Climate Change Act [5], the Green Construction Board, UK has recommended a low carbon route map targeting year 2050 where targets for operational and embodied carbon should be lowered to 113 MtCO<sub>2</sub> by 2025 and 45 MtCO<sub>2</sub> by 2050.

To lower the carbon footprint of a built environment, the goal can be divided into the following: lower down carbon embodied in new building construction and minimizing emission from existing buildings (for both domestic and non-domestic buildings). In 2016, 10% of nation's carbon footprint came from building construction activities and more 10% came only from heating domestic building [6]. Relevant legislations acts and guidelines for

building design, constructions and maintenance that interest environmental aspects are created and curated by different authorities from Government results in heterogeneous data-sources while integrating them in a single platform. To provide stakeholders (engineers, constructors, designers, architects, dwellers etc.) a better access to these legal and regulatory information across different jurisdictions, a domain-neutral common platform is required where structured data will be highly interlinked. From consumers point of view unifying heterogeneous data-sources under same hood in machine readable fashion paves further opportunities such statistical models can be trained to take decisions [7]. In spirit of providing such platform, a knowledge graph development framework for UK Environmental Legislation for Building Construction and Maintenance (UKENV-KG) is proposed based on W3C semantic technologies.

## **1.1 Motivation**

As a knowledge representation technique, Knowledge Graphs [8] have become prevalent in both industry and academia recently. In 2012 Google first introduced the idea of Knowledge Graph that understands real-world entities and the relationship between those entities [9]. The graph enhances Google search by disambiguating search queries and by explorative search suggestions. Also, it provides short summary of topics with structured information.

Knowledge Graph is not only useful in Web Search paradigm but also handy in other systems and applications, including enterprise information management [10]. It is one of the most efficient and effective knowledge integration approaches [11]. Knowledge graphs provide a new paradigm and powerful platform for implementing enterprise knowledge bases using organization knowledge schemas or ontologies (built on basis of business vocabularies). In fact, Knowledge Graphs provide better schema evaluation capabilities comparing to relational platform [12].

Linked data and Knowledge Graphs are often referred to same context while there is significant difference but both approaches are useful when it comes to handling data coming from heterogeneous sources [13]. While working with knowledge representation in

legal paradigm, it is highly likely to confront cases where user needs to deal with heterogeneous data. The Lynx, a European project is currently working building an ecosystem of better managed compliance service based on Knowledge Graph [14]. This project aims to integrate and link heterogeneous compliance data sources in different domain of interest (legislation, case law, standards etc.).

This research focuses on proposing a framework through which the idea of Knowledge Graph can be exploited to create common service which integrates information from different sub-domains of UK Environmental Legislation paradigm related to building design, construction and maintenance.

## **1.2 Problem Definition**

Linked Data or Knowledge Graph approach has been commonly used in Web context [15]. Value of interlinkedness of data is not only limited in complex and open ended systems such as the Web. Managing enterprise information using Linked Data or Knowledge Graph is fairly new and practiced by big players in industry [10]. Inadequate research has been done scoping Linked Data for legal domain.

While working with legal information, time, intervals, and versioning are prime aspects to care. Several mechanisms have been discussed in literature to handle this temporal issue in querying and management of RDF data [16], [17]. Scope of this research is to consider temporal issue in curating linked data in legal domain.

## **1.3 Aim, Research Question and Objectives**

Aim of this research is to propose a framework that unifies knowledge collected from different authorities who are concerned with UK environmental legislation related to building construction & maintenance and represent the knowledge in form knowledge graph.

Following section presents the research questions. Also there are corresponding objectives to each research questions which should be achieved in this research scope.

**RQ1. What are the approaches for handling heterogenous linked data from legal domain?**

**RO1. *Investigation of state-of-the-art***

**RO1.1** Investigation of heterogenous data management

**RO1.2** Investigation of entities in legal domain

**RQ2. How can Knowledge Graph be used in a non-web domain?**

**RO2. *Investigation of scope of Knowledge Graph in different domain***

**RO2.1** Build Knowledge Graph for legal domain

**RO2.2** Consume from the knowledge representation

**RO2.3** Evaluate performance of implementation

## **1.4 Delimitation**

In order to implement a platform combining information from different jurisdictions related to environmental legislation in building construction and maintenance paradigm, a small scope: Domestic Building Service Compliance Guide UK [18] is considered due to the expanse of the entire domain.

Decisions on selection of tools and technologies for implementation is made based on literature study. To create linked data, Resource Description Framework (RDF) is chosen [15]. Based on the performance analysis in literature, SPARQL is used as querying engine [19] that comes with GraphDB graph database [20]. GraphDB is built on top of RDF4J java framework which is widely used in handling RDF data [21].

GraphDB APIs are used to query interlinked data using RDF4J connector with the created web services. RESTful Web services are created with Java Spring MVC Framework and Maven [22].

## 1.5 Contribution

The following are the key contribution of the work:

1. Investigation of different approaches to represent and use heterogenous data coming from various sources
2. Exploitation of Linked Data and Knowledge Graph to build a framework that benefits building constructions and maintenance stakeholders in UK
3. Impact and Sustainability Analysis of the proposed framework

## 1.6 Organization of the Thesis

The structure of this manuscript is as follows:

- **Chapter 1: Introduction** provides an understanding of the background, motivation, objectives and contribution of the work presented in the thesis
- **Chapter 2: Review of Related Work** covers a broad study of Knowledge representation, use of Knowledge Graph as knowledge representation, Knowledge graph for Web vs Enterprise, Knowledge Graph construction approach and how evaluation and maintenance is done for Knowledge Graphs
- **Chapter 3: Methodology** presents reserch methodologies followed, system development lifecycle and lifecycle accounting methodology for Software
- **Chapter 4: Results and Discussion** provides the performance evaluation and sustainability analysis of the work
- **Chapter 5: Conclusion and Future Work** presents the outcome of the thesis and possible direction of the future work

## 2 REVIEW OF RELATED WORK

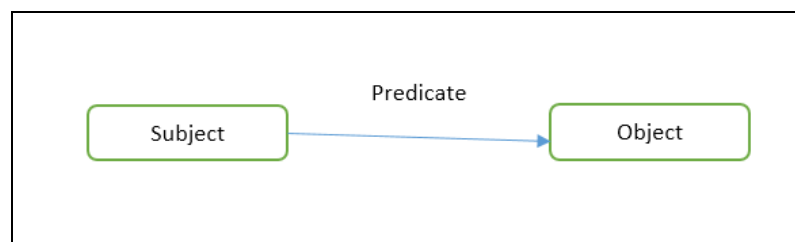
This Chapter is divided into sub sections to discuss the following topics: (a) Knowledge Representation and Ontology, (b) Knowledge Graph, (c) Knowledge Construction, (d) Application of Knowledge Graph for the Legislation Domain, (e) Competency Question.

### 2.1 Knowledge Representation & Ontology

Knowledge Representation is a prominent field in Artificial Intelligence (AI) which helps to represent information in a manner that allows a computer to reason automatically with relevant information for solving complex problems such as decision support [23]. Since the beginning of 1990s, Ontologies have become a topic among Artificial Intelligence research communities while it has been widely used in fields such as information retrieval on the Internet and intelligent information integration [24], knowledge management [25] and so on. Ontologies have widely been used as a schema in many researches regarding to knowledge management [26], [27].

### 2.2 Knowledge Graph

Knowledge graph inherits many classic Knowledge Representation formalisms provide visualizations to represent entities and their relationships [28]. After Quillian first introduced Semantic Networks in his semantic memory model [29], many variants of Semantic Networks have been proposed. Semantic Networks are relatively easy to use and maintain compared to formal knowledge representation and reasoning whereas at times it poses some challenges due to the lack of formal syntax and semantics which fails to preserve the original meaning in many cases [10].



**Fig. 2.** A RDF statement

RDF (Resource Descriptive Framework) is one of the modern standards from W3C which addresses few critical issues in Semantic Networks [30]. Although RDF does not address all the shortcomings of Semantic Networks, it clearly addresses the formal syntax and semantics issue as user can clearly define semantics as per RDF specifications. There are shortcomings of RDF as well since it does not address the concept where OWL (Web Ontology Language) has this flexibility [31], [32]. RDF and OWL are considered as the construction block of Knowledge Graph.

To speed up the development process of Knowledge Graph, existing vocabularies are used in most of the occasions. The idea behind reusing existing vocabularies is to enable optimal reuse of the existing work [33]. At present there are many vocabularies available that meet the requirement to develop knowledge graph in Enterprise Level. Organization Ontology is one of them which supports publishing of organizational information across variable domains [34]. GoodRelations ontology defines terms used in e-commerce scenarios [35]. Another famous ontology in social networking paradigm is Friend-of-a-friend (FOAF) ontology that describes people, their activities and relationship with other people [36].

A big leap was taken by Google, Yahoo! and Bing altogether with the introduction of *schema.org* in early 2011 [37]. All the relevant parties agreed upon a collection of terms that would be used to markup HTML pages in order to improve search results. The data model used in *schema.org* is derived from RDF Schema. It includes a huge set of properties for different domains.

Apart from Web paradigm, Knowledge Graph construction has gained much popularity in the Enterprise domain in recent time [38]. These Knowledge Graphs represents relationships between different classes of data comprising the organizations' data [39]. Usually in large organizations, data or knowledge is in a heterogeneous form. It means data or knowledge are in various formats such as Relational Databases (RDB), and Transaction logs, Web pages etc. These different representations require transformation into a format which is the building block of Knowledge Graphs. RDF is the most popular choice in this regard [33]. This representational transformation is known as a Data Lifting mechanism

[33] and there are various approaches available to perform this. One good specification is RDB2RDF[40] which is helpful for converting legacy data of an organization from relational database into linked data format. RDB2RDF is a W3C standard in data lifting.

