From Research Outputs to Reporting Practice: Towards A Neuro-Symbolic Approach to Managing Emission Conversion Factors for Transparency of Carbon Accounting

Alisa Holm[1,2,*], Milan Markovic[1,2], Matthias Kuhnert[3], and Paul N. Williams[4]

[1]University of Aberdeen, Computing Science, Aberdeen, UK
[2]University of Aberdeen, Interdisciplinary Institute, Aberdeen, UK
[3]University of Aberdeen, Institute of Biological and Environmental Sciences, Aberdeen, UK
[4]Queens University Belfast, Institute for Global Food Security, Belfast, UK
[*]Corresponding author: a.holm.24@abdn.ac.uk.

Introduction

Accurately quantifying and comparing emissions from commercial activities (e.g., manufacturing energy use, fleet mileage) is essential for meeting global climate goals [1], and for compliance with evolving carbon reporting requirements [2, 3]. Emission Conversion Factors (ECFs) enable such comparisons by translating activity data into equivalent carbon dioxide (CO_2e) values (e.g., miles driven multiplied by a specific ECF value) to quantify generated emissions [4, 5].

ECFs are typically research artefacts generated by academic or commercial studies focusing on greenhouse gas emissions (e.g., CO_2, CH_4) related to specific activities (e.g., driving an average petrol car in a specific country) [4, 5]. As such, ECFs are calculated for diverse purposes and may be intended to be used only in certain contexts.

ECFs can be accessed through many sources, such as the IPCC database [6], peer-reviewed literature, and national databases [7]. Companies also frequently employ carbon calculator tools which implement different calculation methodologies and ECFs for a particular domain [8]. However, in all these cases the research context (the intended use) of ECFs or details of their application are not recorded as part of an emission report and cannot be easily retrieved and evaluated using automated means. Due to the lack of such transparency, the ECFs used in emission reports may become disconnected from their intended context and both the comparability and accuracy of emission impacts will reduce [9].

We argue that to improve transparency, contextual information should be recorded by documenting provenance of ECFs and their consecutive application in carbon footprint reports. Such provenance traces would include detailed information about the ECF (e.g., what input data ECF converts, where the ECF can be applied, etc.) as well as details of the calculation process (e.g., what ECFs were used in the calculation, what input data was multiplied using ECF, etc.). Similar calls for better documentation of intended use and other properties of research artefacts to ensure their correct use have also emerged, for example, in the Machine Learning domain including Model Cards for machine learning models [10] and Datasheets for datasets [11]. However, these approaches often face a barrier to adoption due to the limited automation of information recording mechanisms, which makes the documentation process labour-intensive.

Therefore, our research focuses on:

- Preserving contextual information about ECFs to guide their correct application in practice
- Automating execution and documentation of carbon footprint calculations using the most appropriate ECFs
- Enabling automated querying and comparison of carbon footprint reports

Semantic Technologies and Transparency

Semantic web technologies such as ontologies and knowledge graphs can support solutions for challenges related to transparency, as these can be used to effectively record and access provenance of different data modifications [12, 13]. Provenance of Emission Calculation Ontology (PECO) and the Emission Conversion Factor Ontology (ECFO) [14], provide the means to represent emission calculations (including ECFs) and record their provenance. Previous work on integrating data for emission reporting [15] and capturing provenance of reports has used similar semantic approaches [16]. However, these studies show that knowledge graphs are hard for non-experts to adopt because of the specialised skills required for managing and querying them. Past research into the use of semantic technologies in this domain has also highlighted the need to address challenges with the labourious manual alignment of ECFs to the correct activity data, and the creation and retrieval of information about the calculation processes and its provenance trace [14, 17].

