

UNIVERSIDADE ESTADUAL DE PONTA GROSSA
SETOR DE CIÊNCIAS AGRÁRIAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

WESLEY VINICIUS FERNANDES

FRAMEWORK PARA INTEGRAÇÃO DE PADRÕES DE METADADOS
AGROMETEOROLÓGICOS

PONTA GROSSA

2026

WESLEY VINICIUS FERNANDES

FRAMEWORK PARA INTEGRAÇÃO DE PADRÕES DE METADADOS
AGROMETEOROLÓGICOS

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, como requisito parcial para a obtenção do título de Mestre em Computação Aplicada.

Área de Concentração: Computação para Tecnologias em Agricultura.

Orientadora: Prof.^a Dr.^a. Maria Salete Marcon Gomes Vaz

PONTA GROSSA

2026

F363 Fernandes, Wesley Vinicius
 Framework para integração de padrões de metadados agrometeorológicos /
 Wesley Vinicius Fernandes. Ponta Grossa, 2026.
 182 f.

 Dissertação (Mestrado em Computação Aplicada - Área de Concentração:
 Computação para Tecnologias em Agricultura), Universidade Estadual de Ponta
 Grossa.

 Orientadora: Profa. Dra. Maria Salete Marcon Gomes Vaz.

 1. Metadados. 2. Interoperabilidade. 3. Agrometeorologia. I. Vaz, Maria
 Salete Marcon Gomes. II. Universidade Estadual de Ponta Grossa. Computação
 para Tecnologias em Agricultura. III.T.

CDD: 005.3



UNIVERSIDADE ESTADUAL DE PONTA GROSSA
Av. General Carlos Cavalcanti, 4748 - Bairro Uvaranas - CEP 84030-900 - Ponta Grossa - PR - <https://uepg.br>

TERMO

DE APROVAÇÃO

Wesley Vinicius Fernandes

“FRAMEWORK PARA INTEGRAÇÃO DE PADRÕES DE METADADOS AGROMETEOROLÓGICOS”

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora:

Profa. Dra. Maria Salete Marcon Gomes Vaz (PPCA/UEPG - Presidente)

Prof. Dr. José Carlos Ferreira da Rocha (PPCA/UEPG)

Prof. Dr. Luciano Heitor Gallegos Marin (PPGGI/UFPR)

Ponta Grossa, 02 de julho de 2026.



Documento assinado eletronicamente por **Maria Salete Marcon Gomes Vaz, Coordenador(a) do Programa de Pós-Graduação em Computação Aplicada - Mestrado**, em 21/07/2026, às 22:14, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Jose Carlos Ferreira da Rocha, Coordenador(a) do Programa de Pós-Graduação em Computação Aplicada - Mestrado**, em 22/07/2026, às 11:28, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



Documento assinado eletronicamente por **Luciano Heitor Gallegos Marin, Usuário Externo**, em 24/07/2026, às 15:37, conforme Resolução UEPG CA 114/2018 e art. 1º, III, "b", da Lei 11.419/2006.



A autenticidade do documento pode ser conferida no site <https://sei.uepg.br/autenticidade> informando o código verificador **3213891** e o código CRC **BB277E74**.

Agradeço a cada pessoa que esteve ao meu lado durante toda a minha trajetória para chegar até a conclusão desta etapa. Cada um sabe da importância e da minha gratidão.

É preciso sonhar, mas com a condição de crer em nosso sonho, de observar com atenção a vida real, de confrontar a observação com nosso sonho, de realizar escrupulosamente nossas fantasias. Sonhos, acredite neles.
(Lênin)

RESUMO

A crescente demanda por dados agrometeorológicos confiáveis frente às mudanças climáticas impulsiona a agricultura de precisão, mas a exploração dessas informações esbarra na heterogeneidade e no volume massivo gerado na era do *Big Data*. A falta de interoperabilidade entre diferentes padrões de metadados cria as chamadas "ilhas de informação", dificultando análises conjuntas devido às divergências estruturais e de granularidade, especialmente entre os padrões de dados climáticos/geoespaciais, como o *WMO Core Metadata Profile* e a ISO 19115, e os de agrobiodiversidade, como o *Darwin Core*. Diante dessa problemática, este trabalho tem como objetivo desenvolver um *framework*, denominado AgroMeta, para a integração de metadados agrometeorológicos, promovendo a interoperabilidade estrutural e semântica entre esses ecossistemas. A pesquisa guia-se pelas diretrizes da Metodologia *NeOn* para a concepção de uma Ontologia de Alinhamento, desenvolvida para atuar como uma ponte semântica interdomínios. O *framework* é implementado computacionalmente em linguagem Python, na forma de uma *API RESTful*, operando como um motor de triplificação, seguindo uma arquitetura de entrada, transformação e saída, capaz de extrair e converter formatos sintáticos divergentes, como XML, GeoJSON e arquivos CSV, para o modelo estrutural de grafos em RDF. Para sanar a ambiguidade de termos e alcançar a interoperabilidade terminológica, a solução orquestra o uso de ontologias fundacionais, como GeoSPARQL, SOSA e *OWL-Time*, aliado ao vocabulário controlado multilíngue AGROVOC. A validação prática do Grafo de Conhecimento unificado foi conduzida no ambiente *GraphDB* por meio de requisições SPARQL, orientadas por Questões de Competência (CQs) aplicadas a um cenário agrônomo crítico: o monitoramento do complexo praga-cultura da *Spodoptera frugiperda* e do milho. Os resultados atestam a arquitetura, demonstrando potencial para resolver o conflito de granularidade, harmonizando catálogos climáticos macroscópicos com observações biológicas e medições de sensores em resolução fina. Além disso, garantiu a extração da proveniência institucional em conformidade estrita aos princípios FAIR. O modelo supera a mitigação superficial da sintaxe e atua como um orquestrador de conhecimento e Sistema de Suporte à Decisão (DSS), transformando bases isoladas em inteligência analítica e preditiva para o monitoramento agrícola.

Palavras-chave: Metadados; Interoperabilidade Semântica; Agrometeorologia; Darwin Core; WMO Core Metadata Profile; AGROVOC.

ABSTRACT

The growing demand for reliable agrometeorological data in the face of climate change drives precision agriculture; however, harnessing this information is hindered by the high heterogeneity and massive volume generated in the Big Data era. The lack of interoperability between different metadata standards creates so-called "information islands," making joint analyses difficult due to profound divergences in structure and granularity. This is especially true between climatic and geospatial data standards, such as the WMO Core Metadata Profile and ISO 19115, and agrobiodiversity standards, such as Darwin Core. Faced with this issue, this study aims to develop a framework, named AgroMeta, for the integration of agrometeorological metadata, promoting structural and semantic interoperability between these ecosystems. The research is guided by the NeOn Methodology guidelines for designing an Alignment Ontology, developed to act as an interdomain semantic bridge. The framework is computationally implemented in Python as a RESTful API, operating as a triplication engine. It follows an input, transformation, and output architecture capable of extracting and converting divergent syntactic formats, such as XML, GeoJSON, and CSV files, into the RDF graph structural model. To resolve term ambiguity and achieve terminological interoperability, the solution orchestrates the use of foundational ontologies, such as GeoSPARQL, SOSA, and OWL-Time, combined with the AGROVOC multilingual controlled vocabulary. The practical validation of the unified Knowledge Graph was conducted in the GraphDB environment using SPARQL queries, guided by Competency Questions (CQs) applied to a critical agronomic scenario: the monitoring of the *Spodoptera frugiperda* and maize pest-crop complex. The results attest that the architecture successfully resolved the granularity conflict, harmonizing macroscopic climate catalogs with biological observations and fine-resolution sensor measurements. Furthermore, it ensured the extraction of institutional provenance in strict compliance with the FAIR principles. In conclusion, the model surpasses superficial syntax mitigation, acting as an authentic knowledge orchestrator and Decision Support System (DSS), transforming isolated databases into analytical and predictive intelligence for agricultural monitoring.

Keywords: Metadata; Semantic Interoperability; Agrometeorology; Darwin Core; WMO Core Metadata Profile; AGROVOC.

LISTA DE FIGURAS

Figura 1	–	Classes de esquema de metadados.....	42
Figura 2	–	Fluxograma PRISMA do processo de busca e seleção de estudos.....	97
Figura 3	–	Hierarquia de classes e modelada na Ontologia de Alinhamento	114
Figura 4	–	Definição de Domínio e Alcance das Object Properties no Protégé	116
Figura 5	–	Hierarquia das propriedades de dados modeladas na Ontologia de Alinhamento	117
Figura 6	–	Hierarquia de classes inferida após a execução do raciocinador HermiT.....	119

LISTA DE QUADROS

Quadro 1	– Elementos fundamentais e representação estrutural de uma ontologia	29
Quadro 2	– Tipos de termos que compõem um grafo RDF	30
Quadro 3	– Exemplo Representativo do Dicionário de Dados ISO 19115	44
Quadro 4	– Classes do Darwin Core.....	46
Quadro 5	– Propriedades e Funções do WCMP	50
Quadro 6	– Os 25 conceitos de nível superior do AGROVOC	55
Quadro 7	– Análise Comparativa dos Padrões de Metadados Agrometeorológicos	59
Quadro 8	– Síntese das Iniciativas e Frameworks de Integração Analisados.....	74
Quadro 9	– Estratégias de Busca	94
Quadro 10	– Questões de Competência e os Padrões de Metadados Agrometeorológicos Envolvidos	102
Quadro 11	– Prefixos e Namespaces configurados na Ontologia de Alinhamento ...	113
Quadro 12	– Matriz de Atributos Tipados (Data Properties) da Ontologia de Alinhamento.....	117
Quadro 13	– Inicialização do Grafo de Conhecimento e parse da TBox no rdflib....	123
Quadro 14	– Processo de triplificação estrutural, instanciação de URIs e formatação de Literais	125
Quadro 15	– Módulo de Desambiguação Terminológica e Ancoragem Semântica..	126
Quadro 16	– Módulos de extração e instanciação de Meteorologia e Proveniência .	130
Quadro 17	– Estabelecimento de interligações interníveis e serialização do Grafo de Conhecimento	131
Quadro 18	– Consulta SPARQL elaborada para a validação da CQ01	137
Quadro 19	– Resultados inferidos pela execução da consulta SPARQL da CQ01 ...	138
Quadro 20	– Consulta SPARQL elaborada para a validação da CQ02	141
Quadro 21	– Resultados inferidos pela execução da consulta SPARQL da CQ01 ...	143
Quadro 22	– Consulta exploratória direcionada ao endpoint oficial do AGROVOC	144
Quadro 23	– Extensão semântica modelada no Grafo de Especialista	145
Quadro 24	– Consulta SPARQL elaborada para a validação da CQ03	148
Quadro 25	– Resultados inferidos pela execução da consulta SPARQL da CQ03 ...	149
Quadro 26	– Consulta SPARQL elaborada para a validação da CQ04	153
Quadro 27	– Resultados da correlação entre clima extremo e impacto cultural (CQ04)	155

LISTA DE ABREVIATURAS E SIGLAS

ABox	<i>Assertional Box</i>
ACDD	<i>Attribute Convention for Dataset Discovery</i>
AEMET	<i>Agencia Estatal de Meteorología</i>
AgMES	<i>Agricultural Metadata Element Set</i>
API	<i>Application Programming Interface</i>
APSIM	<i>Agricultural Production Systems sIMulator</i>
ARC	<i>Annotated Research Context</i>
BFO	<i>Basic Formal Ontology</i>
BGP	<i>Basic Graph Patterns</i>
BI	<i>Business Intelligence</i>
CF	<i>Climate and Forecast</i>
CPM	<i>Complex Properties Model</i>
CQ	<i>Competency Question (Questão de Competência)</i>
CSV	<i>Comma-Separated Values</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
D-entailment	<i>Datatype entailment</i>
DSS	<i>Decision Support System (Sistema de Suporte à Decisão)</i>
DSSAT	<i>Decision Support System for Agrotechnology Transfer</i>
DwC	<i>Darwin Core</i>
DwC-A	<i>Darwin Core Archive</i>
ECMWF	<i>European Centre for Medium-Range Weather Forecasts</i>
EML	<i>Ecological Metadata Language</i>
Embrapa	<i>Empresa Brasileira de Pesquisa Agropecuária</i>
EnvO	<i>Environment Ontology</i>
ERA5	<i>ERA5 hourly single level reanalysis data</i>
ETL	<i>Extract, Transform, Load (Extração, Transformação e Carga)</i>
FAIR	<i>Findable, Accessible, Interoperable, Reusable</i>
FAO	<i>Food and Agriculture Organization of the United Nations</i>