Features	Pure RDF datasets	Linked Data	Knowledge Graph
Machine Readability	Y	Y	Y
Human Readability	NN	NN	Y
Data distribution	N	Y	NN
Inter-dataset linkage	L	Y	Y
Data integration	NN	NN	Y
Data consistency	NN	NN	Y
Reliability	NN	NN	Y
High quality	NN	NN	Y

Y: Yes; L: Limited; N: No; NN: Not Necessarily

Table 1. RDF Datasets vs. Linked Data vs. Knowledge Graph [33]

In many literature, Linked Data and Knowledge Graph are mentioned in similar context. In RDF datasets, there are data collections which is represented in RDF format, i.e. in subject, predicate and object format where subject and predicate are two identifiers (URIs) and an object can be any value or is an identifier itself. Linked Data refers to multiple RDF datasets which are developed, maintained and distributed independently though they use the same URIs to refer each other. In the case of a Knowledge Graph, there has to be an ontology as its schema and not necessarily needs to be linked with any other Knowledge Graph. Table 1. gives a comparison between RDF Datasets, Linked Data and Knowledge Graph.

The aim of the introduction of Knowledge Graph by Google in 2012 [9] was to enhance Google Web search results. Construction of such Knowledge Graphs is based on two prime technologies. One is using an ontology backed by non-complicated hierarchy of types similar to the collection of *schema.org*. Another is crowd sourced data gathering approach based on open encyclopedic sources and available APIs. Such sources for Google Knowledge Graphs are CIA World Factbook, Wikipedia, Freebase, Wikidata etc. [9].

Features	KG for Web searching	KG for enterprise
Data source	Distributed	Usually centralized
Openness of data	Open to public	Private
Size of data	Huge	Big
Data acquisition	Harder	Easier
Quality of data	Low	High
Ontology Language	Simple	More expressive
Knowledge	Generic	Domain specific

Table 2. Knowledge Graph for Web search vs. for enterprise [33]

In the IT industry, managing enterprise knowledge is one of the most challenging tasks. In most of the cases, there are many departments working under the same roof with different data infrastructures. To share informations from different sources, existing solutions are to build applications on top of relational databases that deal with data acquisition, curation, and maintenance of different schemas. Knowledge Graphs provide a new platform to create enterprise knowledge bases. This knowledge base is based on a common schema that is more effective and powerful. Generally backed by RDF, knowledge graphs provide better schema evolution capabilities with respect to relational platforms [12]. Table 2 gives a comparison of two use cases of knowledge graph (for web and enterprise).

### 2.3 Application of Knowledge Graph for Legislation Domain

Much work has not been done for building Knowledge Graph in Legislation domain. Lynx Project, funded by European Union’s Horizon 2020 Research and Innovation program, working on building the legal knowledge graph for the smart compliance service in multilingual Europe since December 2017 [41]. Aim of the project is to provide more effective way to access huge amounts of heterogeneous regulatory compliance documents in form of legal knowledge graph. Smart services for compliance include search, linking, recommendation, extraction, translation and summarization. The first pilot of the project is with the data protection law of European Union. Second pilot is to work with oil, gas energy law and last pilot project is to work with labor law.

Currently there is no knowledge base for UK environmental legislation related to building construction and maintenance. There is scope to contrast the scenario with Knowledge Graph built for Enterprises [42].

## **2.4 Knowledge Construction**

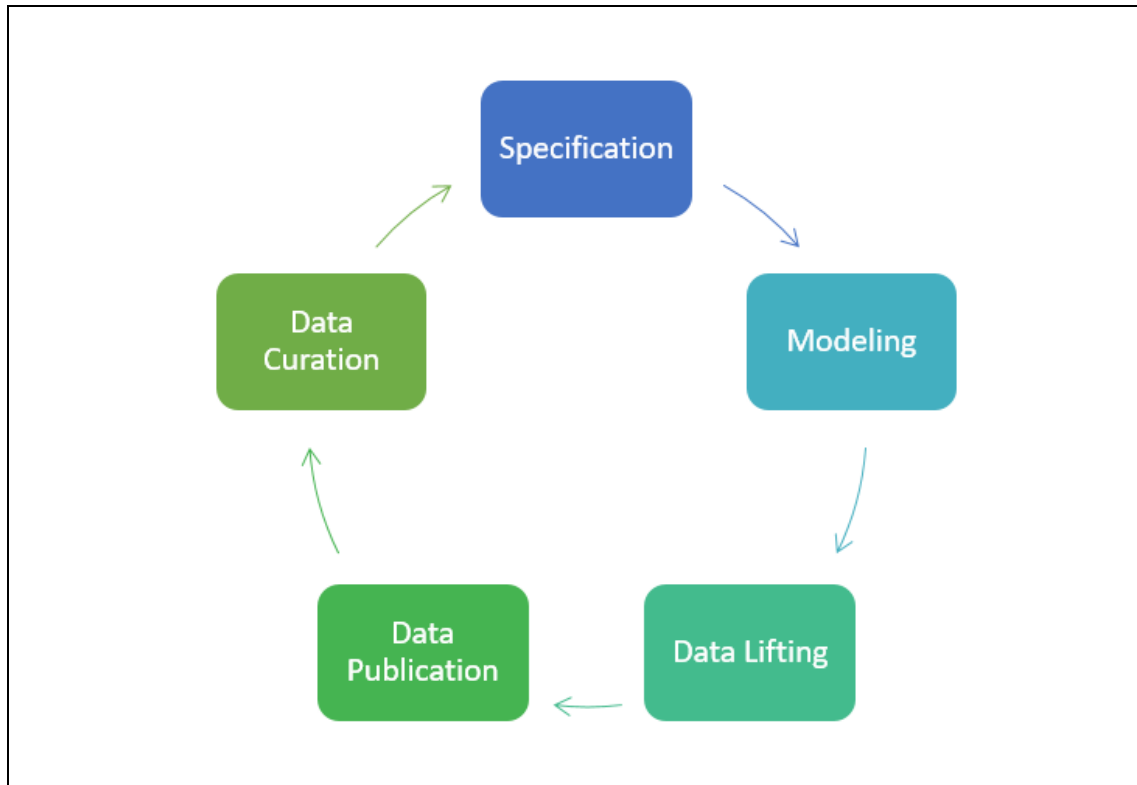
Completeness, data quality and accurateness are key features of a useful Knowledge Base. A Knowledge Graph is a special form of Knowledge Base. The construction method has been divided into four main groups [7]:

1. Curated Approach (RDF triples are created manually by a closed group of experts);
2. Collaborative Approach (triples are created manually by volunteers in an open group);
3. Automated Semi-Structured Approach (triples are extracted from semi-structured source using hand-crafted rules);
4. Automated Un-structured Approach (triples are extracted from unstructured sources automatically using machine learning and natural language processing techniques) [43].

In literature, a higher level overview has been given on related uses of Knowledge Graphs [44]:

1. Construction
2. Storage
3. Consumption

An abstract reference architecture for knowledge management in an organizational paradigm can be divided into three layers: 1) Knowledge Acquisition and Integration Layer, 2) Knowledge Storage Layer and 3) Knowledge Consumption Layer [44]. The Acquisition and Integration layer deals with ontology development and data lifting. The Consumption Layer incorporates analysis of use-cases of graph. It also includes understanding of entities and exploiting the knowledge through semantic search, question answering or query generation.



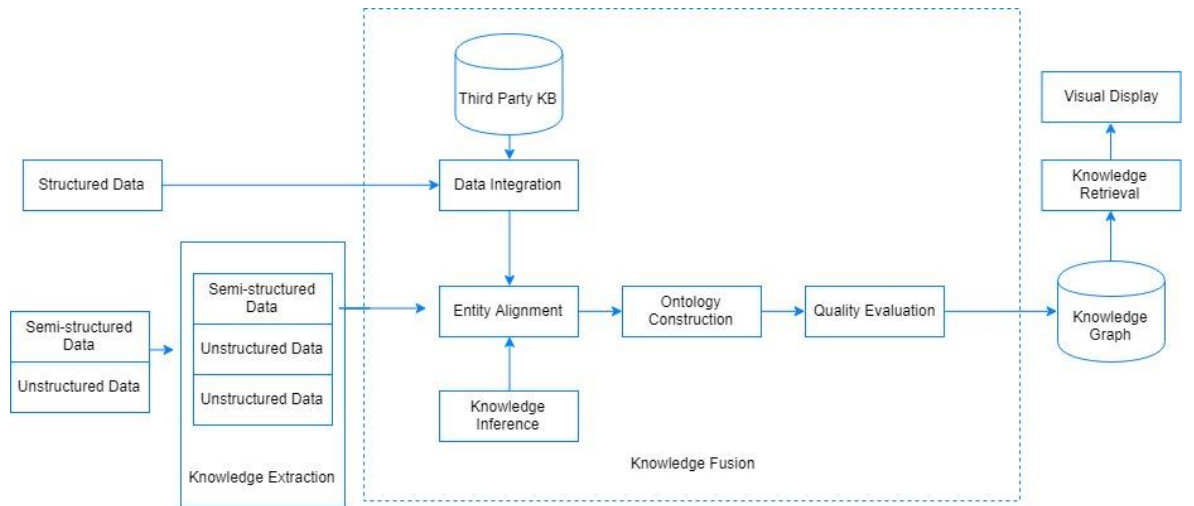
**Fig. 3.** Knowledge construction and maintenance lifecycle [45]

Knowledge construction and maintenance lifecycle is divided in five phases: Specification, Mideling, Data Lifting, Data Publication and Data Curation [45]. Specification phase mostly deals with scoping the use cases of application and identification and analysis of the data sources. In an organisational paradigm, identification of data source is straight forward. After identification, selected data sources needs to be compiled and schema of resources needs to be identified [46].

