Ontology-based Framework for Carbon Reduction Assessment

S. Pandey, G. Beydoun, M. Bandara, B. McCusker, C. Devalence

Abstract

Carbon certificates provide a mechanism to recognise actions conducted by organisations to reduce their carbon footprint. While assessment certificates offer a huge promise in responding to climate change, a reliable approach to the issuance of certificates with an accurate carbon reduction estimate remains elusive. We believe that by collecting and maintaining high-quality and comprehensive datasets on historic carbon reduction projects, and utilising them effectively in planning future projects, we can improve the accuracy of carbon reduction estimations. This paper proposes a reuse framework to achieve this goal. The framework combines the use of ontologies and knowledge graph to consolidate access diverse data sources into a singular platform. The platform will also provide an interactive interface for organisations to design and develop their projects using historical data and insights. Knowledge graph-based platforms are proven to foster knowledge-sharing, encourage best practices, and enable transparent decision making. This paper aims to bring those benefits into modern organizations, supporting their goal for sustainability and the collective goal of a low-carbon economy.

Keywords: Carbon Reduction, Carbon Reduction Assessment, Data-Driven Decision Making, Ontology Modelling, Knowledge Graph.

1 Introduction

To attain the goals of Paris Agreement and curtail of greenhouse gas (GHG or CO2) emissions, the signatories have primarily cantered their strategies on implementing a tax on GHG emissions (Zhou & Li, 2019). This requires attributing a singular price to CO2 emissions, irrespective of their origin (United Nations Climate Change, 2022) and is not governed by any one single party rather a global market mechanism. It has a multitude of different stakeholders players with overlapping roles. These include: Carbon reduction project developers, Carbon credit registries and standard bodies, Carbon credit verifiers, Carbon credit brokers and marketplace, Carbon credit ratings agencies and possibly others (Walsh & Toffel, 2023).

The efficient administration of a carbon reduction project poses significant challenges, extending from the consolidation of the market's objectives, and the formulation of relevant policies, rules, and regulations, to the management of voluminous and various datasets of interest to different players involved in this marketplace. Key issues that obstruct smooth operation for any carbon reduction project include but are not limited to data integration from different sources, environmental integrity, and functional efficacy (Mehling, 2009). Carbon credits still do not have a centralized or sizable market with universally accepted standards. This poses challenges for carbon reduction project developers tasked with producing those credits. For instance, in the current operating environment, selecting specific methodologies to be followed by the project is quite a serious challenge due to the lack of standard platforms or tools that would allow them to evaluate the most suitable method. This makes it quite difficult to use past reduction projects to inform their decisions about their own. Comparing features associated with the locations of a current project with corresponding features associated with past projects is cumbersome if not impossible in many cases. This makes the process of identifying and creating carbon credits difficult to understand and reuse. This is further complicated by different rules of various carbon credit verifying bodies that must be satisfied.

Our primary interest and research focus on providing a data management platform to mitigate challenges against lack of standardizations, the wide array of datasets and methods in seeking carbon accreditation processes. We aim to provide a unified access point to all carbon reduction projects, including both past completed projects and those currently under verifications by different registries. Our unifying access point will include details of individual projects as well as meaningful links between them. Our approach will rely on a unifying ontology operationalized with a knowledge graph. The main contribution of this paper is the Knowledge Base Carbon Reduction Assessment (CRA) framework. CRA framework will link details associated with the carbon reduction projects to specifically assist project developers in selecting appropriate methodologies for their carbon reduction project development process. They can compare relevant rules and regulations using knowledge from past projects. While our illustrative exemplars focus on the carbon project developers, our approach will also enable other players to evaluate various trade-offs when seeking carbon accreditations, thus ensuring technical barriers for carbon accreditation processes are removed.

The rest of the paper is organized as follows: Section 2 is about the background and related work. Section 3 talks about methodology, where we talked about the design science research method, ontology modelling and knowledge graph construction process. Similarly, section 4 is related to the demonstration and evaluation of the knowledge graph, and section 5 is about the contributions and limitations. Finally, section 6 is about the conclusion and future enhancements.

