CarbOnto: Data Integration Towards Net Zero - Extended Abstract – CTDGSI 2025

Luiz Santos[1], Regina Braga[1], José Maria N. David[1]

[1]Departamento de Ciência da Computação - Universidade Federal de Juiz de Fora (UFJF) Caixa Postal 20.010 – 36036-900 – Juiz de Fora – MG – Brazil

fernando.santos@estudante.ufjf.br, regina.braga@ufjf.br (advisor), jose.david@ufjf.br (co-advisor)

Abstract. *This work presents an architecture called CarboFarm, that encompasses an ontology for the syntactic and semantic integration of heterogeneous databases related to the agricultural domain. By utilizing ontology, we contribute to the standardization and interpretation of domain concepts and add semantic information to support the generation of GHG (Greenhouse Gas) inventories on farms. The Artificial Intelligence (AI) component processes integrated data to derive GHG information. The inventories can identify imbalances between emissions and gas stocks, helping search for solutions to neutralize emissions. We conducted a that using CarboFarm can support the balancing of gases.*

Resumo. *Este trabalho apresenta uma arquitetura, chamada CarboFarm, que engloba uma ontologia, para integração sintática e semântica de bancos de dados heterogêneos relacionados a dados agrícolas. Com a ontologia, pretendemos contribuir para a padronização e interpretação de conceitos do domínio, adicionando informações semânticas para suporte à geração de inventários de GEE (Gases de Efeito Estufa) em fazendas. Um componente de Inteligência Artificial (IA) processa os dados integrados para derivar informações de GEE. Os inventários podem identificar desequilíbrios entre emissões e estoques de gases, auxiliando na busca de soluções para neutralizar as emissões. Realizamos um estudo de caso, mostrando que o uso da CarboFarm pode dar suporte ao balanceamento dos gases.*

1. Topic Overview

This work presents an architecture called CarboFarm that encompasses an ontological model – named CarbOnto – with the objective of syntactically and semantically integrating sets of heterogeneous data related to agriculture. CarboFarm also has an AI component for analyzing integrated data to generate knowledge to support decisions in searching for more socio-ecologically sustainable agriculture and generating carbon credits.

The agricultural sector is one of the leading global emitters of GHG. Land use and land use change (LUC), caused by the conversion of native habitats into agricultural land, is considered the second largest source of greenhouse gas emissions on the planet, representing approximately 23% of the total [IPCC, 2021] [Fankhauser *et al.* 2022].

Care is needed in land use and livestock farming to find solutions to implement balanced food production with sustainable development.

Net zero carbon emissions, or simply net zero emissions, refers to a state where the net amount of GHG emissions released into the atmosphere is zero. The aim is to balance gas emissions and their removal so that the net balance is zero. We must align net-zero emissions with broader sustainable development goals, which entail an equitable transition to net-zero emissions, socio-ecological sustainability, and the pursuit of broad economic opportunities [IPCC, 2021] [Fankhauser *et al.* 2022]. Nevertheless, challenges in calculating greenhouse gas emissions on rural properties, as many variables are involved, and we need methods for calculating a complete inventory for an entire farm.

This work addresses social, economic, environmental, and technical issues, considering the interdependence between these areas in the study's context, with an interest in offering a solution to a real-world problem. In this way, it has a direct relationship with the *"Sociotechnical View of Information Systems"* one of the "Grand Research Challenges in Information Systems in Brazil 2016 - 2026" [Boscarioli *et al.* 2017].

2. Objectives and Relevance of the Research

As a specific contribution, this research proposes:

- An ontological model for data integration focusing on the standardization and interoperability of information between measurement systems;
- An AI component designed to process integrated data, derive information and generate knowledge;
- Using the knowledge generated to support constructing GHG inventories, facilitate decision-making, and assist the generation of carbon credits.

3. State of the Art and Practice

After analyzing the literature, we found some studies related to our approach. Some studies and projects [Harris 2021] [MapBiomas 2023] have developed methodologies for calculating GHG emissions, but do not focus on farm inventories. Other studies that use or reference ontologies [Konys 2018] [Di *et al.* 2022], address sustainability-related aspects but they focus on different domains. None of them presents a solution that involves theoretical and practical approaches, like software applications, related to semantic data integration, artificial intelligence, and decision support systems related to agricultural GHG emissions.

Our work proposes a global and transparent solution that, by addressing gaps found in the agricultural GHG emissions domain, can generate more comprehensive farm inventories, provided that the datasets are available. Furthermore, with the results obtained, we hope to verify, at a macro level, the collective impact of climate policies related to agriculture.

4. Scientific Methods Employed

We used DSR (Design Science Research) [Hevner *et al.* 2007] to develop CarboFarm. DSR can be characterized as an approach to building solutions by elaborating,

developing, and evaluating artifacts. It begins with identifying problems in a real-world application environment, considering that DSR focuses on designing, implementing, and verifying solutions that meet specific objectives. We carried out two DSR cycles, using the DSR-Model [Pimentel *et al.* 2020], with the artifact's representation, the application's context, and the practical and theoretical approaches to answer the research questions proposed for the study.