FDO	<i>FAIR Digital Objects</i>
GBIF	<i>Global Biodiversity Information Facility</i>
GDC	<i>Global Discovery Catalogue</i>
GEMET	<i>General Multilingual Environmental Thesaurus</i>
GGBN	<i>Global Genome Biodiversity Network</i>
GODAN	<i>Global Open Data for Agriculture and Nutrition</i>
GRIB	<i>GRIdded Binary</i>
GSC	<i>Genomic Standard Consortium</i>
GUI	<i>Graphical User Interface</i> (Interface Gráfica do Usuário)
HDF	<i>Hierarchical Data Format</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IA	Inteligência Artificial
IDI	<i>Interoperability through Data Integration</i>
INMET	Instituto Nacional de Meteorologia
IoT	<i>Internet of Things</i> (Internet das Coisas)
IRI	<i>Internationalized Resource Identifier</i>
ISA	<i>Investigation/Study/Assay</i>
ISC	<i>International Chronostratigraphic Chart</i>
ISO	<i>International Organization for Standardization</i>
JSON	<i>JavaScript Object Notation</i>
JSON-LD	<i>JSON for Linking Data</i>
LOBD	<i>Linked Open Biodiversity Data</i>
LOD	<i>Linked Open Data</i>
MIxS	<i>Minimum Information about any (x) Sequence</i>
MQTT	<i>Message Queuing Telemetry Transport</i>
NAL	<i>National Agricultural Library</i>
NALT	<i>National Agricultural Library Thesaurus</i>
NetCDF	<i>Network Common Data Form</i>
NISO	<i>National Information Standards Organization</i>
OBDA	<i>Ontology-Based Data Access</i>
OBIS	<i>Ocean Biogeographic Information System</i>

OECD	<i>Organisation for Economic Co-operation and Development</i>
OGC	<i>Open Geospatial Consortium</i>
OWL	<i>Web Ontology Language</i>
PaaS	<i>Platform as a Service</i>
PCR	Pesquisa Computacional Reprodutível
PLN	Processamento de Linguagem Natural
PRISMA	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
PROV-O	<i>Provenance Ontology</i>
QB	<i>RDF Data Cube</i>
QUDT	<i>Quantities, Units, Dimensions and Data Types</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>RDF Schema</i>
RDI	<i>Research Data Infrastructures</i>
REST	<i>Representational State Transfer</i>
ROA	<i>Resource-Oriented Architecture</i> (Arquitetura Orientada a Recursos)
RPW	<i>Red Palm Weevil</i>
RSL	Revisão Sistemática da Literatura
SaaS	<i>Software as a Service</i>
SEO	<i>Search Engine Optimization</i>
SHACL	<i>Shapes Constraint Language</i>
SKOS	<i>Simple Knowledge Organization System</i>
SOA	<i>Service-Oriented Architecture</i> (Arquitetura Orientada a Serviços)
SOSA	<i>Sensor, Observation, Sample, and Actuator</i>
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
SQL	<i>Structured Query Language</i>
SSN	<i>Semantic Sensor Network</i>
SSSOM	<i>Simple Standard for Sharing Ontology Mappings</i>
SWEET	<i>Semantic Web for Earth and Environmental Terminology</i>
SWRL	<i>Semantic Web Rule Language</i>
TBox	<i>Terminological Box</i>

TDWG	<i>Biodiversity Information Standards</i>
TSDB	<i>Time Series Database</i>
UML	<i>Unified Modeling Language</i>
URI	<i>Uniform Resource Identifier</i>
VKG	<i>Virtual Knowledge Graph</i>
VOS	<i>Virtuoso Open-Source Edition</i>
VRE	<i>Virtual Research Environment</i> (Ambiente Virtual de Pesquisa)
W3C	<i>World Wide Web Consortium</i>
WCMP	<i>WMO Core Metadata Profile</i>
WIS2	<i>WMO Information System 2.0</i>
WMO	<i>World Meteorological Organization</i>
XML	<i>Extensible Markup Language</i>
XSD	<i>XML Schema Definition</i>
YAML	<i>YAML Ain't Markup Language</i>
YAMAML	<i>YAMA Mapping Language</i>

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	CONTEXTUALIZAÇÃO	15
1.2	OBJETIVOS	18
1.2.1	Objetivo Geral.....	18
1.2.2	Objetivos Específicos	18
1.3	ESTRUTURA DA DISSERTAÇÃO.....	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	METADADOS E INTEROPERABILIDADE	21
2.2	TECNOLOGIAS E ABORDAGENS PARA INTEGRAÇÃO SEMÂNTICA...	27
2.2.1	Arquitetura e Padrões da Web Semântica.....	28
2.2.2	Aplicações Práticas de Integração Semântica.....	35
2.3	PADRÕES DE METADADOS UTILIZADOS NA AGROMETEOROLOGIA	40
2.3.1	Padrão ISO 19115 e o contexto geoespacial.....	41
2.3.2	Padrão Darwin Core e a agrobiodiversidade	45
2.3.3	Padrão WMO Core Metadata Profile e as observações meteorológicas	49
2.3.4	Padrão AgMES e a evolução para o AGROVOC.....	54
2.3.5	Convergências e Divergências dos Padrões para a Integração Agrometeorológica	58
2.4	ESTADO DA ARTE	63
2.4.1	Aplicações de Padrões em Dados Climáticos e Geoespaciais	63
2.4.2	Aplicações de Padrões em Agrobiodiversidade.....	70
2.4.3	Iniciativas e Frameworks para Integração de Padrões de Metadados	73
2.5	DESAFIOS E LACUNAS NA LITERATURA	90
3	MATERIAL E MÉTODOS	92
3.1	REVISÃO SISTEMÁTICA DA LITERATURA.....	92
3.2	ESPECIFICAÇÃO DE REQUISITOS DA ONTOLOGIA.....	99
3.3	ESPECIFICAÇÃO DO FRAMEWORK E ESTRATÉGIA DE MODELAGEM	103
3.4	AMOSTRAGEM E COLETA DE DADOS.....	106
4	DESENVOLVIMENTO E IMPLEMENTAÇÃO	112

4.1	CONSTRUÇÃO DA ONTOLOGIA DE ALINHAMENTO	112
4.2	IMPLEMENTAÇÃO COMPUTACIONAL DO MOTOR DE TRIPLIFICAÇÃO	119
4.3	VALIDAÇÃO, RESULTADOS E DISCUSSÃO	133
4.3.1	Auditoria Estrutural do Grafo de Conhecimento Gerado	134
4.3.2	Validação da CQ01: Resolução do Conflito de Granularidade	136
4.3.3	Validação da CQ02: Desambiguação Terminológica e Lacunas Ontológicas	140
4.3.4	Validação da CQ03: Garantia de Proveniência e Reprodutibilidade.....	147
4.3.5	Validação da CQ04: Correlação Agrometeorológica e Suporte à Decisão	151
5	CONSIDERAÇÕES FINAIS	158
	REFERÊNCIAS.....	165
	APÊNDICE A – CÓDIGO-FONTE DO MOTOR DE TRIPLIFICAÇÃO EM PYTHON	172
	APÊNDICE B – ONTOLOGIA DE ALINHAMENTO COMPLETA.....	179

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Em um cenário de mudanças climáticas, a crescente demanda por dados agrometeorológicos confiáveis para a gestão agrícola evidencia a importância de sistemas agrícolas baseados em dados precisos e integrados. Conforme apontado pela OECD e FAO (2023), os ecossistemas agrícolas enfrentam o desafio de promover a resiliência e a sustentabilidade ambiental frente a essas mudanças.

Nesse contexto, eventos climáticos extremos representam um risco direto à produtividade das culturas, exigindo o monitoramento integrado para a criação de sistemas de alerta antecipado, como ressaltam Zeginis *et al.* (2024), por exemplo. Consequentemente, a exploração estruturada dessas informações impulsiona a adoção de tecnologias de agricultura de precisão, tornando-se uma necessidade para a tomada de decisão (Krisnawijaya *et al.*, 2025; Sigler *et al.*, 2025).

A crescente digitalização do setor tem aumentado a produção de dados de diversas fontes, como sensores, satélites e estações meteorológicas, das quais a exploração e a análise possibilitam a agricultura de precisão práticas mais eficientes e sustentáveis (Krisnawijaya *et al.*, 2025). No entanto, Putrama e Martinek (2024) mostram que o volume massivo desses dados torna esse campo de estudo cada vez mais complexo, especialmente na era do chamado *Big Data*, conceituado como o armazenamento especulativo de grandes quantidades de dados que podem ser úteis no futuro (Kleppmann e Riccomini, 2026). Esse acúmulo contínuo e diversificado de dados evidencia a necessidade de integrar dados heterogêneos para viabilizar a reprodutibilidade e a encontrabilidade.

A integração dos dados é dificultada pela diversidade de padrões de metadados utilizados para descrevê-los. No setor agrícola, essa falta de interoperabilidade favorece a formação das chamadas “ilhas de informação”, definidas por Roussaki *et al.* (2023) e Sadeghi *et al.* (2024) como tecnologias incapazes de interagir e compartilhar dados de forma efetiva. Beretta *et al.* (2021) reforçam essa problemática ao destacarem que muitos padrões específicos de domínio adotam estritamente a perspectiva técnica de seus produtores, encapsulando a semântica em jargões próprios, limitando sua compreensão e reutilização por outras áreas do conhecimento.

Na Agrometeorologia, esse desafio torna-se particularmente evidente devido às diferenças estruturais e de granularidade existentes entre os padrões utilizados. Enquanto os

dados meteorológicos e climáticos são descritos pelo *WMO Core Metadata Profile* (WCMP) e pela ISO 19115 em nível de conjunto de dados (*dataset*), as informações relacionadas à agrobiodiversidade são representadas pelo padrão *Darwin Core* (DwC), cuja granularidade está centrada na medição individual. Como consequência, a heterogeneidade dessas estruturas dificulta a integração direta e a realização de análises conjuntas entre variáveis climáticas e indicadores de produtividade agrícola.

Nesse cenário, pode-se citar o *Agricultural Metadata Element Set* (AgMES), desenvolvido pela *Food and Agriculture Organization of the United Nations* (FAO) e concebido para promover a padronização da descrição de recursos informacionais no domínio agrícola (FAO, 2010). Porém, o avanço das tecnologias da Web Semântica evidenciou a necessidade de mecanismos mais robustos para representação semântica e integração entre dados heterogêneos. Como alternativa, passou-se a utilizar vocabulários controlados formais, como o vocabulário multilíngue AGROVOC (FAO, 2021). Contudo, conforme Takezaki *et al.* (2020) ressaltam, o AGROVOC não atua como um padrão de metadado estrutural, mas opera como uma ponte semântica para unificar divergências descritivas, evidenciando a necessidade de uma arquitetura ou modelo de integração intermediário capaz de orquestrar a sua semântica em conjunto com a estrutura dos demais padrões.

Diante disso, para a obtenção e análise de dados agrometeorológicos, torna-se necessário integrar as observações meteorológicas, tipicamente estruturadas no padrão WCMP, com os dados biológicos e agrícolas, registrados em *Darwin Core*, e com dados espaciais, estruturados no Padrão ISO 19115. Uma integração que exige transformações devido às estruturas e granularidades distintas desses padrões de metadados. Conforme demonstrado por Arsenault *et al.* (2020), essa incompatibilidade estrutural primária impede a interoperabilidade direta, exigindo que as diferentes fontes passem por um processo computacional de conversão de formato e harmonização. A automação dessa comunicação entre plataformas, segundo Roussaki *et al.* (2023), reduz a dependência de processos manuais intensivos, garantindo, como ressaltam Zeginis *et al.* (2024), que a integração adequada das informações viabilize a análise precisa da interação entre as variáveis climáticas e a produção agrícola, promovendo a interoperabilidade.

A interoperabilidade corresponde à capacidade de diferentes sistemas ou aplicações trabalharem de forma conjunta e coordenada, visando o compartilhamento de informações de maneira eficiente e sem ambiguidade (Sadeghi *et al.*, 2024). Na prática, iniciativas como o *FAIRagro*, na Alemanha, têm demonstrado o potencial de *frameworks* semânticos para promover a interoperabilidade, adotando os princípios FAIR (*Findable, Accessible,*

Interoperable, Reusable) para viabilizar essa integração entre dados agrometeorológicos (Brizuela *et al.*, 2024).

Entretanto, ainda há lacunas significativas no estado da arte, pois essas federações continuam dependentes de curadoria manual detalhada, evidenciando a necessidade de um modelo de integração para articular esses padrões sem exigir transformações e cruzamentos manuais.

Desse modo, o objetivo desta pesquisa é desenvolver um *framework* para a integração de metadados agrometeorológicos, visando promover a interoperabilidade entre padrões. Para alcançar esse propósito, a condução do estudo desdobra-se em etapas metodológicas específicas. Primeiramente, busca-se analisar as estruturas dos padrões de metadados WCMP, ISO 19115 e *Darwin Core* para identificar as equivalências semânticas e as lacunas de interoperabilidade entre os domínios meteorológico, agrícola e biológico. A partir dessa análise, define-se o mapeamento entre os elementos dos padrões originais e as ontologias de referência selecionadas, estabelecendo a correspondência necessária para a integração estrutural dos dados. Somado a isso, o estudo visa estabelecer regras de desambiguação terminológica utilizando o vocabulário AGROVOC, com o intuito de normalizar os termos descritivos provenientes de fontes heterogêneas.