2 Background

Edgar W. Schneider, an Austrian linguist, initially coined the term Knowledge Graph in 1972 while he was discussing how to create modular instructional methods for courses (Schneider, 1973). Later in 2012, Google unveiled their knowledge graph to improve the search engine's capacity to produce results that are more precise and pertinent to user needs (Singhal, 2012). This was a useful addition to Google's string-based search that enjoyed substantial success, and the phrase was more widely used, and since then, knowledge graphs have been integrated across various sectors such as finance, medicine, energy and power (Singhal, 2012).

Knowledge graphs can be general or domain specific. In our research, we are interested in creating a domain-specific knowledge graph for our domain (Carbon Reduction Assessment). Prior work on KG has been evidence that KG has been successfully implemented in various domains such as healthcare, education, environment, ICT, sciences, engineering, etc., (Abu-Salih, 2021; Zou, 2020). A knowledge graph can be constructed using ontologies. An ontology can be used to structure the raw data and create consistent semantics and formalisation (Wu et al., 2022). In the sustainability assessment domain, Tang et al. (Tang et al., 2023) proposes Petro-KG, a knowledge graph developed starting with ontology and entities. Yu et al. (Yu et al., 2024) also propose ESGMKG, a knowledge graph for managing environmental, social, and government metrics, that showcase the value of knowledge graphs and ontologies in managing complex information needs in sustainability reporting.

The carbon marketplace faces many challenges due to the lack of standards and adequate access to relevant and scattered datasets. It is inherently complex. A comprehensive ontology libraries can assist in guiding reusable and concerted approaches that cater for the carbon reduction assessment. An important contribution to this effort is by Peng, Y. (Peng et al., 2023), who proposed a comparative analysis of China's pilot carbon market based on a knowledge graph. But various stakeholders e.g. verifying and validating agencies, rating agencies, brokers and even carbon credit buyers all face daunting tasks when pursuing suitable carbon accreditations that meet all the criteria associated with the Sustainable Development Goals (SDG), carbon credit quality, carbon credit being issued and retired, etc. This paper explores the utilization of a Knowledge Graph to address the issue of integrating and managing carbon market data from various stakeholders, aiming to enhance the national carbon market mechanism. The utilisation of a knowledge graph is underpinned by the identification of a suitable seeding ontology that provides an initial set of concepts and relationships. This is a critical part of our CRA framework. This part of our work uses ground work from Su et al. which analyses emerging keywords associated with carbon mitigation, such as carbon capture and storage, biomass energy, clean energy, etc. (Su & Wang, 2022). They used the power of knowledge graph to find the top keywords with the strongest citations from 1991 to 2021.

3 Methodology

3.1 Research Methodology

To design and develop our Ontology based Knowledge Management CRA Framework proposed in this paper, we follow the design science research (DSR) methodology. DSR offers a structured approach to solving real-world problems by creating innovative artifacts (Hevner et al., 2004; March & Smith, 1995). More specifically, we employ the DSR methodology process model for information systems research detailed in Peffers et al. (Peffers et al., 2007), this research is conducted. The DSR process model for our research is thus shown in Figure 1.

