

Conference Abstract

ChatGPT as a Semantic Engineering Assistant: Lessons from Ontology Design in the Agricultural Biodiversity Domain

Filipi Miranda Soares^{‡§}, Luís Ferreira Pires[§], Luiz Olavo Bonino da Silva Santos[§], Debora Pignatari Drucker[†], Dilvan de Abreu Moreira[¶], Fernando Elias Corrêa[¶], Kelly Rosa Braghetto[¶], Alexandre Cláudio Botazzo Delbem[¶], Antonio Mauro Saraiva[¶]

‡ INRAE, Montpellier, France

§ University of Twente, Enschede, Netherlands

| Embrapa Agricultura Digital, Campinas, Brazil

¶ Universidade de São Paulo, São Paulo, Brazil

Corresponding author: Filipi Miranda Soares (filiwigp2011@gmail.com)

Received: 03 Dec 2025 | Published: 17 Dec 2025

Citation: Soares F, Ferreira Pires L, Santos LBdaS, Drucker D, Moreira DA, Corrêa FE, Braghetto K, Delbem AB, Saraiva A (2025) ChatGPT as a Semantic Engineering Assistant: Lessons from Ontology Design in the Agricultural Biodiversity Domain. Biodiversity Information Science and Standards 9: e181272.

<https://doi.org/10.3897/biss.9.181272>

Abstract

Modeling species names in biodiversity ontologies is particularly difficult in multilingual contexts, where semantic conflation often occurs. A good example is the common name "pimenta." In Brazilian Portuguese, experts usually refer to *Capsicum* spp. (chili peppers), while its direct translation "pepper" in English often denotes *Piper nigrum* (black pepper) (Soares et al. 2025a). In Brazilian markets, however, *Piper nigrum* is more accurately associated with "pimenta-do-reino" ("pimenta-negra"). This issue was observed on Wikipedia, when translating the Portuguese page for "pimenta" into English, the entry switches from *Capsicum* spp. to black pepper (*Piper nigrum*), showing how easily semantic drift can appear in multilingual data modeling.

The correct association between common names used in the agricultural market with species would be a way to avoid the misunderstanding of these cultural differences. However, another challenge in vocabulary management emerges, which is how to

manage species names in ontologies to keep them updated as the taxonomy itself updates. Some agriculturally controlled vocabularies, such as Agrotermos (Telles et al. 2024) lack automated mechanisms for updating taxonomic classifications. For example, *Prochilodus cearensis*, *Prochilodus scrofa*, and *Prochilodus margravii* are all listed in Agrotermos as preferred terms, i.e., the authorized, standard term selected to represent a concept in a controlled vocabulary, while according to the Global Biodiversity Information Facility ([GBIF](#)) Backbone Taxonomy (GBIF Secretariat 2023) these names are synonyms, as shown in Table 1.

Table 1.

Examples of synonymous scientific names and their corresponding accepted names identified by ChatGPT after sending a request to the GBIF API (Soares et al. 2025a).

Target name	Synonym of
<i>Prochilodus cearensis</i>	<i>Prochilodus brevis</i>
<i>Prochilodus scrofa</i>	<i>Prochilodus lineatus</i>
<i>Prochilodus margravii</i>	<i>Prochilodus argenteus</i>

When developing the [Agricultural Product Types Ontology \(APTO\)](#), which was designed to represent products traded in Brazilian agricultural markets based on Agrotermos and AGROVOC, we proposed two approaches using generative AI, specifically [OpenAI](#)'s ChatGPT-4, as a semantic engineering assistant to automate the inclusion of scientific names in the ontology:

1. Prompt-based queries with a plugin accessing the GBIF API
2. A ChatGPT-generated Python script that converted GBIF taxonomy data into Web Ontology Language (OWL) format

These AI-supported methods automated the construction of APTO's "Organism" module, integrating taxonomic hierarchies and managing synonyms. ChatGPT effectively identified synonymy (e.g., see Table 1) and reduced manual labor in ontology development. The first approach is no longer reproducible since OpenAI has replaced plugins by GPTs. As such, we are currently developing a GPT named Taxonomy OWLizer 2.0*¹, which is an evolution of the first approach described in that paper.

Concerns about scalability, reproducibility, and hallucinations (false, made-up information) remain, highlighting the need for expert oversight throughout the process. When ChatGPT was used without API access, hallucinations appeared more frequently. For instance, when asked to check a list of plant species names for typos, it incorrectly suggested that *Euterpe edulis* was a synonym of *Euterpe oleracea*, even though both are recognized as distinct species in widely used catalogues such as the GBIF Backbone Taxonomy (Soares et al. 2025a).

This case study demonstrates that generative AI can support but not yet replace human-led ontology development. It also emphasizes AI's potential contribution to biodiversity informatics, particularly for managing evolving and multilingual vocabularies. All tools and source code related to our work are archived on Zenodo (Soares et al. 2025b). Detailed protocols are provided in Soares et al. 2025a.

Keywords

generative AI, LLM, Large Language Model, ontology, semantic web, agriculture, product types, scientific names

Presenting author

Filipi Miranda Soares

Presented at

Living Data 2025

Grant title

São Paulo Research Foundation. Grant numbers 21/15125-0 and 22/08385-8

Conflicts of interest

The authors have declared that no competing interests exist.

References

- GBIF Secretariat (2023) GBIF Backbone Taxonomy. Checklist dataset. URL: <https://doi.org/10.15468/39omei>
- Soares FM, Saraiva AM, Ferreira Pires L, et al. (2025a) Exploring a Large Language Model for Transforming Taxonomic Data into OWL: Lessons Learned and Implications for Ontology Development. Data Intelligence 7 (2): 265-302. [In en]. <https://doi.org/10.3724/2096-7004.di.2025.0020>
- Soares FM, Saraiva AM, Ferreira Pires L, et al. (2025b) Supporting Data for "Exploring ChatGPT-4 for Transforming Taxonomic Data into OWL: Lessons Learned and Implications for Ontology Development". v.2. Zenodo. Release date: 2025-3-06. URL: <https://doi.org/10.5281/zenodo.14982527>
- Telles MA, Jonquet C, Almeida BTd, et al. (2024) Embrapa's contributions to integrate Brazilian agricultural vocabularies: Agrotermos in AgroPortal. In: Fonseca CM, et al. (Ed.) Proceedings of the 17th Seminar on Ontology Research in Brazil (ONTOBRAS

2024) and 8th Doctoral and Masters Consortium on Ontologies (WTDO 2024), 3905. Seminar on Ontology Research in Brazil, Vitória, 7-10 October 2024. CEUR-WS.org, 5 pp. [In en]. URL: <https://ceur-ws.org/Vol-3905/short8.pdf>

Endnotes

*1 <https://chatgpt.com/g/g-68c7c49f900c8191aa8856659572545d-taxonomy-owlizer-2-0>