Na sequência, a operacionalização deste *framework* ocorre pelo desenvolvimento de um Motor de Triplificação, estruturado como um *pipeline* de ETL Semântico (Extração, Transformação e Carga). Implementada sob a forma de uma *API RESTful* em linguagem Python, a solução computacional é responsável por extrair os dados brutos oriundos de sintaxes divergentes, como XML, GeoJSON e CSV, e convertê-los dinamicamente para o modelo estrutural de grafos em RDF. O emprego dessas tecnologias da Web Semântica permite a consolidação de um Grafo de Conhecimento unificado, provendo a infraestrutura necessária para o modelo atuar como repositório e orquestrador de conhecimento passível de inferência lógica.

Finalmente, valida-se a arquitetura proposta por meio de Questões de Competência, definidas especificamente para verificar a efetividade da integração semântica alcançada entre os dados biológicos e meteorológicos. Para materializar essa validação e atestar a aderência da solução a problemas reais do campo, a arquitetura proposta é instanciada em um cenário agrônomo de criticidade e relevância econômica, o monitoramento do complexo praga-cultura da lagarta-do-cartucho (*Spodoptera frugiperda*) e do milho (*Zea mays*) na região do Paraná. A aplicação das Questões de Competência a este estudo de caso demonstra, na prática, a capacidade do *framework* em cruzar a observação biológica granular da praga com as matrizes

climáticas e espaciais, influenciando o seu desenvolvimento e evidenciando o potencial tecnológico do modelo como um Sistema de Suporte à Decisão (DSS) para a Agrometeorologia.

Dessa forma, o *framework* contribui para reduzir o tempo e o esforço manual necessários para integrar e processar grandes volumes de dados agrometeorológicos oriundos de diversas fontes heterogêneas. Ao superar essas barreiras de interoperabilidade, a solução permitirá avanços significativos na tomada de decisão agrícola, fomentando a expansão da ciência de dados e de análises preditivas aplicadas à Agrometeorologia.

1.2 OBJETIVOS

Diante do cenário apresentado e da necessidade latente de mitigar as barreiras estruturais e semânticas, isolando os ecossistemas de dados agrícolas, biológicos e climáticos, esta seção estabelece os objetivos para guiar a condução da pesquisa. Para tanto, os propósitos deste estudo dividem-se em um objetivo geral, norteador da solução para o problema de interoperabilidade, e em objetivos específicos, delineando as etapas metodológicas, conceituais e de engenharia de software necessárias para a consecução da proposta.

1.2.1 Objetivo Geral

- Desenvolver um *framework* para integração de metadados agrometeorológicos, promovendo a interoperabilidade entre padrões.

1.2.2 Objetivos Específicos

Os objetivos específicos são os seguintes:

- Analisar as estruturas dos padrões de metadados WCMP, ISO 19115 e *Darwin Core* para identificar as equivalências semânticas e as lacunas de interoperabilidade entre os domínios meteorológico, agrícola e biológico.
- Definir o mapeamento entre os elementos dos padrões originais e as ontologias de referência selecionadas, estabelecendo a correspondência para a integração estrutural dos dados.
- Estabelecer regras de desambiguação terminológica utilizando o vocabulário AGROVOC para normalizar os termos descritivos provenientes de fontes heterogêneas.

- Implementar computacionalmente a ontologia proposta por meio de um motor de triplificação ETL Semântico, capaz de ingerir dados heterogêneos brutos, como XML, GeoJSON e CSV, e convertê-los para o modelo estrutural de grafos em RDF.
- Validar a arquitetura proposta por meio de Questões de Competência definidas para verificar a efetividade da integração semântica entre os dados agrícolas, biológicos e meteorológicos.

1.3 ESTRUTURA DA DISSERTAÇÃO

Este trabalho está estruturado em cinco capítulos, incluindo este introdutório, os quais detalham desde a fundamentação do problema até a validação computacional do modelo proposto.

O Capítulo 2 destina-se à Fundamentação Teórica e ao Estado da Arte, estabelecendo o alicerce conceitual da pesquisa. Inicialmente, o capítulo aborda os preceitos sobre metadados, interoperabilidade e a aplicação dos princípios FAIR. Na sequência, exploram-se as tecnologias da Web Semântica, como o modelo RDF, a linguagem OWL e o protocolo SPARQL, voltadas para a integração de dados heterogêneos. O texto aprofunda-se no detalhamento dos principais padrões adotados na agrometeorologia: ISO 19115, *Darwin Core* e WCMP, além do vocabulário controlado AGROVOC. A partir de uma Revisão Sistemática da Literatura (RSL), são analisadas criticamente as iniciativas práticas contemporâneas, culminando no mapeamento das lacunas estruturais, semânticas e de granularidade, justificando o desenvolvimento do *framework* proposto.

O Capítulo 3 detalha os procedimentos metodológicos para guiar a condução da pesquisa. A seção descreve o protocolo da Revisão Sistemática da Literatura PRISMA e a estratégia de modelagem apoiada nas diretrizes da Metodologia *NeOn*. Neste capítulo, são estabelecidos os requisitos da ontologia, a formulação das Questões de Competência (CQs) para orientar a validação, a seleção dos recursos não ontológicos de amostragem e a arquitetura planejada para o *pipeline* de integração.

O Capítulo 4 apresenta o desenvolvimento prático e a implementação computacional do *framework*. A seção descreve a construção da Ontologia de Alinhamento (*TBox*) no ambiente *Protégé* e a operacionalização do motor de triplificação (ETL Semântico) desenvolvido em linguagem Python. Na sequência, realiza-se a validação rigorosa do Grafo de Conhecimento unificado (*ABox*) por meio de consultas SPARQL no ambiente *GraphDB*. Os resultados discutem a resolução empírica das Questões de Competência, demonstrando a eficácia da

desambiguação via AGROVOC, a orquestração interdomínios e a descoberta de lacunas estruturais nas bases globais, elevando a arquitetura a uma ferramenta de auditoria ontológica e de suporte à decisão.

Para finalizar, o Capítulo 5 traz as Considerações Finais da pesquisa. O texto retoma o problema central das "ilhas de informação", atesta o cumprimento integral dos objetivos geral e específicos, e consolida as contribuições acadêmicas e metodológicas do estudo. Além disso, apresenta as perspectivas de trabalhos futuros, indicando caminhos para a expansão do mapeamento ontológico e a evolução da pesquisa no ecossistema agrícola.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo estabelecer o alicerce conceitual e tecnológico, fundamentando o desenvolvimento do *framework* proposto nesta pesquisa. Para tanto, a estrutura do texto foi delineada de modo a guiar a compreensão desde os conceitos básicos até a análise crítica das soluções contemporâneas. Inicialmente, discutem-se os preceitos de metadados, a busca pela interoperabilidade e a aplicação dos princípios FAIR. Na sequência, exploram-se as tecnologias da Web Semântica, seguidas pelo detalhamento estrutural dos principais padrões de metadados adotados na Agrometeorologia. Para finalizar, apresenta-se o estado da arte das iniciativas práticas de integração, culminando no mapeamento rigoroso dos desafios e das lacunas na literatura, justificando a necessidade da nova arquitetura delineada neste estudo.

2.1 METADADOS E INTEROPERABILIDADE

Os metadados são conceituados por Gilliland (2008) como “dados sobre os dados”, ou seja, são informações estruturadas para organizar, descrever e localizar objetos de informação em qualquer nível de agregação. Um objeto de informação, por sua vez, é algo acessado e manipulado como uma entidade, podendo ser um conjunto de vários itens, ou até mesmo um banco de dados ou um sistema de registro (Gilliland, 2008).

No contexto da Pesquisa Computacional Reprodutível (PCR), Leipzig *et al.* (2021) expandem essa definição, ressaltando que os metadados podem ser quaisquer dados estruturados em algum padrão que auxiliem na reprodutibilidade. Leipzig *et al.* (2021) definem a PCR como a prática de publicar os resultados científicos juntamente com o conjunto completo de dados, códigos-fonte e especificações do ambiente computacional utilizado em sua produção, possibilitando a obtenção de resultados idênticos aos reportados por outros pesquisadores.

Desta maneira, Leipzig *et al.* (2021) mostram que os metadados desempenham um papel central na sistematização e no gerenciamento de dados, conferindo estrutura e contexto aos recursos. Conforme Haynes (2004), compreender os metadados e suas funções é importante para o desenvolvimento de atividades em setores voltados ao conhecimento, à informação, à cultura e à aprendizagem.

Aprofundando essa necessidade de preservação contínua, Habermann (2020) alerta para o risco da chamada Entropia da Informação, fenômeno definido como o declínio natural e

progressivo dos dados, da informação e da compreensão ocorrida logo após a publicação de um estudo, em que os detalhes vão sendo perdidos ou esquecidos pelos pesquisadores originais com o passar do tempo, fazendo a informação “desaparecer na obscuridade”. Como a mera descoberta dos dados não é suficiente para frear esse processo de degradação, Habermann (2020) destaca que os metadados de qualidade atuam como um “antídoto crítico”, pois evitam a perda do conhecimento, fornecendo a documentação do conteúdo, contexto, qualidade e acessibilidade necessária para se entender e usar os dados de maneira eficaz em longo prazo.

Ao atuarem como esse antídoto, Habermann (2020) ressalta que os metadados permitem que os dados avancem no chamado *Continuum* da Compreensão, ou seja, transitando do estado de dados brutos para a informação, seguindo para o conhecimento e para a sabedoria coletiva. Além de frear a perda de detalhes, os metadados possibilitam o reuso das informações por novas comunidades, gerando a chamada Negentropia (ou entropia negativa), processo no qual o valor e o conteúdo informacional dos dados não se degradam, pelo contrário, eles se expandem e aumentam. A partir de toda essa documentação de contexto e estrutura, Leipzig *et al.* (2021) destacam os metadados como “cola” da reprodutibilidade, pois permitem a transição do dado isolado para o conhecimento robusto e generalizável.

Nesta perspectiva, a PCR exige uma compreensão da análise científica como uma “pilha analítica” composta por cinco níveis: *inputs* (dados brutos), ferramentas (*software*), relatórios (*notebooks*), *pipelines* (fluxos de trabalho) e publicações, como ressaltado por Leipzig *et al.* (2021). Para os objetivos desta pesquisa, foca-se primordialmente no nível de *inputs*, onde os metadados garantem a descoberta, o registro da proveniência e a interoperabilidade inicial.

Nessa necessidade de descoberta dos dados na camada de *inputs*, Řezník *et al.* (2022) trazem o conceito de encontrabilidade, definido como a capacidade dos metadados de atuar em uma camada de busca, permitindo a descoberta autônoma de recursos relevantes para os dados serem efetivamente encontrados em serviços e repositórios. A garantia dessa busca eficiente é um elemento importante para a tomada de decisão baseada em dados, estabelecendo a ligação entre o acesso a dados confiáveis e a confiabilidade do resultado da decisão final.

Diante disso, visando facilitar a reprodutibilidade e a encontrabilidade, Leipzig *et al.* (2021) e Ulrich *et al.* (2022) classificam os metadados em três categorias principais, tendo como base o padrão da *National Information Standards Organization* (NISO):

- a) Metadados descritivos: descrevem um recurso com o objetivo de identificação e descoberta, como um autor e título de livro;
- b) Metadados estruturais: descrevem o modelo de dados, como a organização de um livro em capítulos;

- c) Metadados administrativos: descrevem informações sobre a gestão de um recurso, como a data de publicação e informações de direitos autorais.

Embora o conceito e as categorias de metadados estejam bem estabelecidos, Putrama e Martinek (2024) mostram que a crescente produção de dados torna esse campo de estudo cada vez mais complexo, especialmente na era do *Big Data*, visto a necessidade de soluções tecnológicas cada vez mais avançadas para integrar dados heterogêneos com diversidade de fontes e formatos. No contexto desta pesquisa, superar esse desafio técnico de integração é um passo para viabilizar, na prática, a reprodutibilidade e encontrabilidade.

Um exemplo prático dessa necessidade está no setor agrícola, conforme exemplificado por Krisnawijaya *et al.* (2025), em que a crescente digitalização do setor tem aumentado a produção de dados de diversas fontes como sensores, satélites, estações meteorológicas, entre outros. A exploração e análise desses dados, em especial dos dados agrometeorológicos, impulsiona a agricultura de precisão ao possibilitar práticas mais eficientes e sustentáveis. Para isso, Zeginis *et al.* (2024) destacam a integração desses dados como relevante para analisar a interação entre as variáveis climáticas e a produção agrícola. Esse processamento integrado permite o monitoramento em tempo real e a previsão de produtividade das culturas, além de apoiar a criação de sistemas de alerta antecipado para proteger as plantações contra danos causados por eventos climáticos extremos.

Essa necessidade de integrar dados de fontes variadas reflete uma tendência de busca da "análise holística", conceituado por Sigler *et al.* (2025), como a capacidade de combinar modelos de diferentes áreas para simular cenários complexos. Enquanto a análise de dados tradicionais pode responder "o que aconteceu?", a análise holística responde a perguntas preditivas como "o que vai acontecer?" e "e se acontecer?". A integração desses modelos, no entanto, é dificultada pela falta de padrões de metadados uniformes para descrever modelos e algoritmos, impedindo a interoperabilidade entre os recursos computacionais (Sigler *et al.*, 2025).