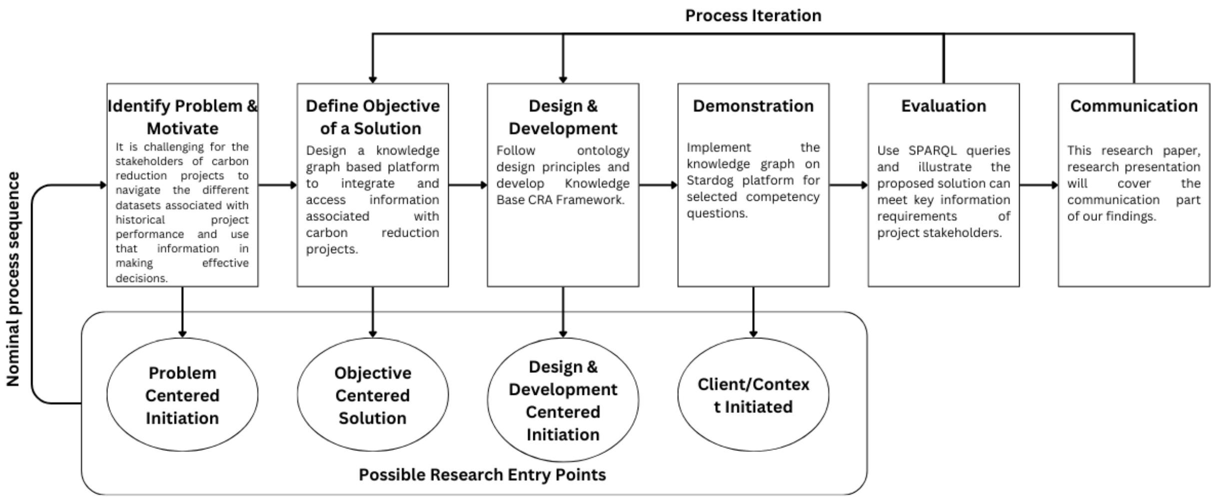

Figure 1: Steps of the DSR Process model adapted from (Peffers et al., 2007)

The DSR Process model illustrated in Figure 1 has six distinct steps. It prescribes how our research is conducted. In the first step, *"Identify Problem & Motivate"*, we analysed the challenges faced in implementing carbon reduction projects by different stakeholders. The focus in this paper is on the two stakeholders, *carbon reduction project developer* and *project manager*. The challenges faced by those stakeholders are identified from the literature on the current practice, as detailed in Section 1 and 2.

The second step, *"Define the Objective of a Solution"*, is associated with defining a solution or artifacts that helps to tackle the problem identified in the first step. In the context of our research, the artifact is the Knowledge Base Carbon Reduction Assessment (CRA) framework, which consists of CRA ontology and CRA Knowledge Graph. Within the semantic web standards, an ontology is used to represent the data so it can be understood by the human and computer alike. Ontologies are the data models that are used to explain the semantics of domain concepts using ontological terminology, such as classes (entities) and relationships (properties) (Yahya et al., 2021). Knowledge graph is where ontology is expanded with data to capture real world information. The proposed ontology and knowledge graph together can be used as a tool to resolve challenges in managing fragmented and diverse carbon market datasets and provide an interactive medium for analysing comprehensive data.

In the *"Design and Development"* step, we developed a Knowledge Base Carbon Reduction Assessment (CRA) framework. Ontology modelling and knowledge graph construction are the two main components within the Knowledge Base CRA framework. We follow ontology development methods and best practices to develop an ontology model (Noy & McGuinness, 2001; Suárez-Figueroa et al., 2011). The process of ontology development can be approached in three different ways, namely top–down, bottom–up, and combination(Tamašauskaitė & Groth, 2023). The top–down approach begins with the identification of the most general concepts in a domain, followed by the specialisation of these concepts into more detailed ones. On the other hand, the bottom–up approach starts with the identification of the most specific classes or leaves of the hierarchy, which are then grouped into more general concepts. We followed a bottom-up approach in identifying entities and relationships for the ontology, by consolidating the metadata from different carbon market data sources accessed through different voluntary carbon registries. Constructing this ontology model helped us define a common vocabulary, resolving any labelling or semantic inconsistencies, so that data from the different registries could be linked within our knowledge graph. This ontology modelling process of CRA framework is detailed in section 3.2 and Knowledge graph construction process is detailed in section 3.3.

As the fourth *"Demonstration"* step of the DSR process model, we implemented this ontology in the cloud-based Stardog enterprise knowledge graph platform. This implementation is done following the RDF/OWL notation and the semantic web standards. RDF (Resource Description Framework) and OWL (Web Ontology language) are the foundation for semantic web standards and a data model designed for interchanging data on the web, which helps integration and sharing of heterogeneous unstructured data or structured data with different schema models systematically[1]. In our research, we use Stardog to create and manage knowledge graphs. The data model in Stardog is a directional graph, and supports the ontology model based on RDF[2].

Then the knowledge graph was populated with data coming from multiple data sources, creating a powerful knowledge graph representing the carbon market projects. Details of the implementation are provided in Section 4.1.

As the fifth step, for evaluating the knowledge graph, we developed a set of competency questions (CQs) that capture the information needs of different stakeholders and used the SPARQL query to query the knowledge graph and observe the capability of the knowledge graph to answer the competency questions. Details of evaluation and results are provided in Section 4. This paper and research presentation are used as the medium of communication of our proposed approach and findings, completing the last step of the DSR process.