After data sources have been indentified and analysed, an ontology or vocabulary is required to model the domain of the data sources. One useful approach is to reuse any of the existing vocabularies which speeds up the development process [47]. Swoogle[48] and Linked Open Vocabularies (LOV)[49] are popular vocabulary repositories where vocabularies for various domain can be found and to be reused. If any existing vocabulary of interest can not be found, ontology needs to be created from scratch. There are a good number of methodologies to build ontologies from scratch and NeOn Methodology [47] is one of them.

In the transformation phase data source to be converted into RDF considering two important aspects: full conversion, so that all queries are possible to make and reflects target vocabulary or ontology structure as closely as possible [45]. There are many tools available for transforming data sources into RDF complying with given vocabulary. For CSV (Comma-separated values) and spreadsheets, RDF Extension of OpenRefine (former Google Refine) [50], XLWrap [51] is well used. An important step afterwards in this phase is linking which refers to creating links with external knowledge graphs. This link creation is done in possible ways: 1) Supervised link creation, 2) Unsupervised link creation [45].

Data publication refers to the publication of RDF data and metadata. There are numerous tools [46] available for RDF data storage: Jena [52], rdf4j [21], Virtuoso Universal Server [53] etc. The last phase of Knowledge graph construction and maintenance lifecycle includes Data Curation. Data curation implies cleaning and maintaining data for reuse over the course of time. There is much noise that hinders application in exploiting structured information collected from knowledge graph [54].



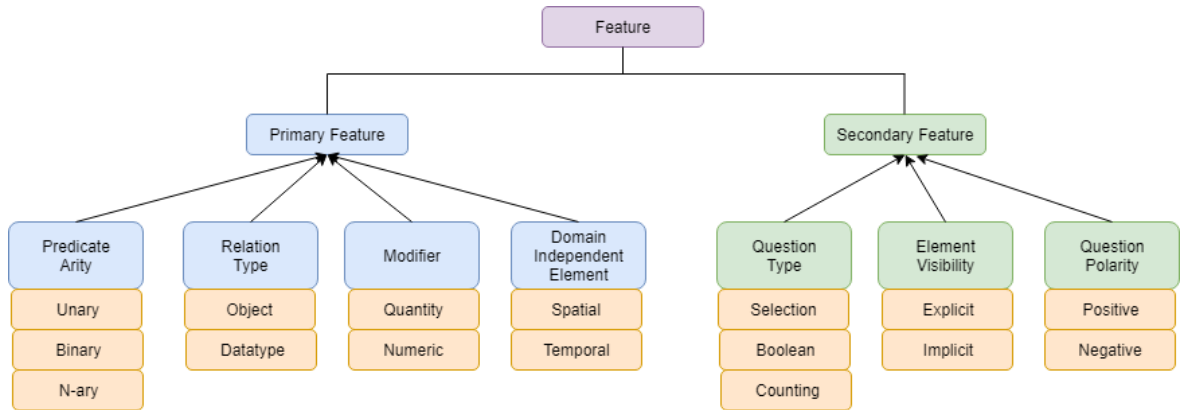
**Fig. 4.** Architecture of Knowledge Graph Construction (bottom up approach) [55]

In literature Knowledge Graph construction approach is divided into: 1) Top down and; 2) Bottom up [55]. In the top down approach, the vocabulary or ontology is created first and

then the knowledge instances are added into the knowledge base. In the bottom up approach, knowledge instances are extracted from existing Linked Open Data (LOD) or different knowledge resources. After fusion of knowledge instances, the top-level ontology is created. Fig.4 shows the bottom up approach of knowledge graph construction.

## 2.5 Competency Question

After constructions of an ontology, Competency Questions (CQs) are expressions that it must be able to answer [56]. CQs are regarded as functional requirement that must be statisfied by an ontology. In ontology authoring, CQs plays a vital role by capturing Scope, Content and Form of Evaluation. Also CQs helps authors to determine the granularity of ontology to identify the most important classes, properties and their relations [33].



**Fig. 5.** CQ feature hierarchy.

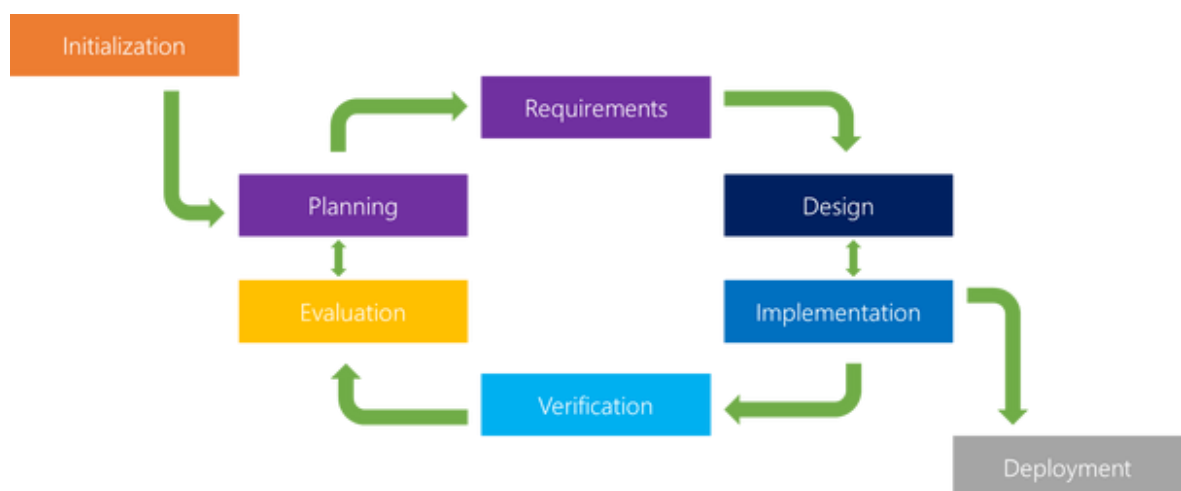
Competency Question-driven Ontology Authoring (CQOA) [57] leverages ideas of competency questions and testing driven software development. Analyzing different CQs patterns, Fig.5 shows a feature based-modelling hierarchy [58].

### 3 METHODOLOGY

Methodology discusses the systematic approach of the work, architectural design in detail. Purpose of research is to discover answers through application of scientific procedures aiming to find out the truth that has not been discovered yet [59]. Research methodology is a systematic effort to discover new information.

#### 3.1 System Development Lifecycle

The research methodology followed that has been followed is of iterative nature. The initialization the system development begins with multiple iterations of requirement analysis, designing, implementing and verifying implementation.



**Fig. 6.** Iterative Lifecycle Model.

Applying iterative methodology allows to frequent refinement of the process [60]. Fig.6 gives the overview of iterative lifecycle method that has been followed to design the framework and implementing the solution.

The development of the system is divided in four phases:

**Phase 1:** Requirement analysis, Ontology scoping and Data lifting

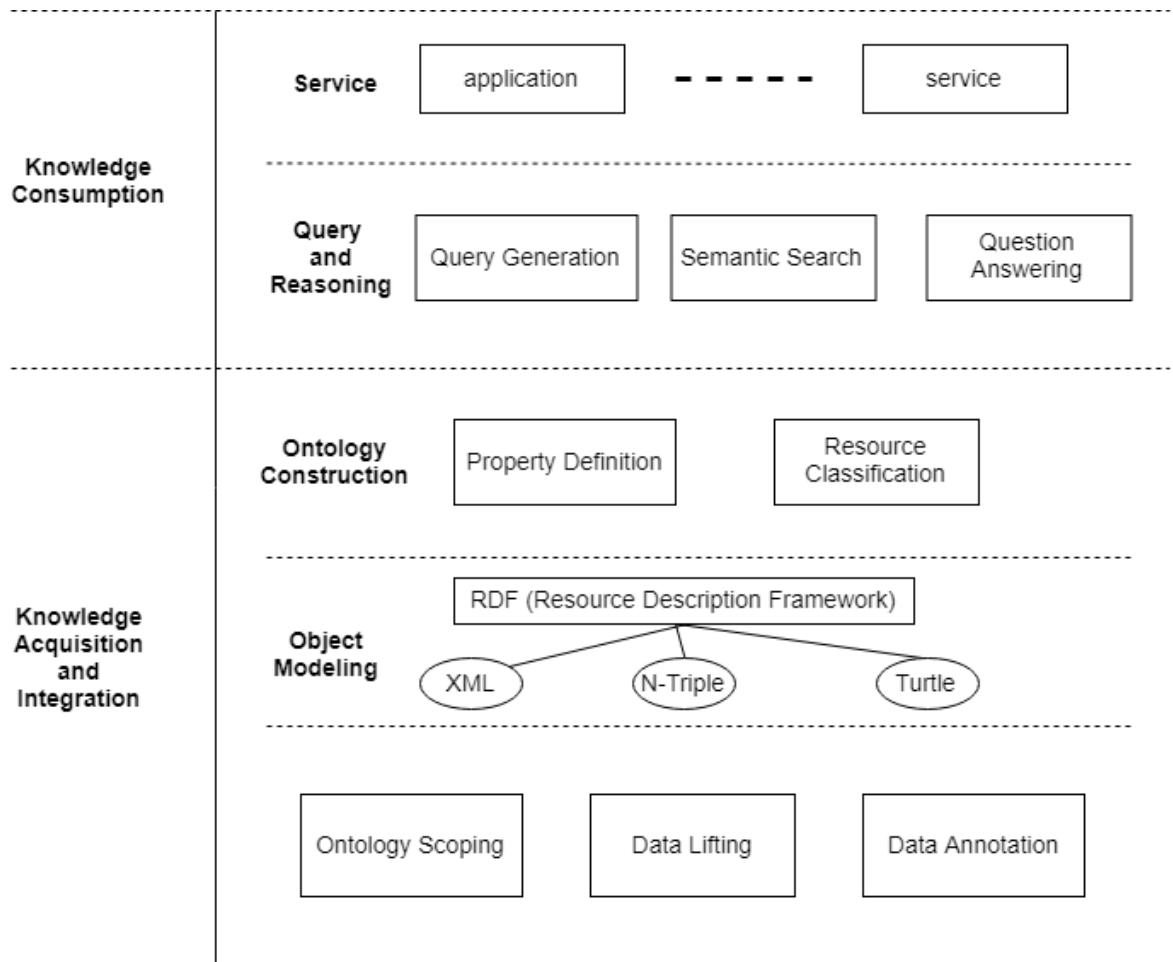
**Phase 2:** Ontology construction (reusing existing ontology) and object modelling

**Phase 3:** Knowledge graph construction and querying

**Phase 4:** Evaluation of the system and Use case analysis

### 3.2 System Architecture

The system should be able to represent knowledge in form of knowledge graph. The general architecture of implemented solution that satisfies the requirements is represented in Fig.7. The architecture is divided in two main layers: Knowledge Acquisition and Integration layer, Knowledge Consumption layer.