Segundo Sadeghi *et al.* (2024), a interoperabilidade em ecossistemas de software vai além da simples capacidade de trocar dados. Ela é definida como o entendimento recíproco e a compatibilidade entre dois ou mais sistemas heterogêneos, permitindo que a comunicação operada por eles, colaborem e utilizem as funções uns dos outros de forma eficaz. Roussaki *et al.* (2023) complementam essa visão ao destacarem que a comunicação entre plataformas não deve se limitar ao simples acúmulo e troca de dados brutos. Os autores afirmam a necessidade de garantir a capacidade de os sistemas interagirem e compreenderem esses dados de forma

automatizada, transformando a abundância de informações em conhecimento produtivo para apoiar a tomada de decisão e reduzir a necessidade de processos manuais intensivos.

Khan e Abonyi (2022) apontam a interoperabilidade em duas dimensões:

- Interoperabilidade técnica: relacionada à capacidade de diferentes sistemas trocarem dados sem a necessidade de intervenção humana;
- Interoperabilidade semântica: relacionada à garantia dos dados serem transmitidos e recebidos com um mesmo significado compartilhado e inequívoco.

A busca pela interoperabilidade é um desafio, tendo em vista o aumento de repositórios e a diversidade de padrões que tornam a compatibilidade um entrave para a integração e reutilização (Sadeghi *et al.*, 2024). No setor agrícola, especialmente, a falta de interoperabilidade é um dos desafios à digitalização, levando à criação das “ilhas de informação”, ou seja, tecnologias que não conseguem interagir e compartilhar dados (Roussaki *et al.*, 2023).

A interoperabilidade semântica, na visão de Habermann (2020), integra os dados heterogêneos, pois garante à informação ser compreendida e reutilizada de forma consistente e produtiva em diferentes domínios, prevenindo a perda de conhecimento. Para isso, Sadeghi *et al.* (2024) especificam, nessa abordagem, a utilização de ontologias, vocabulários e terminologias¹ para fornecer uma especificação de conceituação, permitindo aos sistemas recuperarem o significado exato de um conceito de forma automática e sem ambiguidades. Aplicada ao setor agrícola, Putrama e Martinek (2024) apontam que essa estruturação semântica é capaz de fornecer um vocabulário comum, atuando como ponto de referência, e superando as barreiras técnicas causadas pela heterogeneidade de fonte e pela falta de padrões de metadados uniformes.

Apesar de seus benefícios, Putrama e Martinek (2024) advertem que a construção de ontologias a partir do zero, especialmente de forma manual, é uma tarefa cara, propensa a erros e que exige um alto nível de conhecimento especializado sobre o domínio. Por outro lado, Sadeghi *et al.* (2024) afirmam que a tentativa de contornar essa complexidade semântica força a adoção de uma tecnologia única ou de um modelo centralizado. Os autores reiteram que as abordagens centralizadoras são impraticáveis em grande escala, ao exigirem que os sistemas e partes interessadas adotem o mesmo padrão, tornando-se inviável em ecossistemas de mercado competitivos, dinâmicos e heterogêneos.

¹ Conceitos abordados na Seção 2.2

Para solucionar o impasse do alto custo de desenvolvimento e da inviabilidade de adoção de um modelo centralizado, a alternativa técnica é fomentar o reuso de modelos conceituais e ontologias já existentes. Um exemplo prático da eficácia do reuso de ontologias é apresentado por Porter *et al.* (2025) ao lidarem com a heterogeneidade de metadados em conjuntos de dados ambientais. A automação do processamento de dados é frequentemente prejudicada pela grande variação arbitrária nas representações das unidades de medida, como, por exemplo, diferentes grafias para "gramas por metro quadrado". Em vez de desenvolverem um novo sistema do zero, os pesquisadores optaram por reutilizar uma ontologia já consolidada, a QUDT (*Quantities, Units, Dimensions and Types*), para anotar e padronizar essas variáveis. A adoção dessa ontologia preexistente permitiu vincular dados heterogêneos a conceitos únicos, garantindo a legibilidade por máquina e a interoperabilidade. Esse caso ilustra que a conexão de dados às ontologias já estabelecidas é uma estratégia eficiente para contornar a irregularidade das informações primárias, fomentando a integração sem a necessidade de impor novos padrões centralizados.

Contudo, como advertem Sales *et al.* (2023), para pesquisadores e sistemas conseguirem reaproveitar esses modelos, eles não podem permanecer isolados ou mal documentados, mas devem ser otimizados para que a descoberta, o acesso, a interoperabilidade e o próprio reuso sejam processados por máquinas com o mínimo de intervenção humana.

Desse modo, Krisnawijaya *et al.* (2025) ressaltam a aplicação dos princípios FAIR, que visam tornar os dados "*Findable*" (Encontráveis), "*Accessible*" (Acessíveis), "*Interoperable*" (Interoperáveis) e "*Reusable*" (Reutilizáveis), surgem como um norte para mitigar esses desafios. Sales *et al.* (2023) caracterizam como tornar os dados FAIR, ou seja, como tornar a aplicação dos dados otimizada para a descoberta, acesso, interoperabilidade e reuso com intervenção humana mínima.

A Encontrabilidade é um dos princípios FAIR, conforme mostrado por Sales *et al.* (2023), e consiste em atribuir identificadores globais, únicos e persistentes a cada dado e seus metadados. Além disso, os dados devem possuir metadados ricos, incluindo o identificador, e devem ser indexados em plataformas de busca.

Ainda segundo Sales *et al.* (2023), outro princípio é Acessibilidade, conceituado por exigir o uso de protocolos de comunicação padronizados, como o HTTP (*Hypertext Transfer Protocol*), além de condições de acesso claras para os dados. Já a Interoperabilidade, por sua vez, é aprimorada publicando os dados em formatos comuns de representação do conhecimento, incluindo referências qualificadas a outros dados.

Por fim, a Reutilização é garantida pela descrição dos recursos com atributos relevantes, com informações claras de licenciamento, dados de proveniência e a adesão a padrões comuns relevantes (Sales *et al.*, 2023).

Nesse contexto, para Ulrich *et al.* (2022), os princípios FAIR são vistos como diretrizes para a gestão de dados e para a associação desses dados com metadados. Porém, embora os metadados sejam um pilar dos princípios FAIR, Batista *et al.* (2022) argumentam que eles precisam ser ricos e consistentes. Os autores ressaltam que as diretrizes tradicionais para a criação de metadados são, em sua maioria, voltadas para o consumo humano, tornando-as ambíguas para uma validação automatizada. Para superar essa limitação, os metadados precisam ser legíveis e acessíveis por sistemas, permitindo a compreensão de forma consistente e inequívoca, otimizando sua reutilização.

Para alcançar essa compreensão inequívoca pelas máquinas, Vogt *et al.* (2025) citam que o simples uso de ontologias e vocabulários controlados não é suficiente para garantir a adesão aos princípios FAIR. Como a criação de um padrão semântico universal é inviável, a interoperabilidade semântica deve ser tratada sob uma ótica linguística, dividida em duas camadas: a terminológica e a proposicional.

Vogt *et al.* (2025) definem a interoperabilidade terminológica para a criação de mapeamentos de entidades para ensinar à máquina quais termos de vocabulários distintos possuem o mesmo significado. Já a interoperabilidade proposicional é definida para o uso de cruzamentos de esquemas, alinhando as estruturas e as regras dos dados para que os diferentes sistemas consigam interagir logicamente. Dessa forma, supera-se a ambiguidade promovendo um ecossistema de serviços FAIR capaz de integrar modelos de metadados modulares e reutilizáveis.

A implementação dos princípios FAIR não é um processo simples, especialmente em domínios ambientais, como a Agrometeorologia. Para sanar esse problema, Devaraju e Huber (2021) propuseram a Ferramenta *F-UJI*, objetivando traduzir os princípios FAIR em métricas quantificáveis e testes práticos para medir e avaliar programaticamente o grau de conformidade de um conjunto de dados com os princípios FAIR, uma métrica referida como *FAIRness*.

A Ferramenta *F-UJI* propõe uma harmonização tecnológica combinada com as infraestruturas de pesquisa, com o intuito de superar as ineficiências de abordagens manuais e tornar os dados FAIR para análise automática por ferramentas computacionais. Para viabilizar essa automação, a *F-UJI* funciona como uma Interface de Programação de Aplicações (API – do inglês, *Application Programming Interface*) fundamentada na arquitetura *REST* (*Representational State Transfer*, ou Transferência de Estado Representacional). Na prática,

essa estrutura atua como uma ponte de comunicação padronizada na Web, a partir de um identificador, extraíndo metadados e verificando se incluem os principais elementos, informações de licença e proveniência, além de analisar o conteúdo dos dados para verificar se os metadados fornecidos correspondem aos arquivos reais (Devaraju; Huber, 2021).

No entanto, apesar de suas contribuições, a ferramenta apresenta algumas limitações. Devaraju e Huber (2021) destacam sua avaliação focada em métricas, não abordando critérios FAIR específicos de domínio, como as especificidades dos dados agrometeorológicos. Além disso, a ferramenta não avalia explicitamente aspectos de qualidade e usabilidade dos dados para sua reutilização.

Petrosyan *et al.* (2023) evidenciam essas limitações ao utilizar a Ferramenta *F-UJI* para avaliar o grau de conformidade com os princípios FAIR em um grande conjunto de dados agrícolas. Os resultados indicaram a baixa pontuação de conformidade com o princípio da Reutilização, sugerindo que, embora os metadados básicos possam estar presentes, os aspectos específicos como licenças claras, documentação detalhada e formatos recomendados ainda são insuficientes, com nenhum dos repositórios atingindo 50% de conformidade.

As pesquisas analisadas evidenciam que, o alcance da interoperabilidade por meio da aplicação prática dos princípios FAIR, é um processo complexo e multifacetado. Embora os repositórios consigam garantir a presença de metadados básicos para a Encontrabilidade, avaliações automatizadas em larga escala, como as conduzidas por Devaraju e Huber (2021) em repositórios multidisciplinares e corroboradas por Petrosyan *et al.* (2023) no setor agrícola, demonstram que o princípio da Reutilização é o maior gargalo tecnológico.

O setor agroalimentar e ambiental caracteriza-se como domínios complexos e heterogêneos (Urdu *et al.*, 2024; Zeginis *et al.*, 2024). A presença de descrições genéricas não é suficiente para a integração de sistemas, refletindo diretamente num desafio para a Agrometeorologia. Faz-se necessário, então, um aprofundamento na camada semântica, de modo a fomentar o reuso de ontologias, o mapeamento de entidades e o alinhamento de esquemas, para que as ferramentas computacionais possam acessar, compreender e reutilizar esses dados de forma automatizada, inequívoca e livre de ambiguidades.

2.2 TECNOLOGIAS E ABORDAGENS PARA INTEGRAÇÃO SEMÂNTICA

Com a evolução tecnológica, a compreensão da integração semântica exige a análise tanto dos seus fundamentos teóricos quanto de sua aplicabilidade em cenários reais. Para isso, esta seção está dividida em duas partes principais. A Seção 2.2.1 detalha a arquitetura e os

padrões da Web Semântica, abrangendo desde o modelo de dados até as linguagens de ontologia e consulta, visando fornecer as camadas necessárias para o raciocínio lógico e a representação formal do conhecimento. Em seguida, a Seção 2.2.2 apresenta as aplicações práticas de integração semântica, ilustrando como essa infraestrutura teórica tem sido implementada para superar as heterogeneidades estruturais e semânticas em experimentos e integrações de sistemas no mundo real.

2.2.1 Arquitetura e Padrões da Web Semântica

Os desafios de interoperabilidade de dados não se restringem à simples agregação de conjuntos de dados, mas demandam, também, a migração para a Web Semântica, dada a evolução da Web como uma plataforma de comunicação e informação. Conforme Zárate *et al.* (2020), a Web Semântica tem como objetivo a representação da informação para ser usada e reutilizada autonomamente por máquinas para integração, possibilitando aos sistemas de informação encontrar dados e compreender o significado e as relações entre diferentes fontes.

A fim de solucionar o problema da heterogeneidade e permitir a verdadeira integração da informação, a Web Semântica apoia-se no conceito de *Linked Data* (Dados Conectados). De acordo com Isotani e Bittencourt (2015), o *Linked Data* atua como um conjunto de boas práticas e padrões tecnológicos internacionais recomendados pelo W3C (*World Wide Web Consortium*) para a publicação e a conexão de dados estruturados na Web. É esse alinhamento metodológico que garante o conhecimento acessível, interligado e livremente reutilizado de forma automática por qualquer agente de software ou sistema computacional.

Para que os princípios de dados conectados sejam materializados e o conhecimento ganhe significado interpretável pelas máquinas, Meckler *et al.* (2023) ressaltam a necessidade dos dados ser enriquecidos semanticamente por meio de ontologias e vocabulários compartilhados.

Segundo Zhang *et al.* (2024), ontologia corresponde a uma especificação formal e clara de um modelo conceitual compartilhado. Na área da ciência da computação, a construção dessa ontologia ocorre por meio da captura do conhecimento de um domínio específico, estruturado de forma lógica e matemática. Complementando essa visão estrutural, Tayur e Suchithra (2022) ressaltam que as ontologias são arquitetadas para modelar um domínio de conhecimento particular, definindo formalmente os conceitos (classes), as descrições (propriedades) desses conceitos e os relacionamentos entre eles.