3.2 Ontology Modelling

When developing the ontology for our CRA framework, we followed four stages: Identifying domain concepts, defining relationships, specifying properties, and ontology refinement. At the first stage of ontology modelling, we identified concepts associated with carbon reduction project, resulting in 14 concepts (Classes), using metadata extracted from different carbon market data sources including, Gold Standard[3], Verra[4] and American Carbon Registry[5]. Different voluntary registries were observed to use different nomenclature to represent the same concept. For example, Verra uses the term "Proponent,"

[1] https://www.w3.org/RDF/
[2] https://docs.stardog.com/tutorials/rdf-graph-data-model
[3] https://www.goldstandard.org/
[4] https://verra.org/
[5] https://americancarbonregistry.org/

and Gold Standard uses "Project developer," and both terms refer to entities who developed the carbon project. Considering this, we developed the concepts with formal definitions that map into data source vocabularies so that ambiguity of labels and semantics would not be a problem when using different registry datasets. The identified classes are shown as circles in Figure 2, and they form the foundational building blocks of our ontology. Then we established the relationships between the classes (shown by arrows in Figure 2) and attributes associated with each class.

Upon completing this task, the first version of the ontology is built, and we flowed an iterative process to refine the ontology further. We demonstrated the first version to the industry experts and validated its utility against real-world carbon market operations. Their feedback helped us to refine the model and ensure the completeness and accuracy of our ontology.

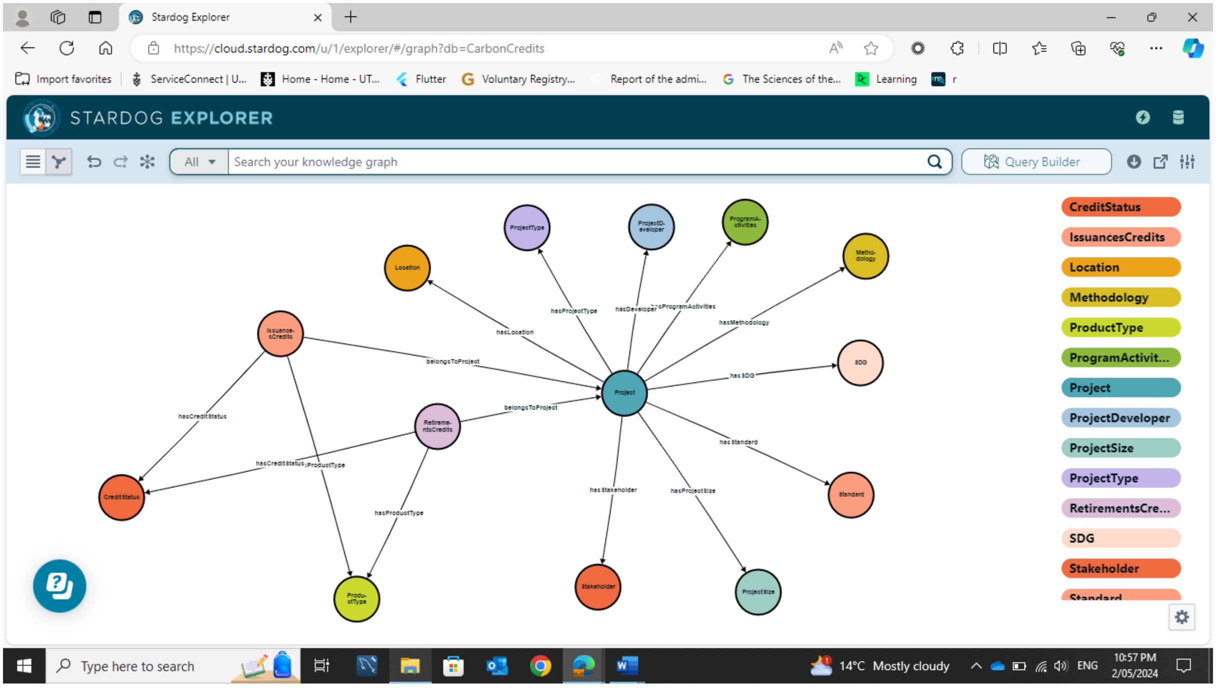

Figure 2: Illustration of the Knowledge Graph within Stardog

3.3 Knowledge graph construction

Schematic Representation of the data pipeline for creating and querying the CRA Framework we followed is shown in Figure 3.