**Fig. 7.** System Architecture.

The *knowledge acquisition and integration layer* constitute scoping, object modelling and ontology construction. Scoping refers to ontology scoping to serve the purpose, data lifting from data sources and data annotation. Ontology is constructed reusing the existing vocabularies to meet the requirements. Ontology construction phase includes appropriate property definition and resource classification. Annotate data is linked in RDF (XML/ N-triple/ Turtle) format following the created vocabulary and stored in RDF store.

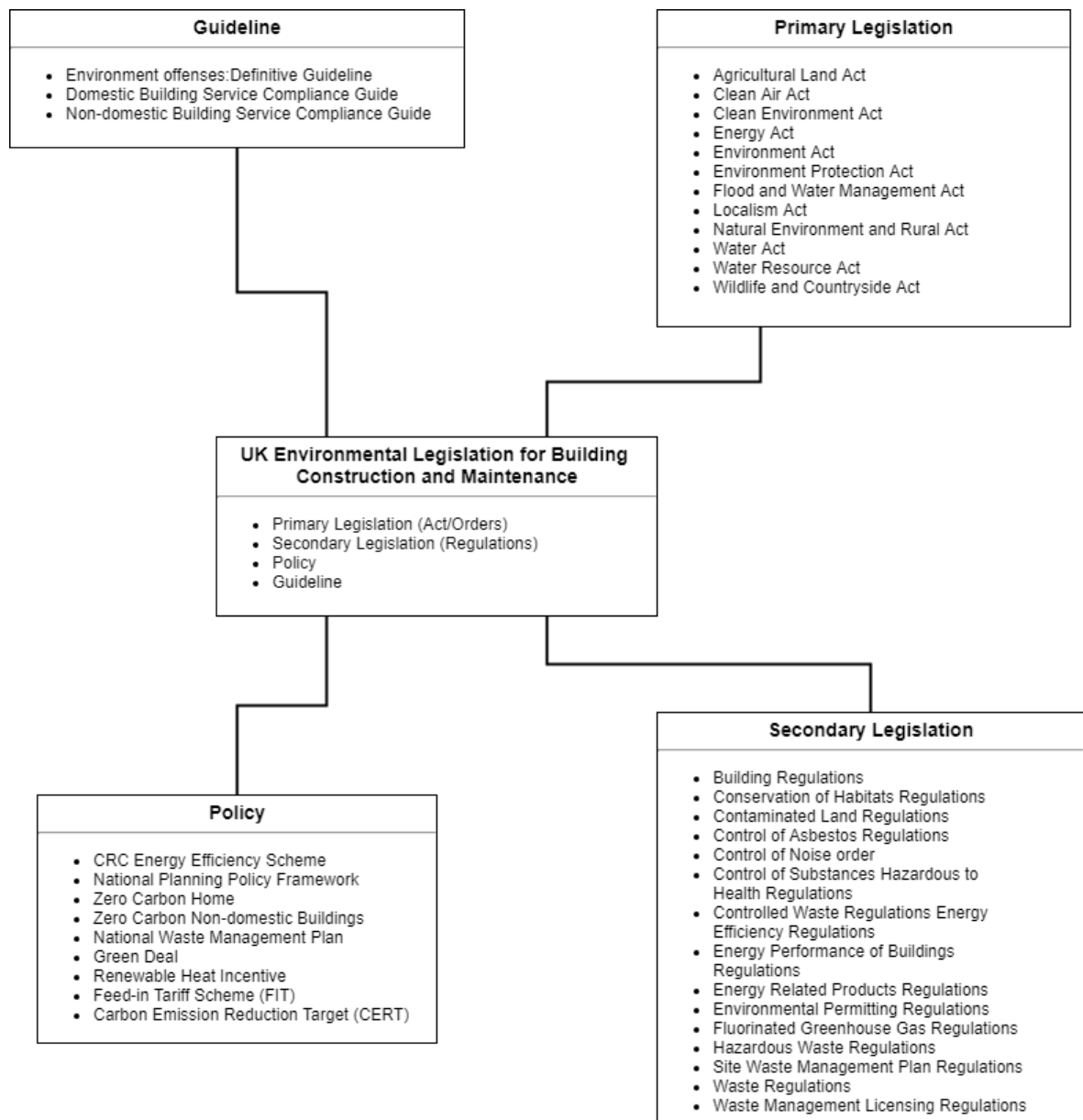
The *knowledge consumption* layer works as interface to user to consume the knowledge representation. This layer is divided in two sub layers: query and reasoning, service. The service layer provides APIs that facilitates query and reasoning.

## 4 REQUIREMENTS AND DESIGN

This chapter discusses the requirements and design of the proposed system.

### 4.1 Requirements

Environmental Legislations for building construction and maintenance can be break down in four sections: Primary Legislations (Acts/Orders), Secondary Legislations (Regulations),



**Fig. 8.** Environmental guidelines, legislations and policies related to building construction and maintenance.

Policy, Guideline. These sections have long list of underneath legislations and policies which are created and maintained by various regulatory authorities from UK Government. The format of data representation is different for acts, orders, regulations, policies and guidelines. Also, data representation for acts, orders, regulations, policies and guidelines varies from one authority to another. Fig.8 gives an overview of the broader picture. As discussed in delimitation of the thesis, proposed solution is to represent knowledge for Guideline e.g. Domestic building service compliance guideline in UK [18].

The domestic building service compliance guideline contains functional requirements, such as that domestic buildings must be structurally stable, constructed and fitted to ensure fire protection and energy efficiency.

	Seasonal efficiency		
	SEDBUK 2009 <sup>b</sup>	SEDBUK 2005/ SEDBUK	ErP <sup>®</sup>
<b>Gas-fired wet central heating</b>			
Condensing boilers in new dwellings	88%	90%	-
Condensing boilers in existing dwellings	-	-	92%
Non-condensing boilers (where permitted): natural gas/LPG	78%/80%	78%/80%	-
Range cooker boilers	75%	75%	-
<b>Gas-fired warm air heating</b>	<b>Efficiency</b>		
	See Table 6		
<b>Gas-fired fixed independent space heaters</b>	<b>Efficiency (gross)<sup>a</sup></b>		
Gas and LPG primary heating	63%		
Gas and LPG secondary heating	63% (new build) 45% (existing build)		
Decorative fuel-effect	Not specified (set to 20% in SAP 2012)		
<b>Gas fires in combined fire/backboilers (replacements)</b>	<b>Efficiency (gross)</b>		
	<b>Natural gas</b>	<b>LPG</b>	
Inset live fuel-effect	45%	46%	
All types except inset live fuel-effect	63%	64%	

**Fig. 9.** A part of domestic building service compliance guideline.

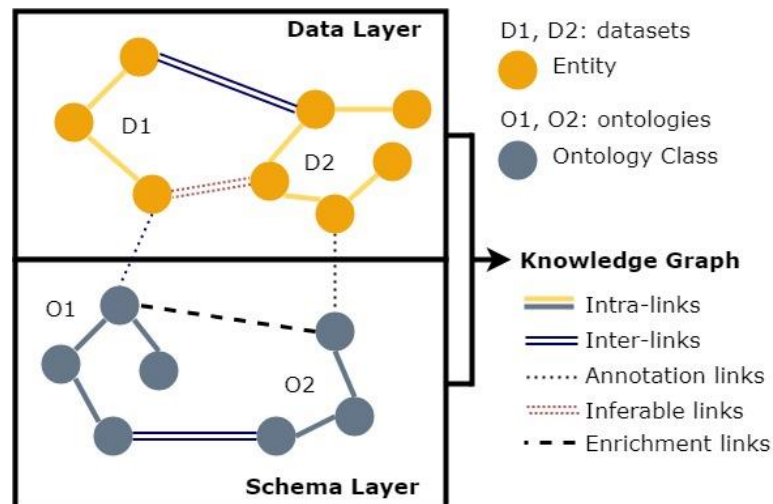
The guide comprises two sections, fuel-based and technology-specific guideline. In both sections, there are information to help with interpreting minimum energy specific provisions needed to comply with the building regulations. The guide contains requirements for both existing and new system and installation of equipment. Also, there are clear indication of recommended efficiency standards.

For building knowledge representation in form of knowledge graph, an appropriate vocabulary is required. Scoping of vocabulary can be done in two ways: creating a new vocabulary, reusing existing vocabulary. After vocabulary is prepared, data needs to be lifted from source. Data lifting process can be manual or automated. In this implementation, data lifting process is manual and taken from service compliance guide book. Lifted data is required to be annotate and stored as RDF triples.