Para os ambientes computacionais processarem essa estrutura, especialmente durante o alinhamento de dados, os componentes ontológicos são convertidos em grafos matemáticos, nos quais os conceitos assumem o papel de vértices e as relações o papel de *links* de conexão, como destacado por Mani e Annadurai (2022). O Quadro 1 sintetiza os elementos de uma ontologia e como são representados computacionalmente.

Quadro 1 – Elementos fundamentais e representação estrutural de uma ontologia

Elemento Ontológico	Descrição Lógica e Conceitual (Zhang <i>et al.</i> , 2024)	Representação Computacional / Estrutural (Mani; Annadurai, 2022)
Classes (ou Conceitos)	Representam o significado de uma entidade, definindo os grupos e conceitos existentes no mundo real.	São extraídas do modelo e representadas como os "nós" (ou vértices) dentro de um grafo.
Relações (ou Propriedades)	Elucidam as características específicas e as interações existentes entre as entidades do domínio.	São representadas como as "arestas" (ou <i>links</i>) que conectam os diferentes nós do grafo.
Instâncias	Correspondem aos dados concretos e específicos que materializam as classes.	São extraídas como os dados individuais atrelados aos nós da ontologia.
Funções	Consistem em tipos especiais de relações lógicas que indicam direcionamentos entre os conceitos.	Representadas pelo direcionamento e orientação (grau de entrada e saída) das arestas no grafo.
Axiomas	Representam os fatos, impõem regras lógicas rigorosas e estabelecem restrições de limite ao modelo.	Atuam como regras de consistência que auxiliam no raciocínio e validação final dos mapeamentos.

Fonte: O autor

Nota: Elaborado pelo autor com base em Zhang *et al.* (2024) e Mani e Annadurai (2022)

Conforme ilustrado no Quadro 1, a integração entre a fundamentação teórica e a representação computacional viabiliza o processamento semântico em ambientes computacionais. Enquanto a estrutura lógica, descrita por Zhang *et al.* (2024), garante que o significado, as regras e as restrições do mundo real sejam preservadas no modelo, a sua conversão para grafos, abordada por Mani e Annadurai (2022), permite aos algoritmos navegarem por essa rede de informações, numa arquitetura baseada em nós e *links*, possibilitando aos sistemas computacionais calcular matematicamente a similaridade estrutural entre diferentes bases de dados.

Dessa forma, ao fornecer essas descrições de conhecimento de forma explícita, compartilhável e reutilizável, as ontologias atuam como uma solução para a heterogeneidade dos dados, permitindo a experimentação semântica e a inferência de novas informações com base em regras lógicas (Abhilash; Mahesh, 2023).

Para alcançar o nível de processamento e inferência mencionado, a complexidade da construção da ontologia exerce um papel determinante. Inserida no contexto de evolução da Web Semântica, McGuinness (2003) propõe um espectro, categorizando as ontologias com base em seu uso e expressividade estrutural. Em uma extremidade encontram-se as ontologias simples, envolvendo modelos rudimentares, como vocabulários controlados e taxonomias básicas. Embora sejam úteis para a organização da informação e permitam uma interoperabilidade inicial entre sistemas, essas abordagens são insuficientes para raciocínios complexos.

Por outro lado, para viabilizar as funções avançadas dos ambientes computacionais, exige-se a adoção de ontologias estruturadas, conforme McGuinness (2003). Este tipo de ontologia possui uma formalização rigorosa, fazendo uso de relações complexas, restrições de valor e axiomas lógicos, responsáveis por permitir a verificação automática de consistência dos dados, a integração semântica profunda e a inferência automatizada.

Para essa formalização rigorosa ser efetivamente processada e compartilhada no ambiente da Web Semântica, as ontologias são descritas por meio de linguagens e vocabulários padronizados (Meckler *et al.*, 2023). Os autores citam o RDF (*Resource Description Framework*) como o modelo de dados universal para representar informações semanticamente enriquecidas, garantindo o significado dos dados legível em ambientes computacionais.

De acordo com a documentação do consórcio W3C, o RDF define um modelo abstrato, independentemente de qualquer forma de serialização ou formato específico de armazenamento. Essa separação entre o modelo lógico e a codificação garante interoperabilidade e flexibilidade, permitindo sua representação em diferentes sintaxes textuais, como *Turtle*, *N-Triples*, *JSON-LD (JavaScript Object Notation)* e *RDF/XML (eXtensible Markup Language)*, sem que o significado original da informação seja alterado. Tal característica reforça a noção do RDF atuar como um modelo semântico, e não apenas sintático, voltado à representação do conhecimento de forma compreensível tanto para humanos quanto para máquinas (W3C, 2026a).

Para viabilizar essa comunicação, a infraestrutura do RDF é baseada na formulação de triplas lógicas, nas quais cada afirmação no modelo consiste na junção de um sujeito, um predicado (propriedade) e um objeto (o valor da propriedade), atuando como o alicerce sintático necessário para descrever recursos e mapear as suas inter-relações na Web Semântica de forma consistente (Fensel *et al.*, 2003).

O W3C (2026a) define que o conjunto dessas triplas constitui com um grafo, podendo ser visualizado como um diagrama de nós e ligações direcionadas. O Quadro 2 sintetiza os

quatro tipos principais de termos formadores desses grafos, detalhando suas funções estruturais e exemplos práticos.

Quadro 2 – Tipos de termos que compõem um grafo em RDF

Tipo de Termo	Descrição Estrutural	Exemplo Prático
IRIs (<i>Internationalized Resource Identifiers</i>)	São os " <i>links</i> " ou "CPFs" dos recursos. Eles identificam entidades de maneira global e única na Web, garantindo que máquinas diferentes saibam que estão falando exatamente da mesma coisa.	Em vez de usar apenas a palavra "Brasil" (que pode ser ambígua), o sistema usa o endereço único http://dbpedia.org/resource/Brasil para representar inequivocamente o país.
Literais	São os valores concretos da informação. Enquanto os IRIs são os endereços, os literais são os dados em si: textos, números, datas ou valores booleanos.	O nome de uma pessoa expresso como o texto "João Silva"; uma idade expressa numericamente como "35"^^xsd:integer; ou a palavra "gato"@pt indicando especificamente o idioma português.
Nós em Branco (<i>Blank Nodes</i>)	São "nós anônimos". Eles indicam que algo existe, mas não possuem um identificador global (IRI). Servem para agrupar informações dependentes que não precisam de um endereço próprio na Web.	Para dizer que uma pessoa tem um endereço, mas o endereço não tem um " <i>link</i> " próprio, o sistema cria um nó anônimo temporário (ex: <code>_:endereço1</code>) apenas para agrupar os dados "Rua X" e "Cidade Y".
Termos de Tripla (<i>Triple Terms</i>)	É a capacidade de "embutir" uma tripla inteira dentro de outra. Isso permite que a máquina faça afirmações sobre outras afirmações (um processo chamado de reificação).	Em vez de apenas registrar o fato direto (A Terra, gira em torno do, Sol), o sistema embute isso como o objeto de outra frase para modelar: Galileu afirmou que (A Terra, gira em torno do, Sol).

Fonte: O autor

Nota: Elaborado pelo autor com exemplos próprios baseados nos conceitos de W3C (2026a).

Como se observa no Quadro 2, a complementaridade desses quatro elementos confere ao modelo RDF a sua expressividade estrutural. Os IRIs garantem não haver ambiguidade na identificação universal dos recursos, já os literais assumem o papel de armazenar os valores concretos da informação.

O W3C (2026a) ressalta o IRI surgindo para tratar com as limitações de caracteres do URI (*Uniform Resource Identifier*). Enquanto o URI restringe a identificação de recursos ao uso do alfabeto latino básico (caracteres ASCII), o IRI é um superconjunto do URI, permitindo o uso de caracteres do *Unicode*, como letras com acentos, caracteres de outros alfabetos, tais como o cirílico, árabe, chinês, entre outros. URIs e IRIs cumprem a mesma função de identificação global e única dos recursos, base para a criação dos *links* de dados e inter-relações na estrutura de triplas RDF.

Em paralelo, os nós em branco conferem a flexibilidade necessária para agrupar dados localmente, e os termos de tripla elevam a capacidade do modelo ao permitir a máquina

processar afirmações sobre outras afirmações. Conjuntamente, essa estrutura não apenas descreve abstrações de modo padronizado, mas atua como um alicerce amplamente extensível.

A eficácia dessa arquitetura estrutural em lidar com a heterogeneidade é destacada por Avelino, Cordeiro e Cavalcanti (2020), que utilizam a flexibilidade dos grafos em RDF para interligar dados em diferentes níveis de abstração. O RDF não impõe uma distinção rígida entre dados e metadados, permitindo categorias lógicas (esquemas) e dados concretos (instâncias) coexistirem no mesmo nível do grafo, ou seja, o descritor de uma fonte de dados (metadados) pode atuar como dado em outra fonte. Assim, ao utilizar as triplas para descrever o conhecimento, o RDF torna-se uma abordagem robusta para estabelecer interligações interníveis (entre metadados e dados) e intraníveis (entre metadados ou entre dados), promovendo um alto grau de interoperabilidade em cenários de dados heterogêneos.

No experimento conduzido por Avelino, Cordeiro e Cavalcanti (2020) no âmbito da engenharia de software, o modelo RDF permitiu, por exemplo, conectar diretamente o registro de um software listado como instância em um sistema de catálogo à categoria estrutural (esquema) desse mesmo software em sua ferramenta de origem, viabilizando consultas unificadas sobre a taxa de defeitos e testes.

Além de sua flexibilidade estrutural, conforme o W3C (2026a), a semântica formal do RDF estabelece as condições de interpretação dos grafos, definindo diferentes regimes de inferência, chamados de *entailments*. O *simple entailment* constitui o regime básico, no qual um grafo em RDF G1 implica outro grafo G2 se toda interpretação satisfaz G1 também satisfaz G2 (W3C, 2026a).

Por sua vez, o W3C (2026a) cita o *RDFS entailment*, que define as regras de inferência para o *RDF Schema* (RDFS), uma extensão fornecendo mecanismos para descrever classes, propriedades, hierarquias e restrições de domínio e alcance, possibilitando que novas informações possam ser inferidas a partir de declarações explícitas.

Outra regra de inferência definida pelo W3C (2026a) é o *D-entailment*, ou *Datatype entailment*, entendido como um sistema de inferência que estende o RDF para considerar tipos de dados literais, como inteiros, *strings* e datas, e suas interpretações. O *D-entailment* permite valores literais ser comparados e interpretados de acordo com os tipos de dados definidos, garantindo consistência semântica em operações lógicas e consultas. Assim, enquanto o *RDFS entailment* lida com as relações estruturais e hierárquicas entre recursos, o *D-entailment* amplia a semântica para lidar com a coerência e a inferência sobre dados tipados, integrando a lógica de valores ao raciocínio semântico. Contudo, a própria documentação do consórcio W3C (2026a) ressalta que os modelos RDF e RDFS reconhecem apenas os casos básicos de

inferência. Para contornar essa limitação estrutural em cenários de complexidade, a Web Semântica exige a adoção de linguagens, adicionando regimes de *entailment* mais expressivos e poderosos ao topo dessa infraestrutura de grafos.

Neste cenário, Zhang e Lobov (2024) destacam a linguagem OWL (*Web Ontology Language*), recomendada pelo consórcio W3C como a principal tecnologia para a descrição de classes abstratas, relações e instâncias. De acordo com a especificação oficial do W3C (2012), a OWL é uma linguagem ontológica declarativa destinada à representação explícita de conhecimento sobre um domínio específico. Enquanto o RDF atua como o alicerce de estruturação, a principal função da OWL é aumentar a expressividade semântica desse modelo, fornecendo axiomas e entidades, podendo ser classes, propriedades de objeto e propriedades de dados. Esses axiomas estabelecem, conforme o W3C (2012), as relações entre entidades e permitem descrever formalmente conceitos como subclasse, equivalência de classes, disjunção, restrições de valor e cardinalidade, armazenados e trafegados primariamente como documentos RDF.

O W3C (2012) documenta que a OWL, construída sobre o RDF e sua extensão RDFS, permite descrever características complexas, às quais o RDF não dá suporte, como relações de equivalência de classes, disjunções e restrições rigorosas de cardinalidade e valor. Fensel *et al.* (2003) explicam que essa expressividade possibilita aos sistemas realizar inferências lógicas, descobrindo novos conhecimentos que não estão explicitamente declarados nos dados brutos.

Além dessa estrutura declarativa, o W3C (2012) define uma semântica formal, a *RDF-Based Semantics*, que assegura que toda inferência realizada sobre uma ontologia seja mantida a estrita coerência lógica do modelo. Essa semântica fundamenta os mecanismos de inferência implementados por programas que processam ontologias OWL, os chamados *reasoners*.