The proposed CRA framework is divided into four layers. The bottom layer presents the Raw data: all the data associated with the carbon market, obtained from the voluntary carbon registries. For this study, data collection is done from different sources, including, but not limited to, Gold Standard, Verra and American Carbon Registry. The data comprises details on renewable energy initiatives, the technologies employed, the participating organizations, and the performance measures related to these projects.

The second layer is the data preprocessing unit programmed in Python, where the obtained raw data is cleaned and integrated, identifying inconsistencies, duplicates and incomplete data. To ensure uniformity across various data formats and to ascertain the accuracy and consistency of the data used for building the knowledge graph, we adopted standard data integration techniques. Data integration helps merge data from different source systems, providing a unified view of these combined data, which is critical for maintaining data integrity. Before creating the knowledge graph, all the data were translated into a relational database format. This process aims to ensure structural integrity and allows for efficient storage and manipulation of data, which are critical for the construction and utility of the knowledge graph.

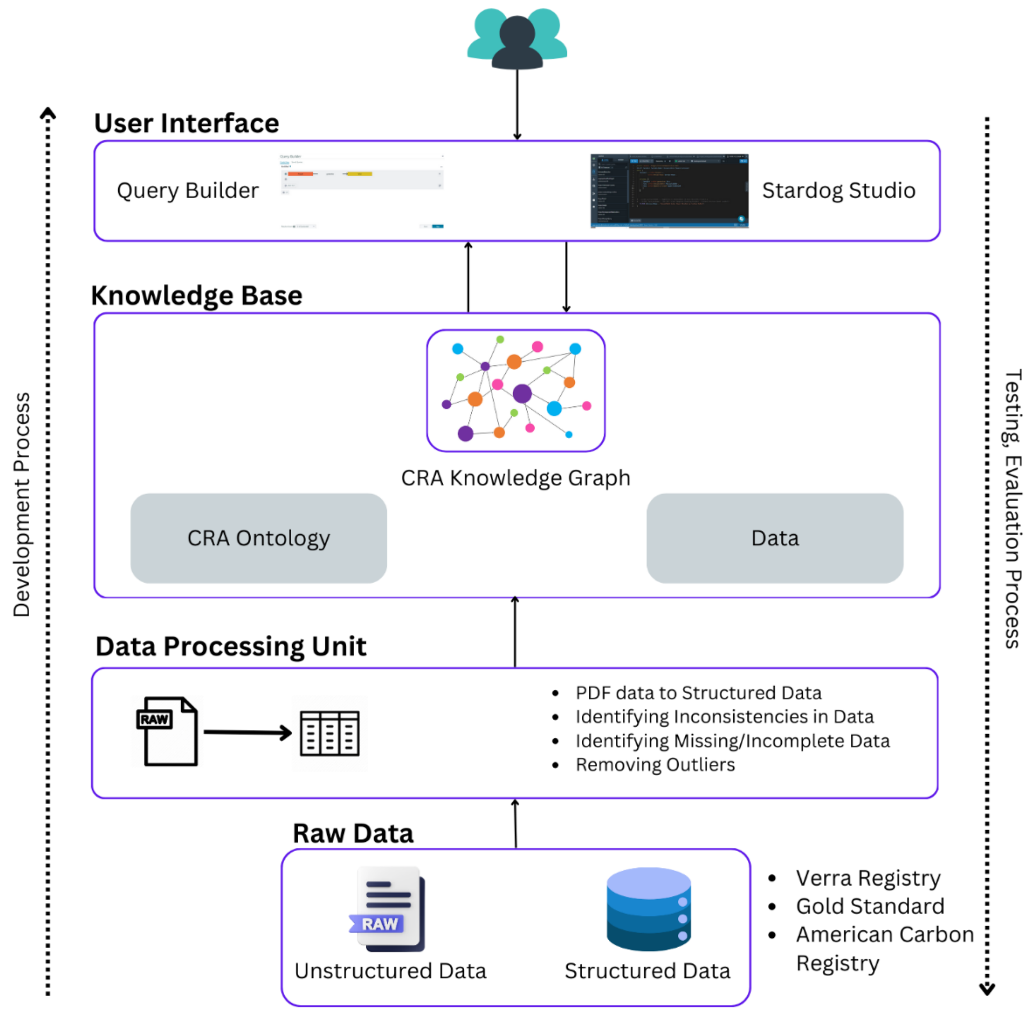

Figure 3: Schematic Representation of CRA Framework including the data pipeline for populating and querying the proposed

Knowledge Base is the core of the CRA framework, that contains the CRA ontology and associated real world data, known combinedly as the CRA knowledge graph. This layer is implemented using the enterprise knowledge graph platform Stardog. The knowledge graph is designed to be dynamic, continuously updated to accommodate new data extracted though lower layers, and to adapt following the changes in the carbon market landscape.