To query and reasoning over linked data store, a query engine and reasoner is required. The consumption of knowledge is required to be served by APIs.

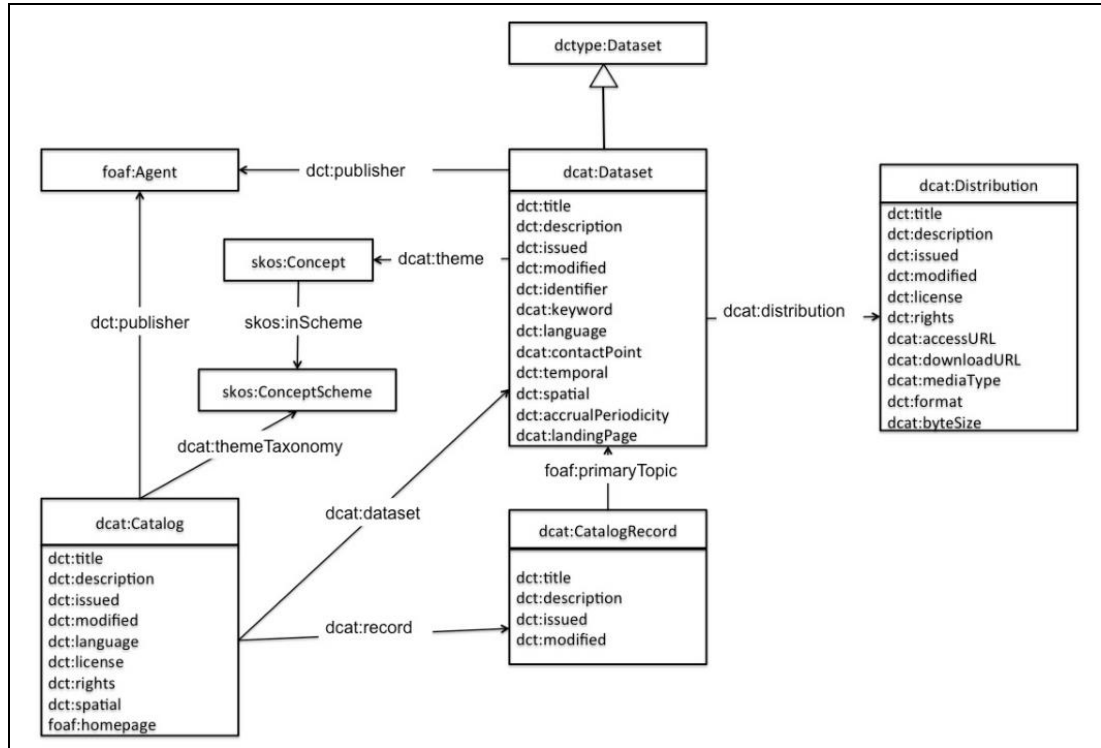
## 4.2 Design

The implemented Knowledge Graph is of two layers: schema or ontology layer, data layer.



**Fig. 10.** Two layer of Knowledge Graph.

Current implementation is done on top of single dataset and single ontology that is to be extended with multiple datasets and ontologies to leverage the full benefit of knowledge graph.

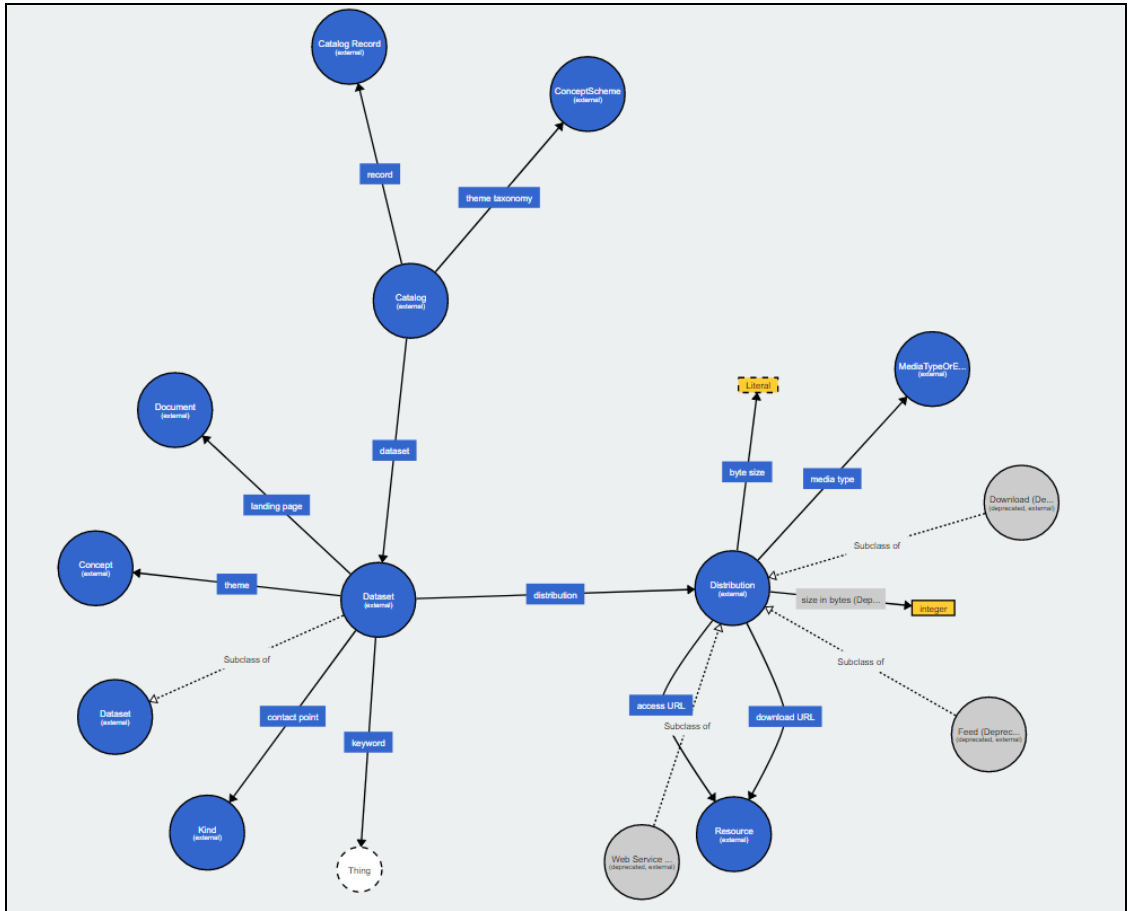


**Fig. 11.** Overview of DCAT (Data Catalogue Vocabulary)

Scoping for appropriate ontology is done in existing ontology repositories including Swoogle [48] and Linked Open Vocabularies [49]. Upon study of relevant vocabularies, DCAT (Data Catalogue Vocabulary) is chosen for reuse with further extension of it.

Overview of DCAT vocabulary is shown in Fig.11. DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web [61]. DCAT is a well-suited RDF vocabulary to represent government data catalogue. It has three main classes:

1. Catalog
2. Dataset
3. Distribution



**Fig. 12.** Onto Visualization of DCAT vocabulary.

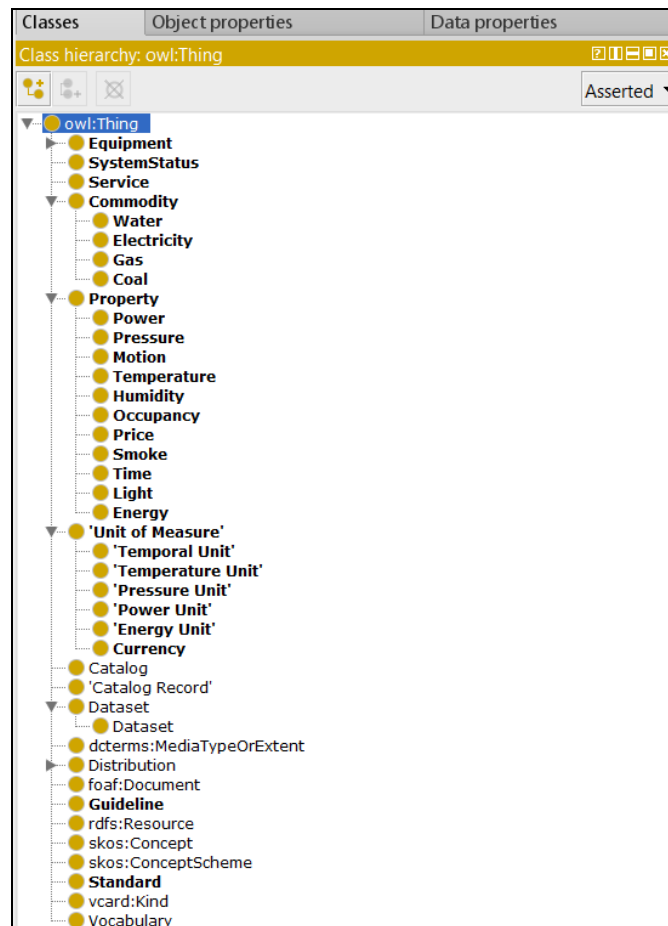
Onto visualization of DCAT vocabulary is showed in Fig.12 using WebVOWL [62] visualization tool. It shows all the classes, object properties and data properties in the vocabulary.

## 5 UKENV-KG SYSTEM IMPLEMENTATION AND EVALUATION

Implementation of UKENV-KG system is divided in following steps: extension of DCAT vocabulary, data annotation and storing, query and reasoning. Evaluation of the implemented system is done analyzing the use cases which will be further extended by analysis of query execution time for multiple data storage and ontology scenarios.

### 5.1 Extending DCAT Vocabulary

As DCAT vocabulary doesn't fully suffice the requirements, the vocabulary needs to be reused and extended.