Fensel *et al.* (2003) destacam que esses motores de inferência permitem avaliar tanto o esquema quanto as instâncias. Conforme a Lógica de Descrição que embasa a fundamentação dessas tecnologias (Fensel *et al.*, 2003; W3C, 2012), essa separação divide a base de conhecimento em dois componentes formais: a *TBox (Terminological Box)*, armazenando o esquema de classes e as regras lógicas; e a *ABox (Assertional Box)*, armazenando os fatos e instâncias concretas do mundo real. É a avaliação conjunta dessas duas camadas que viabiliza processos de raciocínio como a identificação automática de hierarquias de classes (classificação), a verificação de contradições lógicas dentro da ontologia (consistência) e a determinação de a qual classe um indivíduo pertence (realização).

Outro ponto relevante destacado pelo W3C (2012) é a total compatibilidade estrutural entre a OWL e o RDF. Assim, as ontologias OWL podem ser integralmente serializadas em

formato RDF, garantindo a interoperabilidade com outras ferramentas da Web Semântica. Na prática, o W3C (2012) ressalta, ainda, que as condições semânticas estabelecem cada axioma ou expressão em OWL possui uma tradução equivalente em triplas RDF, mantendo o mesmo significado. Como consequência, essa característica consolida a OWL como o nível superior da camada semântica da Web, pois combina a formalidade rigorosa da lógica descritiva com a flexibilidade e a interconectividade inerentes aos grafos em RDF.

Para os usuários e os agentes computacionais interagir com a base de conhecimento enriquecida pelas regras da linguagem OWL, a Web Semântica adota o SPARQL (*SPARQL Protocol and RDF Query Language*) como sua linguagem de consulta e atualização padrão. Como apontam Zhang e Lobov (2024), dentro dessa arquitetura, o SPARQL atua como a interface unificada para recuperar e manipular as informações modeladas nos grafos em RDF. Segundo El Idrissi *et al.* (2022), a estrutura de uma consulta SPARQL é baseada nos *Basic Graph Patterns* (BGP), um conjunto de padrões de triplas. O processamento dessas consultas ocorre por meio da correspondência de subgrafos (*subgraph matching*), na qual o motor de busca avalia como as variáveis do BGP mapeiam-se aos dados do grafo subjacente.

A partir desse mecanismo de correspondência, o W3C (2026b; 2026c) documenta que a linguagem oferece diferentes tipos de consultas, incluindo SELECT, usado para a projeção tabular de resultados; CONSTRUCT, utilizado para a criação de novas triplas e grafos em RDF; DESCRIBE, focado no detalhamento de recursos estruturais; e ASK, voltado à verificação booleana da existência de correspondências. Além da extração de dados, El Idrissi *et al.* (2022) citam que as especificações mais recentes do protocolo permitem a manipulação direta do grafo, com operações de inserção e exclusão de triplas, como INSERT DATA e DELETE DATA, consolidando a gestão dinâmica da base de conhecimento.

A viabilidade prática dessa camada de consulta e atualização é evidenciada no experimento de Avelino, Cordeiro e Cavalcanti (2020), no qual os autores utilizaram requisições em SPARQL sobre um grafo unificado para recuperar, em segundos, métricas transversais de defeitos e testes de engenharia de software, que originalmente residiam em sistemas estruturalmente isolados.

A construção, a manipulação e a consulta dessas estruturas semânticas complexas são suportadas por diversas ferramentas e *frameworks* tecnológicos de código aberto. O *Protégé*, por exemplo, é amplamente reconhecido pela comunidade internacional como um editor de ontologias, facilitando a construção, a visualização e a manutenção de hierarquias de classes e propriedades lógicas (Isotani e Bittencourt, 2015). Já para o desenvolvimento de aplicações computacionais voltadas ao processamento de grafos, os autores destacam bibliotecas e

frameworks de programação dotados de APIs, como o *Apache Jena*. A utilidade prática dessas ferramentas é evidenciada por Tayur e Suchithra (2022), com emprego da biblioteca *Apache Jena* na etapa de pré-processamento de ontologias em formato OWL, permitindo a extração automatizada de características estruturais e propriedades para viabilizar o mapeamento e o alinhamento semântico entre diferentes modelos.

O êxito na integração e na recuperação de informações em cenários como esse, ilustra a Web Semântica operando por meio de uma progressão lógica de camadas tecnológicas complementares. Conforme detalhado por Isotani e Bittencourt (2015), essa arquitetura consolida-se a partir de uma base estruturada no modelo RDF, sobre a qual a linguagem OWL adiciona a camada lógica e de representação ontológica, permitindo, no nível de acesso, a recuperação de informações por meio do SPARQL como linguagem de consulta padrão. É a integração dessa infraestrutura tecnológica, como ressaltam Meckler *et al.* (2023), que materializa os princípios do *Linked Data* abordados inicialmente, viabilizando a interoperabilidade semântica e a verdadeira integração da informação.

2.2.2 Aplicações Práticas de Integração Semântica

A transição da teoria arquitetônica para a implementação prática revela que a simples troca ou publicação de conjuntos de dados não é suficiente para garantir a verdadeira integração e o alcance das diretrizes de dados abertos e FAIR. Conforme argumentam Thalath, Nagamori e Sakaguchi (2025), o compartilhamento de informações perde a sua capacidade se a semântica e as estruturas inerentes aos dados permanecerem ocultas ou não padronizadas, como em formato de tabelas e arquivos CSV (*Comma-Separated Values*).

Para superar essa barreira, os autores propuseram o desenvolvimento de mecanismos de mapeamento e perfilamento, usando a adoção de Perfis de Aplicação de Metadados (*Metadata Application Profiles*), materializada por meio de ferramentas de código aberto como o YAMAML (*YAMA Mapping Language*). Eles destacam que o funcionamento do YAMAML é baseado em regras de conversão escritas em texto simples na linguagem YAML (*YAML Ain't Markup Language*). O usuário cria um perfil básico indicando como as colunas de uma planilha comum se relacionam com as classes e propriedades desejadas. Ao ler o arquivo original, a ferramenta aplica essas regras e converte automaticamente os dados tabulares em um grafo semântico no formato RDF. Para simplificar a adoção, o processo é executado tanto por comandos de texto quanto por uma interface gráfica na web, permitindo aos usuários sem experiência em programação realizarem a integração (Thalath; Nagamori; Sakaguchi, 2025).

Apesar de viabilizar a publicação de dados interoperáveis, a abordagem de Thalath, Nagamori e Sakaguchi (2025) apresenta limitações estruturais e operacionais, devendo ser consideradas em processos de integração. Como os próprios autores ressaltam, por serem concebidas primariamente como provas de conceito focadas na simplicidade, as ferramentas YAMA e YAMAML enfrentam restrições técnicas ao processar perfis complexos ou estruturas de dados profundamente aninhadas.

Além disso, Thalath, Nagamori e Sakaguchi (2025) mostram que a solução tem uma forte dependência do modelo estrutural RDF e que apresentou desafios de escalabilidade no consumo de memória ao mapear volumes massivos de dados. O ponto mais crítico reportado pelos autores é a ausência de um mecanismo nativo para a reconciliação de *links* de entidades, forçando a avaliação das bases sem a devida interconexão exigida pelo *Linked Data*. A exigência de conhecimento prévio na linguagem de programação YAML impõe uma barreira considerável de adoção para usuários não técnicos.

Como alternativa à conversão física de dados legados, outras abordagens práticas têm focado na integração de grandes volumes de informações heterogêneas por meio de consultas virtuais, contornando os custos de armazenamento e a redundância da materialização de dados. Nesse cenário, Ranatunga *et al.* (2025) propõem um *framework* prático voltado para o domínio ambiental e geoespacial baseado no paradigma de Acesso a Dados Baseado em Ontologias (OBDA - *Ontology-Based Data Access*), atuando na formulação de um Grafo de Conhecimento Virtual (VKG). A abordagem visa superar a fragmentação das informações, garantindo que fontes dinâmicas, como sensores meteorológicos em tempo real, e estáticas, como mapas e topologias, sejam alinhadas sob uma mesma ontologia.

Na prática, Ranatunga *et al.* (2025) ressaltam que a arquitetura do *framework* utiliza a ferramenta de código aberto *Ontop* para conectar um banco de dados relacional georreferenciado a uma ontologia estendida a partir do vocabulário padrão GeoSPARQL. Em vez de transformar e salvar os dados originais no modelo estrutural RDF, o sistema cria mapeamentos declarativos entre a ontologia e o banco de dados. Ao utilizar uma interface web, o usuário executa consultas semânticas em SPARQL, as quais são traduzidas dinamicamente pelo sistema para comandos SQL (*Structured Query Language*) nativos e executadas diretamente nas bases originais. Como resultado, os autores mostram o sistema retornando as informações tabulares processando junções espaciais e visualizando graficamente o relacionamento entre os polígonos e os pontos de interesse em mapas integrados.

A despeito dos benefícios de escalabilidade propiciados pela não materialização dos grafos, a integração via OBDA implementada por Ranatunga *et al.* (2025) apresenta desafios

técnicos no mundo real, como gargalos significativos de desempenho e atrasos na renderização ao submeterem o sistema a consultas espaciais complexas, envolvendo uma densidade de pontos e polígonos. A infraestrutura, conforme os autores, também esbarra em limitações de interoperabilidade com formatos não padronizados ou específicos, como arquivos *raster* massivos, *NetCDF* (*Network Common Data Form*) ou *HDF5* (*Hierarchical Data Format*), exigindo frequentemente o pré-processamento manual dos dados com scripts e ferramentas externas antes de sua inserção no banco relacional. Além disso, Ranatunga *et al.* (2025) destacam a complexidade de configuração inicial de toda a pilha tecnológica OBDA, aliada à necessidade de atualizações manuais de certos repositórios isolados, que impõem consideráveis barreiras de adoção para os gestores e usuários sem profundo domínio da engenharia de dados.

Para além do mapeamento estrutural de arquivos e bancos de dados, um desafio significativo na integração semântica diz respeito à rastreabilidade e à anotação de objetos do mundo real, como sensores em ambientes de Internet das Coisas (IoT). No campo da agricultura, uma aplicação direta da unificação de dados e metadados é a chamada Marcação Semântica (*Semantic Tagging*). Neveu, David e Jonquet (2020) destacam que a digitalização do setor exige que os dados sejam marcados com URIs e o significado dessa informação seja formalizado por ontologias. Na prática, a Marcação Semântica utiliza ontologias para criar anotações, ou seja, *links* de um termo da ontologia para um elemento de dado, permitindo aos sistemas a indexação, busca e raciocínio sobre os dados de forma clara. Como exemplo, uma leitura de sensor pode ser marcada com o conceito “Temperatura do Solo” em uma ontologia.

Contudo, realizar essa anotação de forma manual é um processo oneroso e pouco escalável frente ao volume massivo de informações do *Big Data* agrícola. Para automatizar esse processo, Neveu, David e Jonquet (2020) destacam a eficácia de iniciativas como o *AgroPortal Annotator*. Esse serviço web consome as ontologias e os vocabulários contidos no AgroPortal, um repositório centralizado voltado especificamente para as ciências agrícolas, e analisa descrições em texto bruto, inserindo *tags* semânticas baseadas em conceitos de ontologias automaticamente. Dessa forma, os autores ressaltam a transformação de dados brutos em informações coerentes e interligadas pela ferramenta, evidenciando como o modelo RDF unifica os dados e metadados, enquanto as ontologias fornecem o contexto formal e computacional para apoiar a tomada de decisões de forma integrada.

Apesar do avanço proporcionado pela automatização, a implementação de anotações semânticas no mundo real enfrenta barreiras consideráveis de usabilidade e infraestrutura. Neveu, David e Jonquet (2020) advertem que o ecossistema de ontologias vasto, frequentemente sobreposto e sofre mudanças constantes, dificulta a compreensão de suas

estruturas por usuários finais e pesquisadores sem domínio técnico avançado. Além disso, citam a anotação contínua sendo frequentemente vista pelos produtores de dados como uma tarefa exaustiva e sem recompensa imediata.

O estudo levanta uma crítica quanto à infraestrutura arquitetônica exigida, uma vez que mesmo que a identificação global via URIs seja o padrão tecnológico, garantir a persistência desses *links* a longo prazo é um desafio prático, visto que repositórios mudam de domínio, esquemas de nomenclatura são alterados, e instituições se reestruturam, ameaçando a estabilidade e o acesso futuro à base de conhecimento.

Além da heterogeneidade de formatos e da rastreabilidade física, a integração semântica em sistemas de longo prazo enfrenta o desafio da evolução temporal do conhecimento. No campo das geociências, Ma, Ma e Wang (2020) propuseram uma estrutura para controle de versão da Escala de Tempo Geológico (*International Chronostratigraphic Chart – ISC*), onde essa escala é revisada frequentemente para redefinir os limites temporais de determinados períodos geológicos, exigindo um controle robusto para manter a consistência semântica dos dados ao longo do tempo.