The top layer is the User Interface that enables stakeholders to interact with the Knowledge Base. We have implemented this layer utilising two modules in the Stardog platform: Query Builder, which is used for the visual exploration of knowledge graphs and Studio, which is used to run SPARQL (i.e. the standard query language for sematic web) queries and obtain data.

4 Demonstration and Evaluation

To evaluate the proposed CRA framework , we developed a set of requirements, also known as Competency questions (CQs), that represent the information required by stakeholders, who plan to use the proposed solution, presented as questions in natural language. The ability of the knowledge graph to provide answers to these questions adequately through SPARQL queries, indicates that the framework is well designed to meet its intended requirements (Grüninger & Fox, 1995). Set of CQs we designed for the knowledge graph, in consultation with domain experts, are given in Table 1.

Competency Questions	Description
1. How many projects have been implemented in a specific geographic region (e.g. India)?	This query helps us navigate the projects based on the location.
2. What SDGs were met by the project that was successfully implemented (e.g. "Eritrea Community Boreholes" Project)?	It helps us to get insights into what sustainable development goals a particular project has met and contributed to.
3. Who are the project developers involved in developing projects with project type afforestation and reforestation?	Investors interested in certain project types can explore investment opportunities by knowing the right developers associated with the project type.
4. What are the different carbon registries integrated in the Knowledge Graph?	Helps understand the dataset being used by the knowledge graph.
5. What are the methodologies used by the projects (e.g. "Clean and Efficient Cooking and Heating Project China" Project)??	Relating projects with specific methodologies being used to develop the project.
6. How many credits have been issued/retired by a certain project (e.g. "Clean and Efficient Cooking and Heating Project China" Project)?	It helps discover how many credits the project has generated and retired. It helps get insights into the project performance.
7. Which project developer has developed a project with a large project size?	It helps discover and rank the project developers' performance.

Table 1: Competency Questions

In our evaluation, we translated the CQs in Table 1 into SPARQL query, and executed in Stardog Studio. Figure 4 shows an example SPARQL query execution of competency question 3 from Table 1: *Who are the project developers involved in developing projects with project type afforestation and reforestation?*. The SPARQL query retrieves all the project developer names involved in developing projects with project type *afforestation* and *reforestation*.