5. Results Obtained

This work presented a set of concepts related to GHG inventories on farms, syntactic and semantic data integration, ontology, AI, and decision support systems. We identified integrating heterogeneous databases of GHG emissions data and sequestration sources, and AI, can contribute to environmentally and economically sustainable agriculture. Data integration and analysis can generate knowledge to support farmer´s decision-making and possibly entry into the carbon market.

We discussed a case study with integrated public datasets from institutions in Brazil and detailed the land cover and use balance of Brazilian farms. Additionally, our proposal aims to contribute to discussions on more ecological agricultural practices, considering that this sector is one of the leading global emitters of GHG.

Throughout our research, we did not find studies that addressed ontologies and AI components built to address climate issues with an emphasis on GHG inventories in agriculture. CarboFarm can help to fill this gap, offering integrated data for AI components that aim to generate knowledge to support rural producers' decisions in constructing GHG inventories and generating carbon credits.

6. Conclusion

The CarboFarm architecture, notably the CarbOnto ontology, was suitable for integrating data from heterogeneous bases that can contribute to generating GHG inventories. Inventories generated by integrated data allow us to understand rural properties' GHG balance status. Understanding the property's status will enable us to mitigate emissions and make production more sustainable.

In our work, we consider that the generation of agricultural GHG inventories involves technical and social issues, mainly related to small rural producers. The engagement of public and private organizations is also necessary, considering that agriculture can significantly contribute to mitigating climate change and biodiversity by adopting a comprehensive approach incorporating political, social, and economic changes combined with technological advances and innovations. This work is aligned with the “Grand Research Challenges in Information Systems in Brazil 2016 - 2026*”* [Boscarioli *et al.* 2017].

Throughout the development of this research, we published some papers:

- Santos, L., David, J. M. N. and Braga, R. (2023). “Uma abordagem para suporte à decisão no processo de geração de créditos de carbono em propriedades rurais’. In Simpósio Brasileiro de Sistemas Colaborativos (SBSC) (pp. 1-15). SBC. DOI: https://doi.org/10.5753/sbsc.2023.229059.

- Santos, L. F., Gomes, J., Braga, R., David, J. M. N. and Stroele, V. (2023). "Towards a seco for carbon credit control". In *2023 IEEE/ACM 11th International Workshop on Software Engineering for Systems-of-Systems and Software Ecosystems (SESoS)* (pp. 13-21). IEEE. DOI: 10.1109/SESoS59159.2023.00008
- Santos, L., Braga, R., David, J. M. N. and Stroele, V. (2024). CarbOnto: Data Integration Towards Net Zero. *IEEE Access*. DOI: 10.1109/ACCESS.2024.3477259

References

Boscarioli, C.; Araujo, R. M.; Maciel, R. S. P. (2017). "I GranDSI-BR - Grand Research Challenges in Information Systems in Brazil 2016 - 2026". Special Committee on Information Systems (CE-SI). Brazilian Computer Society (SBC). ISBN: [978-85-7669-384-0]. 2017. 184p. DOI: https://doi.org/10.5753/sbc.2884.0.

Di, F., Wu, J., Li, J. and Wang, Y. (2022). "Research on the knowledge graph based life cycle carbon footprint labeling representation model for electromechanical products". *In 5th International Conference on Computer Information Science and Application Technology (CISAT 2022),* vol. 12451, pp. 393-399. DOI: 10.1117/12.2656476.

Fankhauser, S. *et al.* (2022). "The meaning of net zero and how to get it right". Nat. Clim. Chang. 12, 15–21. DOI: 10.1038/s41558-021-01245-w.

GHG Protocol Agricultural Guidance (2014). "Interpreting the Corporate Accounting and Reporting Standard for the agricultural sector". Greenhouse Gas Protocol, 103 p. *World Resources Institute*.

Harris, N.L. *et al.* (2021) "Global maps of twenty-first century forest carbon fluxes". Nat. Clim. Chang. 11, 234–240. DOI: 10.1038/s41558-020-00976-6.

Hevner, Alan R. et al. (2004). "Design science in information systems research". MIS quarterly, p. 75-105. DOI: 10.2307/2518625.

IPCC. Intergovernmental Panel on Climate Change. Climate Change 2021: The Physical Science Basis. Available: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM_final.pdf. Access in: 03 mar 2025.

Konys, A. (2018). "An ontology-based knowledge modelling for a sustainability assessment domain. Sustainability", 10 (2), 300. DOI: 10.3390/su10020300.

MapBiomas. (2023). "Mapeamento anual do estoque de carbono orgânico do solo no Brasil 1985-2021 (coleção beta)", "Annual mapping of soil organic carbon stock in Brazil 1985-2021 (beta collection)", DOI: 10.58053/MapBiomas/DHAYLZ, MapBiomas Data, V1" (in portuguese).

Pimentel, Mariano; Filippo, Denise; dos Santos, Thiago Marcondes. (2020). "Design Science Research: pesquisa científica atrelada ao design de artefatos". RE@ D-Revista de Educação a Distância e eLearning, v. 3, n. 1, p. 37-61.