Para solucionar esse problema, em vez de criar um vocabulário ou servidor separado para cada versão do documento, a arquitetura desenvolvida por Ma, Ma e Wang (2020) utiliza URIs como marcadores de versão, ou seja, o sistema liga o atributo específico, como a idade numérica de um limite geológico, diretamente ao URI da sua versão correspondente, unificando todo o histórico de versões da ISC, abrangendo os anos de 2004 a 2018 em um único Grafo de Conhecimento. Essa abordagem permite o histórico completo da evolução de um conceito ser recuperado em uma única consulta na linguagem SPARQL, comprovando a eficiência do modelo estrutural em grafos para rastrear a proveniência do conhecimento sem a necessidade de consultar múltiplos servidores fragmentados.

Apesar de unificar o histórico do conhecimento e otimizar a infraestrutura, a abordagem estrutural de Ma, Ma e Wang (2020) esbarra em desafios operacionais no mundo real, como o fato das práticas de pesquisa gerarem situações de proveniência de dados muito complexas, como amostras antigas que ganham novos atributos, ou a inserção de novos pesquisadores em testes derivados, tornando o rastreamento contínuo um problema ainda em aberto. Além disso, a execução de buscas temporais para cruzar o histórico das escalas exige profunda fluência na formulação de códigos complexos em SPARQL, impondo uma barreira técnica complexa para os geocientistas e especialistas de domínio sem formação em ciência da computação. Por conseguinte, o estudo evidencia a adoção em larga escala de grafos evolutivos ainda

dependendo estritamente do desenvolvimento de interfaces gráficas (GUIs), ocultando a complexidade das consultas para os usuários finais.

Além dos desafios de conversão de dados legados, a integração na Web Semântica enfrenta uma barreira de mais alto nível, a interoperabilidade entre as próprias ontologias. Para Prudhomme *et al.* (2025) a proliferação de modelos estruturais em um mesmo domínio pode gerar "silos de ontologias", um cenário onde os dados organizados sob ontologias independentes permanecem semanticamente isolados e incompatíveis. Para mitigar esse problema e aproximar os dados do nível máximo de interoperabilidade, os autores propuseram uma metodologia rigorosa de Mapeamento de Ontologias (*Ontology Mapping*), focada em construir um alinhamento semântico entre a Ontologia de Proveniência (PROV-O) e a Ontologia Formal Básica (BFO).

Na prática, a arquitetura de alinhamento proposta por Prudhomme *et al.* (2025) ultrapassa o simples cruzamento de vocabulários, baseando-se em regras lógicas de equivalência e subsunção, relação de hierarquia “é-um” entre classes, para interpretar uma teoria dentro da outra. Utilizando tecnologias recomendadas pelo W3C, como as linguagens OWL e SPARQL, aliadas a regras SWRL (*Semantic Web Rule Language*) e axiomas de cadeias de propriedades, o sistema estabelece uma ponte de tradução formal. Para garantir a precisão estrutural, os autores utilizaram ferramentas de raciocínio automatizado, os *reasoners*, como o HermiT, executadas em processos contínuos para avaliar as regras de alinhamento contra instâncias reais de dados. Como resultado, Prudhomme *et al.* (2025) destacam que essa estrutura permite aos dados de uma ontologia serem processados pela outra, mas também possibilita aos raciocinadores lógicos fazer inferências e descobrir informações implícitas automaticamente.

Apesar do rigor lógico do mapeamento, a abordagem de Prudhomme *et al.* (2025) apresenta limitações significativas de escala e infraestrutura. Por depender da curadoria semiautomatizada e da análise conceitual profunda das condições necessárias e suficientes de cada termo, o processo de alinhamento é exaustivo e oneroso quando comparado a técnicas de correspondência totalmente automatizadas. Adicionalmente, a abordagem esbarrou em limitações expressivas da estrutura base do modelo RDF e da linguagem OWL para o mapeamento de atributos de dados (*data properties*), forçando a equipe a deixar o mapeamento de regras estruturais mais complexas fora do escopo computável.

Prudhomme *et al.* (2025) evidenciaram a aplicação de raciocinadores lógicos rigorosos, revelando que a fusão de ontologias frequentemente expõe inconsistências e erros estruturais ocultos nas próprias bases de dados originais, exigindo revisões e validações complementares.

As aplicações práticas discutidas evidenciam a concretização da Web Semântica e do *Linked Data* além da simples adoção tecnológica. Seja na conversão física de arquivos, na virtualização de bancos de dados, na anotação de sensores, no controle de versões temporais ou no mapeamento lógico de modelos conceituais, o desafio central reside na profunda diversidade e heterogeneidade das informações. Como demonstrado pelos estudos, a proliferação de padrões independentes frequentemente culmina na criação de "silos de ontologias", nos quais dados do mesmo ecossistema permanecem semanticamente isolados.

Esse cenário de fragmentação é especialmente crítico quando se observam os padrões de metadados específicos de domínio, sobretudo na Agrometeorologia. Diante dessa barreira, a próxima seção visa detalhar esses padrões de metadados agrometeorológicos, evidenciando a necessidade de mecanismos e tecnologias que unifiquem essas representações divergentes para viabilizar, de fato, a criação de um grafo de conhecimento coerente e acionável.

2.3 PADRÕES DE METADADOS UTILIZADOS NA AGROMETEOROLOGIA

A integração de dados agrometeorológicos permite, conforme *Parisse et al. (2022)*, analisar de maneira precisa as interações entre as variáveis climáticas e a produção agrícola. *Krisnawijaya et al. (2025)* e *Sigler et al. (2025)* corroboram que a gestão eficiente desses dados impacta diretamente a tomada de decisão, constituindo uma ferramenta para a promoção de práticas agrícolas sustentáveis e precisas.

Sigler et al. (2025) ressaltam, ainda, que a falta de um padrão unificado para a agrometeorologia representa uma barreira significativa para a interoperabilidade e o compartilhamento prático das informações.

No contexto agrícola, ambiental e geoespacial, alguns padrões são amplamente utilizados devido à sua robustez e abrangência para atender aos princípios FAIR. Destacam-se o uso de vocabulários semânticos como o AGROVOC (*Krisnawijaya et al., 2025*), padrões globais de metadados focados em clima e geolocalização, como *WMO Core Metadata Profile* (WCMP) e o ISO 19115 (*Parisse et al., 2022*), e padrões consolidados para o compartilhamento de dados ambientais e de biodiversidade, como o *Darwin Core* (*Wieczorek et al., 2012*).

Desse modo, a integração do ISO 19115, *Darwin Core*, WCMP e do AGROVOC compõe uma abordagem multidisciplinar de metadados, na qual cada padrão resolve um aspecto da interoperabilidade, como a semântica agrícola, o rigor geoespacial-climático e o monitoramento ambiental, garantindo aos sistemas agrometeorológicos como um todo atender aos princípios FAIR. A partir dessa perspectiva, esta seção é dedicada a explorar

detalhadamente cada um desses modelos, evidenciando como eles contribuem individual e coletivamente para a consolidação de um ecossistema de dados interoperável.

2.3.1 Padrão ISO 19115 e o contexto geoespacial

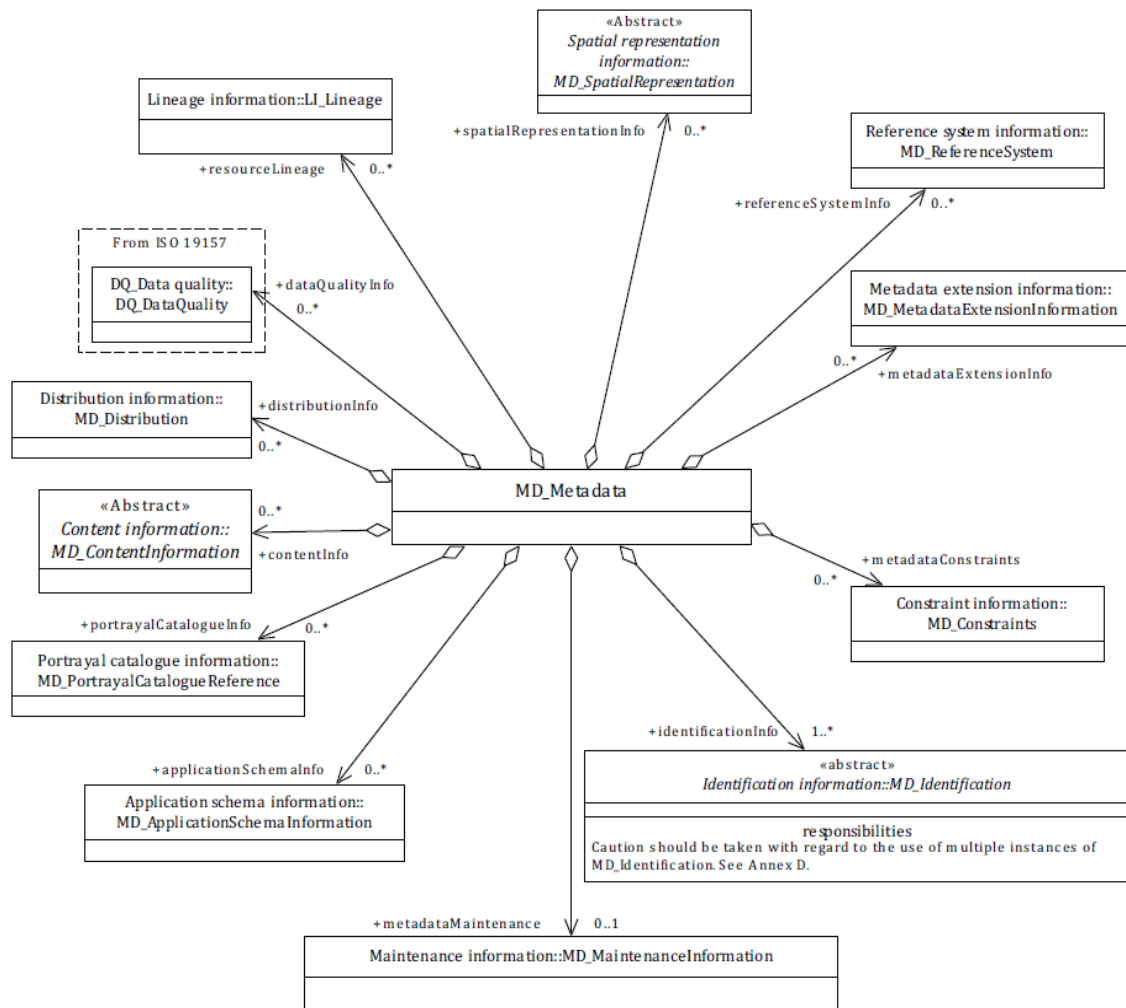
A ISO 19115 é um padrão internacional da *International Organization for Standardization* (ISO) desenvolvido para descrever metadados geoespaciais através de um conjunto de elementos e estruturas para documentar dados geográficos (ISO, 2014). De acordo com a documentação da norma, o objetivo desse padrão é possibilitar aos usuários localizar, avaliar, adquirir e utilizar dados geoespaciais de forma eficiente.

A ISO (2014) define que o padrão é constituído por um esquema de metadados para descrever informações e serviços geográficos, abordando a identificação, a extensão, a qualidade, os aspectos espaciais e temporais, o conteúdo, a referência espacial, a representação visual, a distribuição e outras propriedades de dados e serviços geográficos digitais. Apesar de seu foco primário ser o ambiente digital, os princípios da norma são abrangentes e podem ser estendidos para descrever recursos físicos ou textuais, como mapas impressos e conjuntos de dados não geográficos (ISO, 2014).

Em termos de estrutura, a ISO (2014) concebe esse padrão para ser abrangente e flexível, com aplicabilidade vasta, incluindo a catalogação de todos os tipos de recursos, desde conjuntos e séries de dados até serviços e feições geográficas individuais. A norma é estruturada por meio de um modelo de linguagem UML (*Unified Modeling Language*) para organizar a estrutura dos metadados em 13 pacotes distintos, em que cada pacote possui um diagrama de classe UML próprio, representando uma parte específica da descrição de um recurso, como identificação, restrições e linhagem. Esses diagramas mostram como as classes e elementos de metadados se relacionam.

Para ilustrar essa arquitetura, a Figura 1 demonstra a relação entre a classe de metadado principal, denominada *MD_Metadata*, e suas classes agregadas, a exemplo da representação espacial e das informações de identificação:

Figura 1 – Classes de esquema de metadados



Fonte: ISO, 2014

Como pode ser observado na Figura 1, a arquitetura do padrão tem como núcleo a classe raiz *MD_Metadata*, armazenando as informações sobre o próprio registro de metadados. A partir dela, o esquema completo se desdobra em um agregado de 12 classes adicionais de metadados. A norma estabelece uma estrutura modular para a catalogação, em que pacotes adicionais são incorporados apenas quando houver a necessidade de fornecer metadados suplementares.

Nesse sentido, a Figura 1 ilustra a diferença de cardinalidade entre os pacotes mostrando a classe *MD_Identification*, correspondente a informações de identificação, sendo a única estritamente obrigatória na agregação, indicada pela cardinalidade "1..*", pois fornece as características mínimas exigidas para descobrir e identificar um recurso.

As demais classes operam com cardinalidade "0..*" ou "0..1", como as classes *MD_Constraints* (informações de restrições), *MD_SpatialRepresentation* (representação

espacial), *LI_Lineage* (linhagem) e *MD_Distribution* (distribuição), indicando a implementação como opcional ou condicional às características específicas do dado modelado. Adicionalmente, a representação estrutural evidencia a interoperabilidade do padrão com outras normas da família ISO, como é o caso da classe *DQ_DataQuality* (qualidade de dados), importada de forma integrada da norma ISO 19157.