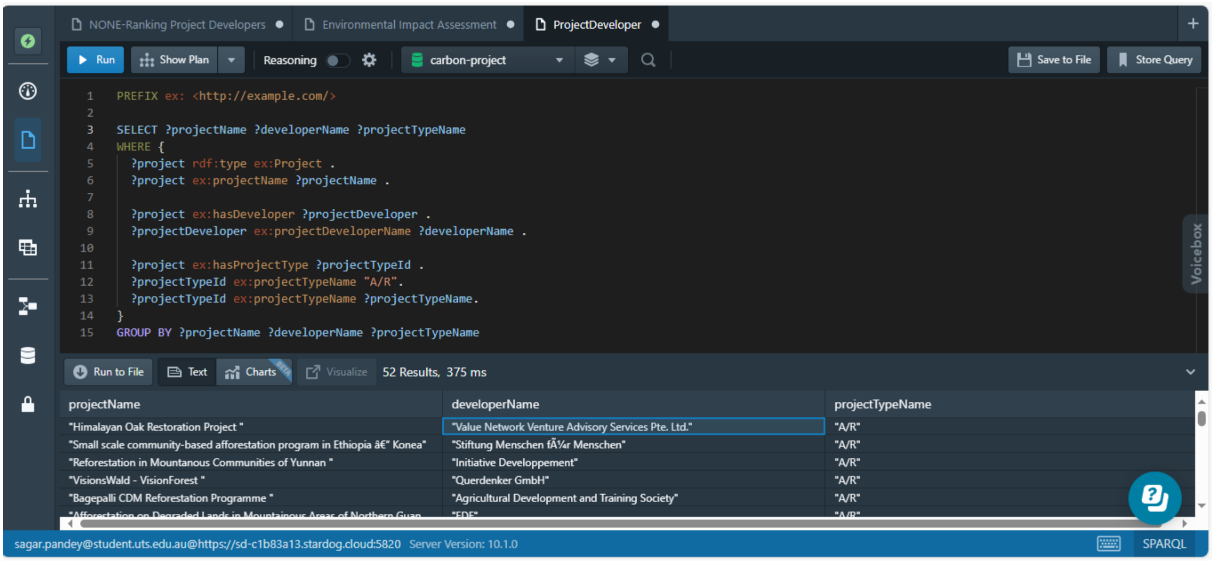

Figure 4: Sample SPARQL query and its output on Stardog Studio for competency question 3 from Table 1

Table 2 summarizes the outputs obtained for all queries associated with seven competency questions, and outcome of its manual validation. Our implementation was able to answer all the competency questions mentioned above, and we cross checked the accuracy of generated answers generated manually though the data sources.

CQs		Query Output	Correct?
1.	How many projects have been implemented in a specific geographic region (e.g.: India)?	500	Yes
2.	What SDGs were met by the project that was successfully implemented (e.g. “Eritrea Community Boreholes” Project)?	Climate action; Good health and well-being; Gender equality; Clean water and sanitation	Yes
3.	Who are the project developers involved in developing projects with project type afforestation and reforestation?	Value Network Venture Advisory Services Pte. Ltd.; Querdenker GmbH; FDF; and so on.	Yes
4.	What are the different carbon registries integrated in the Knowledge Graph?	Gold Standard; Verra; American Carbon Registry	Yes
5.	What are the methodologies used by the projects (e.g. “Clean and Efficient Cooking and Heating Project China” Project)?	GS Methodology for Improved Cook stoves and Kitchen Regimes v2.	Yes
6.	How many credits have been issued/retired by a certain project (e.g. “Clean and Efficient Cooking and Heating Project China” Project)?	1778775	Yes
7.	Which project developer has developed a project with a large project size?	CO2 balance UK ltd; MicroEnergy Credits; Proyecto Mirador; Swiss Carbon Value Ltd.; and so on.	Yes

Table 2: Summary of query result obtained for all the CQs

5 Contributions and Limitations

5.1 Contributions

Developing Knowledge base CRA framework helped address the challenges faced by stakeholders in the carbon reduction projects, which we identified in the initial phase of our research (Problem Identification phase). Moreover, the proposed knowledge graph helps carbon reduction project developers and project managers in their decision-making process and provides an efficient way to access relevant information. It makes information sharing easy and promotes transparency over the carbon credit transactions in the carbon market. In the long run, this will help organisations to promote a low-carbon economy and contribute to climate change mitigation. Finally, this paper provides a new way of thinking about interrelated data that exists across different parties involved in the carbon reduction assessment, that can lead to further innovations in this domain.

5.2 Limitations

Key limitation we see in adapting this work is the effort organisations will have to invest in managing large number of heterogenous data sources feeding into the knowledge graph and keeping the knowledge graph up to date for constantly generated new data. Even though this effort will be well rewarded down the line by saving time spent on individual project assessment tasks, further guidance and automated tools for knowledge graph management and quality checking can also reduce the adaptation barrier. Another challenge is updating the ontology- the underlying schema for knowledge graph, to support future needs of stakeholders, such as new regulatory reporting requirements. Knowledge graph is a flexible and adaptable entity that supports evolving information, so we believe the effort for that challenge will be minimal. Addressing these challenges also requires close interaction with data providers and consistent efforts to keep the knowledge graph up to date.

6 Conclusion and Future Enhancements

In conclusion, our paper presents a novel approach to enhance carbon reduction assessments with the Knowledge graph based CRA framework. The proposed framework enables stakeholders to query and retrieve critical information about renewable energy projects with innovative technologies, fostering data-driven decision-making and collaboration. While certain challenges and limitations exist, the potential for positive impacts on the carbon reduction project is significant. We believe that our work lays the foundation for future research and development in knowledge graph-based systems for sustainability and carbon reduction projects.