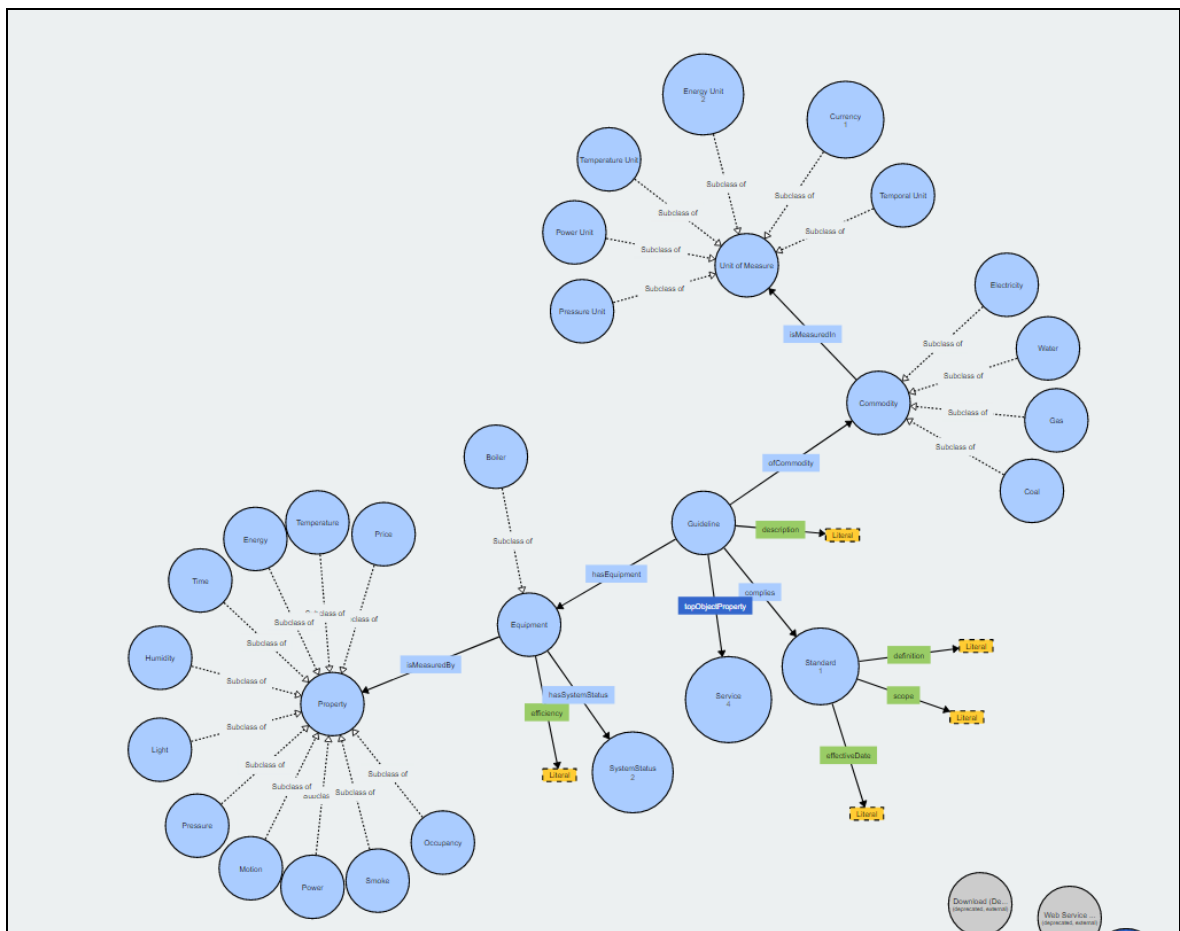


**Fig. 13.** Ontology classes in Protégé.

To represent data lifted from Domestic Building Service Compliance guide DCAT vocabulary is added with following ontology classes:

- Equipment
- SystemStatus
- Service
- Commodity
- Property
- UnitOfMeasure
- Standard
- Guideline

After adding required additional classes, class hierarchy is shown in Fig.13. For extending DCAT vocabulary, the ontology was loaded in Protégé tool. Besides additional ontology classes, data property and object properties are added upon requirement analysis.



**Fig. 14.** Enhancement of DCAT vocabulary to meet requirements.

After adding ontology classes, data property and object property, the visualization of the vocabulary is shown in Fig.14.

## 5.2 Data Annotation, Storage and Querying

After ontology is constructed, lifted data is annotate with new vocabulary. For experimental purpose, annotate data was imported from Protégé tool in different syntax such as RDF/XML, OWL/XML, RDF Turtle. Fig.15 shows a segment from annotate data dumped in RDF Turtle syntax.

```
@prefix : <http://localhost:8800/ukenv-dcat/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@base <http://localhost:8800/ukenv-dcat/> .

□ <http://localhost:8800/ukenv-dcat/> rdf:type owl:Ontology ;
    owl:imports <http://www.w3.org/ns/dcat> .

#####
#   Object Properties
#####

### http://localhost:8800/ukenv-dcat/complies
□ :complies rdf:type owl:ObjectProperty ;
    rdfs:domain :Guideline ;
    rdfs:range :Standard .

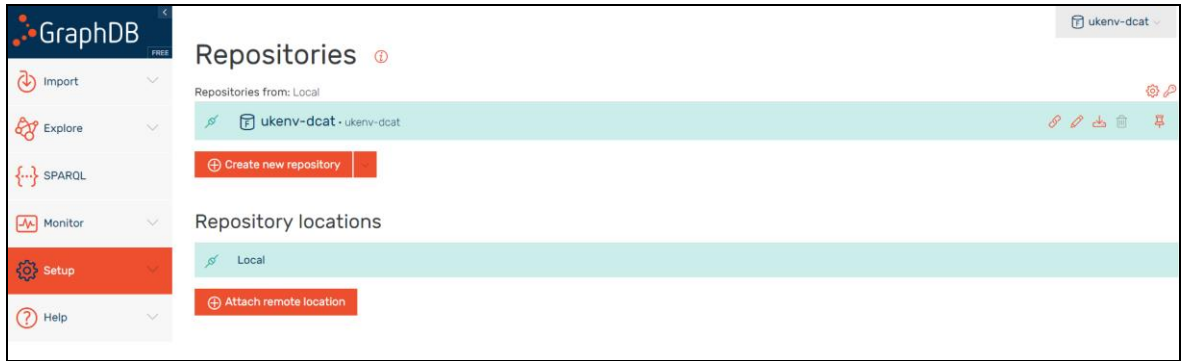
### http://localhost:8800/ukenv-dcat/hasEquipment
□ :hasEquipment rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf owl:topObjectProperty ;
    rdfs:domain :Guideline ;
    rdfs:range :Equipment .

### http://localhost:8800/ukenv-dcat/hasService
□ :hasService rdf:type owl:ObjectProperty ;
    rdfs:subPropertyOf owl:topObjectProperty .
```

**Fig. 15.** Segment of RDF Turtle based on modified vocabulary.

For storing RDF data, GraphDB is used [20]. GraphDB's APIs are built on top of RDF4J Java Framework that is widely used for processing and handling RDF data [21]. GraphDB is run as stand-alone server which comes with a preconfigured web server.

The RDF turtle file is loaded in GraphDB server creating repositories for this specific purpose. Fig.16 shows the interface of GraphDB repository interface where RDF file is uploaded.



**Fig. 16.** GraphDB as RDF storage and SPARQL query engine.

Once data is stored in GraphDB repositories, it can be queried using integrated SPARQL query engine. Feature of querying is extended with integration of RDF4J connector with RESTful Web service.

## 5.3 Evaluation

This section discusses the use cases and performance of the implemented system.

### 5.3.1 Use Cases

Knowledge graph is not the only option while dealing with heterogeneous data. Two popular alternatives to knowledge graph as data model are XML and relational model. XML comes with the tree structure which is a big limiting factor. Storing any graph structure gets heavily limited with the XML's hierarchical expression. In case of relational data model, limitation is more tenuous comparing to knowledge graph.

While integrating two different knowledge graph, we only need an IRI (Internationalized Resource Identifier) whereas it is a considerable problem in case of relational data model because of various complex table operations [63][64]. The benefit of integration of multiple knowledge sources are useful in our context as environmental legislations and guidelines are published and maintained by different authorities.

Another advantage of publishing legislation and guidelines in form of knowledge graph is simply the availability of data. Data hosted in proprietary format under different jurisdiction is harder to put in a common platform using relational data model. Knowledge graphs, however, are easy to publish and platform independent.

Despite the implementation in this research work is only focused to service compliance guideline or legislation domain, similar approach can be adopted in other domain such as enterprise guideline, enterprise information management. Knowledge graph has been widely adopted in many enterprises to store and consume enterprise knowledge. One big advantage of such representation is, a standalone graph can be integrated with other external representations.

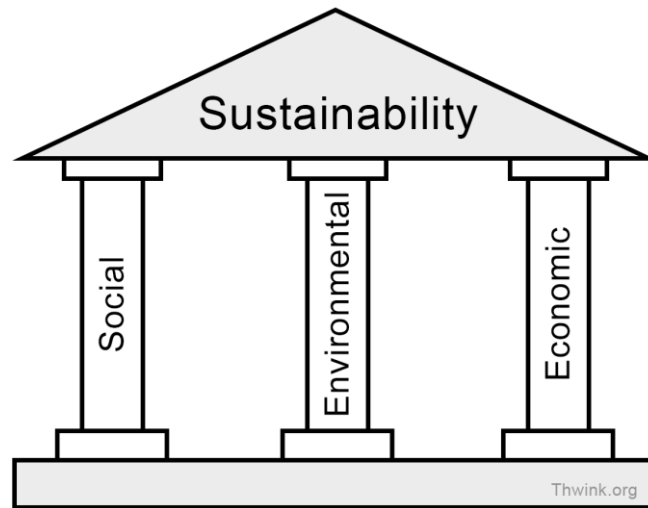
### **5.3.2 Performance Analysis**

Existing implementation is done on basis of single dataset and single ontology. It will be extended to multiple dataset with use of multiple ontology.