LLMs as Automation Agents

Large Language Models (LLMs) have rapidly gained popularity in task automation for their usability in a variety of contexts. The ability of these models to provide natural language answers based on retrieved information (e.g., from a knowledge graph of ECFs) could help improve the adoption of semantic technologies by the industry by automating labourious creation and management of semantic data structures. This is especially the case with enhancements to context awareness by methods such as Retrieval Augmented Generation (RAG) [18, 19]. For example, recent research into the use of RAG to help optimise the ability of LLMs to assist in the creation of SPARQL queries has shown promising results for automation of expert and labourious tasks in retrieving information from a knowledge graph [19,20]. Previous research has also demonstrated that SPARQL queries can be used to create provenance traces and retrieve input data for semantically represented emissions calculation formulas [17]. However, the manual creation of these queries is a significant drawback to the method and would benefit from automation.

Approach

Our research therefore aims to answer the following research question:

"How effectively can Large Language Models generate SPARQL queries that correctly integrate, document and inform about individual elements of the emission calculation process by generating and querying knowledge graphs?"

Since knowledge graphs can be both queried, created, and modified using SPARQL queries, we will explore the ability of LLMs to generate such queries to better sustain a connection between an ECF used in an emission report and the original context for that ECF. We have identified the following tasks from the emissions reporting process that could benefit from LLM-based SPARQL support:

Linking Data and Calculation Methods Formulating queries to retrieve appropriate input data and corresponding ECFs in the context of a specific calculation methodology.

Provenance Trace Creation Generating SPARQL Create queries to annotate the reported emission data with provenance traces containing the information about the ECFs, input data and calculation methods used.

Human-in-the-Loop Compliance Monitoring Translating natural language questions about details of emission reports into SPARQL queries executed over a knowledge graph of provenance records.

Conclusion & Future Work

In this paper we have highlighted the need for additional provenance information to support the correct use of research artefacts such as ECFs in practical settings. We argued that while semantic web technologies provide suitable mechanisms to represent and query the required provenance records, neuro-symbolic approach implementing LLMs may be needed to limit manual human tasks to improve adoption of such solutions.

Our future work will focus on evaluating the capabilities of LLMs for the tasks outlined in this paper in the context of real-world case studies. We will evaluate the LLMs' ability to produce semantically correct SPARQL queries, the effects of hallucination, and novel workflow design patterns combining human-in-the-loop, LLM agents, and semantic rules to enhance the compliance monitoring of carbon footprint reporting.

Acknowledgments

This work was supported by the UKRI AI Centre for Doctoral Training in Sustainable Understandable agri-food Systems Transformed by Artificial INtelligence (SUSTAIN) [grant reference: EP/Y03063X/1]. We also acknowledge the generous support of alumni and friends in establishing the University of Aberdeen's Interdisciplinary Institute, which partly enabled this research, including Dr Jane Hellman Caseley (MBChB 1956), Professor Patrick Meares (DSc 1959), Nancy Miller (MA 1942), Norman Robertson, Dr Ian Slessor (MBChB 1956) and Anne Young (MA 1957)

References

[1] F. Yu, Q. Yuan, X. Sheng, M. Liu, L. Chen, X. Yuan, D. Zhang, S. Dai, Z. Hou, Q. Wang, and Q. Ma, "Understanding carbon footprint: An evaluation criterion for achieving sustainable development," *Chinese Journal of Population, Resources and Environment*, 2024. doi: https://doi.org/10.1016/j.cjpre.2024.11.001 .

[2] A. Amel-Zadeh and Q. Tang, "Managing the shift from voluntary to mandatory climate disclosure: The role of carbon accounting," *The British Accounting Review*, 2025. doi: 10.1016/j.bar.2025.101594 .

[3] E. Union, "Directive (eu) 2022/2464 of the european parliament and of the council of 14 december 2022 amending regulation (eu) no 537/2014, directive 2004/109/ec, directive 2006/43/ec and directive 2013/34/eu, as regards corporate sustainability reporting," Official Journal of the European Union, L322, pp. 15–80, December 2022, accessed: 11 March 2025. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022L2464

[4] M. Bertolini, P. Duttilo, and F. Lisi, "Accounting carbon emissions from electricity generation: A review and comparison of emission factor-based methods," *Applied Energy*, 2025. doi: 10.1016/j.apenergy.2025.125992 .