As future work, we propose to expand the CRA knowledge graph to encompass a broader spectrum of sustainability-related data, including social and environmental impact metrics. This expansion would offer stakeholders a more comprehensive view of project performance. Secondly, we want to focus on integrating geographical maps with our knowledge graph to enable queries based on the precise location of projects, providing stakeholders a map based interface to interact with both the knowledge graph and project locations without the nuisance of Stardog interfaces and SPARQL query. Lastly, we intend to incorporate advanced machine learning techniques to enhance the query system's natural language processing capabilities as well as data import pipeline. This enhancement will enable automated and context-aware queries and data integration, further improving the overall functionality and usability of the system.

7 Reference

Abu-Salih, B. (2021). Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*, *185*, 103076.

Grüninger, M., & Fox, M. S. (1995). The role of competency questions in enterprise engineering. In *Benchmarking—Theory and practice* (pp. 22-31). Springer.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, *15*(4), 251-266.

Mehling, M. (2009). Global Carbon Market Institutions. *An assessment of governance challenges and functions in the carbon market.*

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. In: Stanford knowledge systems laboratory technical report KSL-01-05.

Othman, S., and Beydoun, G. 2010. "A disaster management metamodel (DMM) validated," in: Kang BH., Richards D. (eds) Knowledge Management and Acquisition for Smart Systems and Services. Springer, Berlin, Heidelberg, pp 11–125

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Peng, Y., Guo, J., & Liu, Y. (2023). Comparative Analysis of China's Pilot Carbon Market Based on Knowledge Graph. 2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS),

Schneider, E. W. (1973). Course Modularization Applied: The Interface System and Its Implications For Sequence Control and Data Analysis.

Singhal, A. (2012). *Introducing the Knowledge Graph: things, not strings*. Google. https://blog.google/products/search/introducing-knowledge-graph-things-not/

Su, N., & Wang, Z. (2022). Visual analysis of global carbon mitigation research based on scientific knowledge graphs. *International Journal of Environmental Research and Public Health*, *19*(9), 5766.

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Fernández-López, M. (2011). The NeOn methodology for ontology engineering. In *Ontology engineering in a networked world* (pp. 9-34). Springer.

Tamašauskaitė, G., & Groth, P. (2023). Defining a knowledge graph development process through a systematic review. *ACM Transactions on Software Engineering and Methodology*, *32*(1), 1-40.

Tang, X., Feng, Z., Xiao, Y., Wang, M., Ye, T., Zhou, Y., Meng, J., Zhang, B., & Zhang, D. (2023). Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development. *Geoscience Frontiers*, *14*(5), 101426.

United Nations Climate Change. (2022). *Mechanisms under the Kyoto Protocol*. UNFCCC. https://unfccc.int/process/the-kyoto-protocol/mechanisms#:~:text=The%20Kyoto%20mechanisms%3A&text=Help%20countries%20with%20Kyoto%20commitments

Walsh, V. R., & Toffel, M. W. (2023). *What Every Leader Needs to Know About Carbon Credits*. Harvard Business Review. https://hbr.org/2023/12/what-every-leader-needs-to-know-about-carbon-credits

Wu, J., Orlandi, F., O'Sullivan, D., & Dev, S. (2022). LinkClimate: An interoperable knowledge graph platform for climate data. *Computers & Geosciences*, *169*, 105215.

Yahya, M., Breslin, J. G., & Ali, M. I. (2021). Semantic web and knowledge graphs for industry 4.0. *Applied Sciences*, *11*(11), 5110.

Yu, M., Rabhi, F. A., & Bandara, M. (2024). Ontology-Driven Architecture for Managing Environmental, Social, and Governance Metrics. *Electronics*, *13*(9), 1719.

Zhou, K., & Li, Y. (2019). Carbon finance and carbon market in China: Progress and challenges. *Journal of Cleaner Production*, *214*, 536-549.

Zou, X. (2020). A survey on application of knowledge graph. Journal of Physics: Conference Series,

Walsh, V. R., & Toffel, M. W. (2023, December 15). *What Every Leader Needs to Know About Carbon Credits*. Harvard Business Review. https://hbr.org/2023/12/what-every-leader-needs-to-know-about-carbon-credits