## **5.4 Sustainability Analysis**

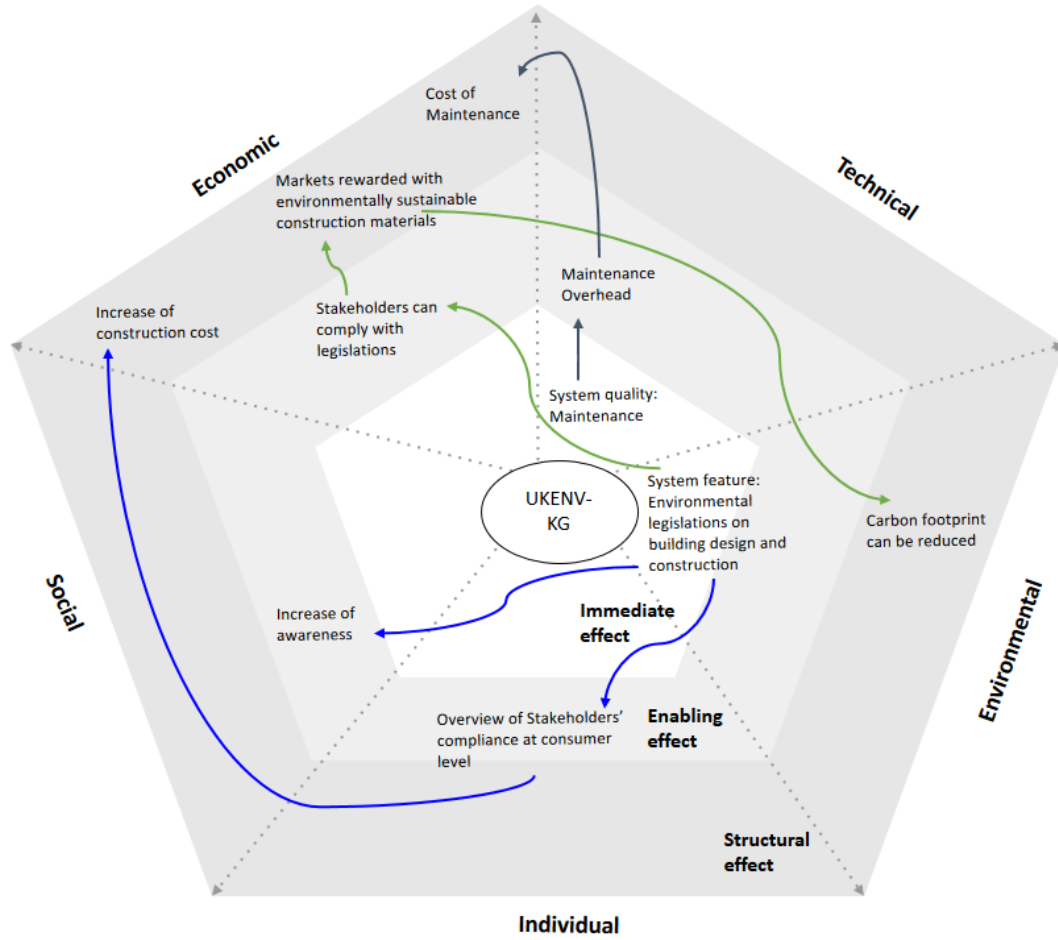
The PERCCOM program aims at combining ICT with environmental awareness to build cleaner, greener and energy efficient cyber-physical systems [65]. As a part of PERCCOM program, this research work adopts the concept of sustainability. A sustainable development process meets the requirements of present without compromising the ability of future generations to meet their own requirements [66].

Sustainable development is built on three main pillars: Economic, Social and Environmental (Fig. 18). Each of the pillars is highly interdependent. The environmental aspect focuses on meeting the present need without disrupting global environmental ecosystem.



**Fig. 18.** Three pillars of sustainability.

The economic pillar emphasizes of the current economic activity not to be disproportionately burden to future generations. Definition of social pillar of a sustainable development highlights on growing a sense of community ownership by participation of citizens to transmit awareness of social sustainability [67].



**Fig. 19.** Sustainability analysis pentagon.

Proposed solution (UKENV-KG) is applied for building a knowledge graph that encompasses UK environmental legislations related to building construction and maintenance. The solution provides stakeholders (constructors, engineers, architects, planner, and dwellers) an ease-of-access platform to go through legislations and compliance guidelines and influences the social pillar of sustainability. Also, complying with environmental guidelines and legislations affects environmental aspects of sustainability. From community perspective, it offers equity of access to services.

Christoph et al. [68] highlights on sustainability analysis of software system in five dimensions (Economic, Social, Technical, Environment, Individual) with immediate, enabling and structural effect. Fig.19 illustrates the pentagon of sustainability analysis for this project. The diagram identifies the immediate, enabling and structural effect of the

proposed solution in Economic, Social, Technical, Environmental and Individual paradigm.

Discussion can be made on direct and indirect effect of proposed software system. To talk about direct effects, as the system would provide an ease of use platform for the stakeholder, the compliance rate would definitely rise which will definitely have a positive impact on environment. Software systems used today cannot be considered as isolated system, rather part of the socio-technical system where the software is deployed [69]. Software solution made based on proposed framework will definitely become a part of socio-technical system with definite direct and indirect impact.

## **6 CONCLUSION AND FUTURE WORK**

This final chapter contains overall conclusions of the thesis work and possible future work.

### **6.1 Conclusions**

The objective of this research work was to leverage idea of knowledge graph in domain of UK environmental legislation related to building construction and maintenance. Due to vastness of the domain of interest, implementation was delimited to service compliance sub-domain. Nevertheless, the proposed framework can be used for extension of graph coverage in legislation domain.

In alignment with predefined requirements, UKENV-KG was developed with reuse and extension of existing ontology which is responsible for conceptually describing the domain. The evaluation of implemented solution gives scope to draw several conclusions.

Knowledge graph is a popular concept in web domain whereas it can also be used in other areas. Usage of semantic technologies in legislation domain is yet not popular due to certain reasons. Ontology construction requires lots of effort and expertise. Also, data lifting in manual fashion is slower and resource consuming.

### **6.2 Future Work**

There are several definite improvements that the system needs. First, the knowledge creation is done in curated approach where data lifting is done manually. Automated data lifting is necessary in large scale knowledge graph development.

One of the big challenges in legislation information management is dealing with temporal information. In this research temporal issue is not handled whereas it can be achieved by versioning or extending ontology with support to temporal information.

The implementation only covers the Domestic Building Service Compliance Guideline of UK. To achieve maximum from knowledge graph, it needs to be extended and cover other related domains.

## 7 REFERENCES

- [1] A.-L. Kor *et al.*, “Education in green ICT and control of smart systems : A first hand experience from the International PERCCOM masters programme,” 2019.
- [2] “Climate change - UKGBC - UK Green Building Council.” [Online]. Available: <https://www.ukgbc.org/climate-change/>. [Accessed: 07-Jun-2019].
- [3] “2010 to 2015 government policy: greenhouse gas emissions - GOV.UK,” 2015. [Online]. Available: <https://www.gov.uk/government/publications/2010-to-2015-government-policy-greenhouse-gas-emissions/2010-to-2015-government-policy-greenhouse-gas-emissions>. [Accessed: 28-May-2019].
- [4] HM Government, “The UK Low Carbon Transition Plan,” 2009.
- [5] “Climate Change Act 2008,” 2008.
- [6] M. Bell *et al.*, “Next steps for UK heat policy,” 2016.
- [7] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, “A review of relational machine learning for knowledge graphs,” *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, 2016.
- [8] F. Van Harmelen, V. Lifschitz, and B. Porter, *Handbook of knowledge representation*. Elsevier, 2008.
- [9] T. Steiner, R. Verborgh, R. Troncy, J. Gabarro, and R. Van De Walle, “Adding Realtime Coverage to the Google Knowledge Graph.”
- [10] J. M. Gomez-Perez, J. Z. Pan, G. Vetere, and H. Wu, “Enterprise Knowledge Graph: An Introduction,” in *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, Cham: Springer International Publishing, 2017, pp. 1–14.
- [11] J. Yan, C. Wang, W. Cheng, M. Gao, and A. Zhou, “A retrospective of knowledge graphs,” *Front. Comput. Sci.*, vol. 12, no. 1, pp. 55–74, 2018.
- [12] R. Angles and C. Gutierrez, “Survey of graph database models,” *ACM Comput. Surv.*, vol. 40, no. 1, pp. 1–39, Feb. 2008.
- [13] X. Wilcke, P. Bloem, and V. de Boer, “The knowledge graph as the default data model for learning on heterogeneous knowledge,” *Data Sci.*, pp. 1–19, Oct. 2017.
- [14] E. Montiel-Ponsoda, V. Rodríguez-Doncel, and J. Gracia, “Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe.”
- [15] T. Heath and C. Bizer, “Linked Data: Evolving the Web into a Global Data Space,”