[5] M. Hiete, U. Berner, and O. Richter, "Calculation of global carbon dioxide emissions: Review of emission factors and a new approach taking fuel quality into consideration," *Global Biogeochemical Cycles*, 2001. doi: 10.1029/2000GB001261 .

[6] I. P. on Climate Change (IPCC), "Ipcc emission factor database (efdb)," Online Database, 2023, accessed: 14.10.2025. [Online]. Available: https://www.ipcc-nggip.iges.or.jp/EFDB/main.php

[7] D. for Energy Security and N. Zero, "Greenhouse gas reporting: conversion factors 2025," https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2025, 2025, accessed: 14 October 2025.

[8] C. Whittaker, M. C. McManus, and P. Smith, "A comparison of carbon accounting tools for arable crops in the united kingdom," *Environmental Modelling Software*, 2013. doi: 10.1016/j.envsoft.2013.03.015 .

[9] C. Adewale, J. P. Reganold, S. Higgins, R. D. Evans, and L. Carpenter-Boggs, "Agricultural carbon footprint is farm specific: Case study of two organic farms," *Journal of cleaner production*, 2019. doi: 10.1016/j.jclepro.2019.04.253 .

[10] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, "Model cards for model reporting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ser. FAT* '19. ACM, 2019. doi: 10.1145/3287560.3287596 .

[11] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. III, and K. Crawford, "Datasheets for datasets," 2021. [Online]. Available: https://arxiv.org/abs/1803.09010

[12] I. Naja, M. Markovic, P. Edwards, W. Pang, C. Cottrill, and R. Williams, "Using knowledge graphs to unlock practical collection, integration, and audit of ai accountability information," *IEEE Access*, 2022. doi: 10.1109/ACCESS.2022.3188967 .

[13] I. M. Putrama and P. Martinek, "Heterogeneous data integration: Challenges

and opportunities," *Data in Brief*, 2024. doi: 10.1016/j.dib.2024.110853 .

[14] M. Markovic, D. Garijo, S. Germano, and I. Naja, "Tec: Transparent emissions calculation toolkit," in *The Semantic Web – ISWC 2023*. Springer Nature Switzerland, 2023. doi: 10.1007/978-3-031-47243-5$_5$.

[15] S. Germano, C. Saunders, I. Horrocks, and R. Lupton, "Use of semantic technologies to inform progress toward zero-carbon economy," in *The Semantic Web – ISWC 2021*, ser. Lecture Notes in Computer Science. Springer, 2021. doi: 10.1007/978-3-030-88361-4_39 .

[16] D. Beloin-Saint-Pierre, A. C. Junior, A. Brilhante, S. Colcher, J.-P. Calbimonte, K. Främling, L. Lefèvre, F. Loureiro, F. Poux, and S. Mayer, "Transparent integration and sharing of life cycle sustainability data with provenance," in *The Semantic Web – ISWC 2020*, ser. Lecture Notes in Computer Science. Springer, 2020. doi: 10.1007/978-3-030-62466-8$_2$3 .

[17] M. Markovic, S. Germano, D. Garijo, P. Edwards, A. Li, T. Ayall, R. Ramsey, and G. Leontidis, "Farm explorer: A tool for calculating transparent greenhouse gas emissions," *CEUR Workshop Proceedings*, 2024.

[18] X. Pan, J. van Ossenbruggen, V. de Boer, and Z. Huang, "A rag approach for generating competency questions in ontology engineering," 2025. [Online]. Available: https://arxiv.org/abs/2409.08820

[19] V. Emonet, J. Bolleman, S. Duvaud, T. M. de Farias, and A. C. Sima, "Llm-based sparql query generation from natural language over federated knowledge graphs," 2025. doi: 10.48550/arXiv.2410.06062 .

[20] X. Pan, V. de Boer, and J. van Ossenbruggen, "Firesparql: A llm-based framework for sparql query generation over scholarly knowledge graphs," 2025. doi: 10.48550/arXiv.2508.10467 .