- Synth. Lect. Semant. Web Theory Technol.*, vol. 1, no. 1, pp. 1–136, Feb. 2011.
- [16] Z. O. Repository, “University of Zurich Applied temporal RDF : efficient temporal querying of RDF data with SPARQL Applied temporal RDF : efficient temporal querying of RDF data with SPARQL,” no. June, pp. 308–322, 2009.
  - [17] Claudio Gutierrez, Carlos A. Hurtado, and Alejandro Vaisman, “Introducing Time into RDF,” *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 2, pp. 207–218, 2007.
  - [18] HM Government, “Domestic Building Services,” 2018.
  - [19] M. Schmidt, T. Hornung, G. Lausen, and C. Pinkel, “SP<sup>2</sup>Bench: A SPARQL Performance Benchmark,” in *2009 IEEE 25th International Conference on Data Engineering*, 2009, pp. 222–233.
  - [20] “Programming with GraphDB — GraphDB Free 8.9 documentation.” [Online]. Available: <http://graphdb.ontotext.com/free/devhub/programming.html>. [Accessed: 07-Jun-2019].
  - [21] “Eclipse rdf4j | The Eclipse Foundation.” [Online]. Available: <https://rdf4j.eclipse.org/>. [Accessed: 29-May-2019].
  - [22] *Expert Spring MVC and Web Flow*. Apress, 2006.
  - [23] G. KONG, D.-L. XU, and J.-B. YANG, “Clinical Decision Support Systems: a Review of Knowledge Representation and Inference under Uncertainties,” *Int. J. Comput. Intell. Syst.*, vol. 1, no. 2, p. 159, 2008.
  - [24] I. B. Arpinar, R. Zhang, B. Aleman-Meza, and A. Maduko, “Ontology-driven Web services composition platform,” *Inf. Syst. E-bus. Manag.*, vol. 3, no. 2, pp. 175–199, Jul. 2005.
  - [25] I. Jurisica, J. Mylopoulos, and E. Yu, “Ontologies for Knowledge Management: An Information Systems Perspective,” *Knowl. Inf. Syst.*, vol. 6, no. 4, pp. 380–401, Jul. 2004.
  - [26] C. Asamoah, L. Tao, K. Gai, and N. Jiang, “Powering Filtration Process of Cyber Security Ecosystem Using Knowledge Graph,” in *2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)*, 2016, pp. 240–246.
  - [27] M. Iannacone *et al.*, “Developing an Ontology for Cyber Security Knowledge Graphs,” in *Proceedings of the 10th Annual Cyber and Information Security Research Conference on - CISR '15*, 2015, pp. 1–4.

- [28] J. Z. Pan *et al.*, *Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering : 12th International Summer School 2016, Aberdeen, UK, September 5-9, 2016, Tutorial Lectures.* .
- [29] M. R. Quillian, “Word concepts: A theory and simulation of some basic semantic capabilities,” *Behav. Sci.*, vol. 12, no. 5, pp. 410–430, Sep. 1967.
- [30] S. Decker *et al.*, “The Semantic Web: the roles of XML and RDF,” *IEEE Internet Comput.*, vol. 4, no. 5, pp. 63–73, 2000.
- [31] J. Z. Pan, “DESCRIPTION LOGICS: REASONING SUPPORT FOR THE SEMANTIC WEB,” 2004.
- [32] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds., *The Description Logic Handbook*. Cambridge: Cambridge University Press, 2007.
- [33] B. Villazon-Terrazas *et al.*, “Knowledge Graph Foundations,” in *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, Cham: Springer International Publishing, 2017, pp. 17–55.
- [34] D. Reynolds, “The Organization Ontology.” [Online]. Available: <https://www.w3.org/TR/vocab-org/>. [Accessed: 27-May-2019].
- [35] M. Hepp, “GoodRelations: The Professional Web Vocabulary for E-Commerce.” [Online]. Available: <http://www.heppnetz.de/projects/goodrelations/>. [Accessed: 28-May-2019].
- [36] D. Brickley and L. Miller, “FOAF Vocabulary Specification.” [Online]. Available: <http://xmlns.com/foaf/spec/>. [Accessed: 28-May-2019].
- [37] *The code4lib journal*. Code4Lib Journal, 2007.
- [38] J. Huang *et al.*, “OmniSearch: a semantic search system based on the Ontology for MicroRNA Target (OMIT) for microRNA-target gene interaction data.,” *J. Biomed. Semantics*, vol. 7, p. 25, 2016.
- [39] Ozonat Mehmet and Bartolini Claudio, “Building a semantics graph for an enterprise communication network,” Jan. 2013.
- [40] A. J. G. Gray, N. Gray, and I. Ounis, “Can RDB2RDF Tools Feasibly Expose Large Science Archives for Data Integration?,” Springer, Berlin, Heidelberg, 2009, pp. 491–505.
- [41] E. Montiel-Ponsoda, V. Rodríguez-Doncel, and J. Gracia, “Building the legal

- knowledge graph for smart compliance services in multilingual Europe,” *CEUR Workshop Proc.*, vol. 2049, no. December 2017, pp. 15–17, 2018.
- [42] “Representing enterprise data in a knowledge graph,” May 2013.
  - [43] G. Weikum and M. Theobald, “From information to knowledge,” in *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems of data - PODS '10*, 2010, p. 65.
  - [44] R. Denaux, Y. Ren, B. Villazon-Terrazas, P. Alexopoulos, A. Faraotti, and H. Wu, “Knowledge Architecture for Organisations,” in *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, Cham: Springer International Publishing, 2017, pp. 57–84.
  - [45] B. Villazon-Terrazas *et al.*, “Construction of Enterprise Knowledge Graphs (I),” in *Exploiting Linked Data and Knowledge Graphs in Large Organisations*, Cham: Springer International Publishing, 2017, pp. 87–116.
  - [46] T. Heath and C. Bizer, “Linked Data: Evolving the Web into a Global Data Space,” *Synth. Lect. Semant. Web Theory Technol.*, vol. 1, no. 1, pp. 1–136, Feb. 2011.
  - [47] M. C. Suárez-Figueroa, “NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse,” Jun. 2010.
  - [48] L. Ding *et al.*, “Swoogle,” in *Proceedings of the Thirteenth ACM conference on Information and knowledge management - CIKM '04*, 2004, p. 652.
  - [49] P.-Y. Vandenbussche, G. A. Atemezinge, M. Poveda-Villalón, and B. Vattant, “Linked Open Vocabularies (LOV): A gateway to reusable semantic vocabularies on the Web,” *Semant. Web*, vol. 8, no. 3, pp. 437–452, Dec. 2016.
  - [50] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens, and R. Van De Walle, “RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data.”
  - [51] A. Langeegger and W. Wöß, “XLWrap – Querying and Integrating Arbitrary Spreadsheets with SPARQL,” Springer, Berlin, Heidelberg, 2009, pp. 359–374.
  - [52] “Apache Jena.” [Online]. Available: <http://jena.apache.org/>. [Accessed: 29-May-2019].
  - [53] “OpenLink Virtuoso.” [Online]. Available: <https://virtuoso.openlinksw.com/>. [Accessed: 29-May-2019].
  - [54] A. Hogan, A. Harth, A. Passant, S. Decker, and A. Polleres, “Weaving the Pedantic

Web LDOW 2010,” 2009.

- [55] Z. Zhao, S.-K. Han, and I.-M. So, “Architecture of Knowledge Graph Construction Techniques,” *Int. J. Pure Appl. Math.*, vol. 118, no. 19, pp. 1869–1883, 2018.
- [56] M. Uschold and M. Gruninger, “Ontologies: principles, methods and applications,” *Knowl. Eng. Rev.*, vol. 11, no. 2, pp. 93–136, Jun. 1996.
- [57] Y. Ren, A. Parvizi, C. Mellish, J. Z. Pan, K. van Deemter, and R. Stevens, “Towards Competency Question-Driven Ontology Authoring,” Springer, Cham, 2014, pp. 752–767.
- [58] S. R. (Stephen R. Palmer and J. M. Felsing, *A practical guide to feature-driven development*. Prentice Hall PTR, 2002.
- [59] C. R. Kothari, *Research methodology : methods & techniques*. New Age International (P) Ltd, 2004.
- [60] N. B. Ruparelia and N. B., “Software development lifecycle models,” *ACM SIGSOFT Softw. Eng. Notes*, vol. 35, no. 3, p. 8, May 2010.
- [61] “Data Catalog Vocabulary (DCAT).” [Online]. Available: <https://www.w3.org/TR/vocab-dcat/>. [Accessed: 07-Jun-2019].
- [62] “WebVOWL - Web-based Visualization of Ontologies.” [Online]. Available: <http://vowl.visualdataweb.org/webvowl.html>. [Accessed: 07-Jun-2019].
- [63] R. Hecht and S. Jablonski, “NoSQL evaluation: A use case oriented survey,” in *2011 International Conference on Cloud and Service Computing*, 2011, pp. 336–341.
- [64] M. A. Hernández, S. J. Stolfo, M. A. Hernández, and S. J. Stolfo, “The merge/purge problem for large databases,” in *Proceedings of the 1995 ACM SIGMOD international conference on Management of data - SIGMOD '95*, 1995, vol. 24, no. 2, pp. 127–138.
- [65] A. Klimova, E. Rondeau, K. Andersson, J. Porras, A. Rybin, and A. Zaslavsky, “An international Master’s program in green ICT as a contribution to sustainable development,” *J. Clean. Prod.*, vol. 135, pp. 223–239, Nov. 2016.
- [66] L. M. Hilty and M. D. Hercheui, “ICT and Sustainable Development,” Springer, Berlin, Heidelberg, 2010, pp. 227–235.
- [67] J. Morelli, “Environmental Sustainability: A Definition for Environmental Professionals,” *J. Environ. Sustain.*, vol. 1.

- [68] C. Becker *et al.*, “Requirements: The Key to Sustainability,” *IEEE Softw.*, vol. 33, no. 1, pp. 56–65, Jan. 2016.
- [69] L. Duboc *et al.*, “Do we really know what we are building? Raising awareness of potential Sustainability Effects of Software Systems in Requirements Engineering.” IEEE Computer Society, 22-Jun-2019.