Além das relações de agregação com os demais pacotes, a classe raiz *MD_Metadata* possui um conjunto de propriedades internas específicas, concebidas para fornecer informações sobre o próprio registro de metadados em si (ISO, 2014). Essa estrutura, referida pela norma como "metadados sobre metadados", atua como a identidade do documento, armazenando características gerenciais para a rastreabilidade e a administração do catálogo. A ISO (2014) mostra que, entre os atributos internos dessa classe, a norma estabelece a obrigatoriedade de identificar a parte responsável pela criação ou manutenção das informações dos metadados (*contact*), assim como a data de criação ou de atualização do registro (*dateInfo*), ambos operando com cardinalidade "1..*".

Adicionalmente, a ISO (2014) ressalta que a classe *MD_Metadata* suporta a inclusão de diversos elementos opcionais ou condicionais, a exemplo de um identificador único para o registro (*metadataIdentifier*), do idioma e da codificação de caracteres adotados no documento (*defaultLocale*), de vínculos hierárquicos com registros superiores (*parentMetadata*), bem como das citações formais referenciando os padrões e os perfis aos quais o arquivo se conforma (*metadataStandard* e *metadataProfile*).

Para complementar os diagramas UML, a norma ISO (2014) fornece um dicionário de dados detalhando cada classe e elemento de metadados. Cada classe possui atributos específicos no dicionário de dados, com o intuito de garantir clareza e consistência. Os atributos incluem:

- a) Nome/Papel: é o rótulo atribuído a uma classe ou a um elemento de metadados, sendo os nomes das classes sempre começam com uma letra maiúscula;
- b) Definição: uma descrição sobre o elemento ou a classe;
- c) Obrigação/Condição: indica se o elemento é obrigatório (M), opcional (O) ou condicional (C), ou seja, exigido apenas se uma condição específica for atendida;
- d) Ocorrência Máxima: o número máximo de vezes de aparição do elemento, representado por "1", quando for uma única ocorrência, e "N", para ocorrências repetidas ou ilimitadas. Caso sejam ocorrências com um número fixo, usa-se o número correspondente, por exemplo, "2" quando for duas ocorrências;
- e) Tipo de Dados: o tipo do valor aceito pelo elemento, como *integer*, *string*, *boolean*, entre outros;

- f) Domínio: conjunto de valores permitidos para um elemento de metadados, como uma lista fechada de opções pré-definidas, intervalo numérico, texto livre e até outra classe de metadados.

Com base nisso, o Quadro 3 foi montado para ilustrar a estrutura do dicionário de dados para elementos de metadados de geoinformação, demonstrando a diversidade de tipos de dados, obrigações e propósitos dos metadados no padrão:

Quadro 3 – Exemplo Representativo do Dicionário de Dados ISO 19115

Nome/Papel	Definição	Obrigaçã /Condição	Ocorrência Máxima	Tipo de Dados	Domínio
<i>MD_Identification.abstract</i>	Breve resumo narrativo do recurso	M	1	<i>Character String</i>	Texto livre
<i>MD_Identification.purpose</i>	Resumo das intenções com que o recurso foi desenvolvi do	O	1	<i>Character String</i>	Texto livre
<i>MD_Identification.status</i>	Status do recurso	O	N	Classe	<i>MD_ProgressCode</i>
<i>MD_Resolution.distance</i>	Distância horizontal da amostra de solo	C – se escala, dist. vertical, dist. angular ou nível de detalhe não documenta dos	1	Classe	Distance
<i>MD_Keywords.keyword</i>	Palavra(s) comument e usada(s) ou formalizad a(s) para descrever o assunto	M	N	<i>Character String</i>	Texto livre
<i>MD_Keywords.type</i>	Assunto usado para agrupar palavras- chave semelhante s	O	1	Classe	<i>MD_KeywordType Code</i>

Fonte: O autor.

O Quadro 3 permite compreender como a ISO 19115 equilibra rigor e flexibilidade na catalogação. Os elementos para a descoberta de um recurso, como as classes

MD_Identification.abstract (resumo) e *MD_Keywords.keyword* (palavras-chave), são classificados como obrigatórios (M), garantindo o nível mínimo de detalhamento exigido pelos princípios FAIR.

Em contrapartida, a norma oferece extensibilidade por meio de elementos opcionais (O), como o propósito e o status da feição, permitindo aos produtores de dados enriquecerem o documento conforme a necessidade. O Quadro 3 também evidencia a robustez do padrão ao lidar com cenários específicos através da condicionalidade (C), demonstrando que a classe *MD_Resolution.distance* (distância da resolução), por exemplo, só se torna obrigatória caso outras métricas de escala ou detalhamento espacial não tenham sido previamente fornecidas.

A restrição dos dados a tipos específicos, como *CharacterString*, e a domínios fechados, como a classe *MD_ProgressCode*, elimina ambiguidades de preenchimento, assegurando aos sistemas computacionais distintos interpretarem os metadados de maneira padronizada e interoperável.

No escopo da descoberta de dados, a classe *MD_Identification* assume um papel central na arquitetura do padrão. De acordo com a ISO (2014), sua finalidade é fornecer as informações para identificar unicamente um recurso, agrupando atributos descritivos como o resumo (*abstract*), o propósito (*purpose*), o status (*status*) e as restrições de uso (*resourceConstraints*).

Sob a ótica dos princípios FAIR, especialmente na encontrabilidade (*Findable*), essa classe revela-se indispensável para a modelagem na Agrometeorologia. Nela, a norma estabelece as propriedades de resolução espacial (*spatialResolution*), resolução temporal (*temporalResolution*) e as extensões geográficas e temporais (*extent*) do dado catalogado. Consequentemente, a estruturação precisa dessas informações para que pesquisadores e sistemas computacionais localizem com exatidão a abrangência territorial e o período de ocorrência de variáveis climáticas registradas, preparando o conjunto de dados para as etapas posteriores de integração.

2.3.2 Padrão Darwin Core e a agrobiodiversidade

O *Darwin Core* (DwC) é um padrão de metadados desenvolvido para facilitar o compartilhamento de dados sobre a biodiversidade, promovendo a interoperabilidade semântica e sintática através do fornecimento de identificadores e definições padronizadas (Wieczorek *et*

al., 2012)². Ele é mantido pelo *Darwin Core Maintenance Group*, uma entidade ligada ao *Biodiversity Information Standards* (TDWG), e é definido como um conjunto de termos, campos e propriedades para descrever ocorrências biológicas, espécimes, eventos de coleta, localizações e identificações taxonômicas (*Darwin Core Maintenance Group*, 2021).

A estrutura do *Darwin Core* é composta por um conjunto de termos padronizados, cada um possuindo uma definição formal e um identificador único (IRI) com metadados associados, assegurando clareza semântica e a rastreabilidade entre sistemas (Wieczorek *et al.*, 2012). Esses termos são organizados em classes, também chamados de conjuntos temáticos, agrupando propriedades relacionadas, permitindo a representação de diferentes dimensões de um registro da biodiversidade (*Darwin Core Maintenance Group*, 2021). O Quadro 4 apresenta as principais classes do *Darwin Core* e suas respectivas finalidades:

Quadro 4 – Classes do Darwin Core

(continua)

Classe	Descrição	Exemplos de principais termos
<i>Record-level</i>	Contém informações gerais aplicáveis a qualquer tipo de registro, como identificadores e data de criação.	<i>dwc:datasetID</i> , <i>dwc:institutionCode</i> , <i>dwc:basisOfRecord</i>
<i>Occurrence</i>	Descreve a ocorrência de um organismo em um determinado local e tempo, sendo uma das classes centrais do padrão.	<i>dwc:occurrenceID</i> , <i>dwc:recordedBy</i> , <i>dwc:individualCount</i> , <i>dwc:sex</i> , <i>dwc:lifeStage</i>
<i>Event</i>	Reúne dados sobre o evento de coleta, observação ou amostragem. Pode incluir data, local, método e condições do evento.	<i>dwc:eventID</i> , <i>dwc:eventDate</i> , <i>dwc:samplingProtocol</i> , <i>dwc:fieldNumber</i>
<i>Location</i>	Abrange as informações geográficas associadas à ocorrência ou ao evento.	<i>dwc:country</i> , <i>dwc:stateProvince</i> , <i>dwc:decimalLatitude</i> , <i>dwc:decimalLongitude</i> , <i>dwc:coordinateUncertaintyInMeters</i>
<i>Taxon</i>	Representa a classificação taxonômica do organismo observado, permitindo vinculação com vocabulários taxonômicos controlados.	<i>dwc:taxonID</i> , <i>dwc:scientificName</i> , <i>dwc:kingdom</i> , <i>dwc:family</i> , <i>dwc:specificEpithet</i>
<i>Organism</i>	Refere-se ao indivíduo biológico em si, possibilitando vincular múltiplas observações ao mesmo organismo.	<i>dwc:organismID</i> , <i>dwc:organismName</i> , <i>dwc:organismScope</i> , <i>dwc:associatedOrganisms</i>
<i>Identification</i>	Descreve o processo de identificação taxonômica do organismo, incluindo quem realizou a identificação e quando.	<i>dwc:identificationID</i> , <i>dwc:identifiedBy</i> , <i>dwc:dateIdentified</i> , <i>dwc:identificationRemarks</i>
<i>MeasurementOrFact</i>	Utilizada para armazenar medições, observações ou fatos associados à ocorrência ou ao organismo.	<i>dwc:measurementType</i> , <i>dwc:measurementValue</i> , <i>dwc:measurementUnit</i>

² Embora datada de 2012, a publicação revisada por pares de Wieczorek *et al.* é designada pela documentação oficial atual do *Darwin Core Maintenance Group* (2021) como a citação normativa e estrutural exigida para referenciar o padrão em trabalhos acadêmicos.

(conclusão)

Classe	Descrição	Exemplos de principais termos
<i>ResourceRelationship</i>	Registra relações entre diferentes recursos, como espécimes relacionados, hospedeiro-parasita, entre outros.	<i>dwc:resourceID</i> , <i>dwc:relatedResourceID</i> , <i>dwc:relationshipOfResource</i>
<i>GeologicalContext</i>	Inclui dados geológicos e estratigráficos que contextualizam a origem dos espécimes ou fósseis.	<i>dwc:earliestAgeOrLowestStage</i> , <i>dwc:lithostratigraphicTerms</i> , <i>dwc:geologicalContextID</i>

Fonte: O autor.

Nota: Elaborado com base em *Darwin Core Maintenance Group* (2021).

Embora o Quadro 4 apresente a totalidade da estrutura do *Darwin Core*, para o contexto da integração agrometeorológica destacam-se classes como Evento (*Event*), Localização (*Location*), Ocorrência (*Occurrence*) e, de forma especial, Medição ou Fato (*MeasurementOrFact*).

A classe *Occurrence* atua como o núcleo do registro biológico ao documentar a existência de um organismo em um local e tempo específicos. As classes *Event* e *Location* fornecem o suporte espaço-temporal para qualquer correlação climática, determinando a ação, o período, o protocolo de amostragem e as coordenadas geográficas exatas do registro georreferenciado. Já a classe *MeasurementOrFact* opera como a principal ponte de interoperabilidade com outros campos de estudos, suportando a associação de medições, fatos e características aos registros de ocorrência ou evento, viabilizando o registro padronizado de variáveis para a agricultura, a exemplo da temperatura, conforme a *Darwin Core Maintenance Group* (2021).

A identificação por um IRI em cada termo permite a integração semântica em ambientes da Web e em arquiteturas baseadas em *Linked Data*, característica essencial para aplicações envolvendo ontologias, favorecendo a interoperabilidade com outros padrões e o reuso de dados em sistemas distribuídos (*Darwin Core Maintenance Group*, 2021).

Além disso, a documentação ressalta o padrão permitindo o uso de vocabulários controlados e identificadores persistentes para garantir consistência e minimizar ambiguidades semânticas, além do alinhamento com os princípios FAIR.

Conforme ilustram Meyer *et al.* (2021), a base semanticamente orientada do *Darwin Core* permite a geração de mapeamentos semânticos e sintáticos para harmonizá-lo e integrá-lo a outros padrões de domínios distintos. Essa característica evita a criação de silos de dados (*data silos*), como consequência da existência de especificações de padrões de metadados separados, impedindo a harmonização total e levam à fragmentação ou isolamento dos dados. Essa ênfase na semântica permite aos dados de biodiversidade serem coesos e corretamente interpretados em plataformas distribuídas.

Para Wiecezorek *et al.* (2012), o *Darwin Core* é implementado, principalmente, em dois formatos: o *Simple Darwin Core* e o *Darwin Core Archive* (DwC-A). O *Simple Darwin Core* é conceituado pelos autores como uma representação tabular, geralmente em formato CSV, em que as colunas correspondem aos nomes dos termos do padrão e as linhas representam registros individuais. Já o *Darwin Core Archive* é caracterizado por Wiecezorek *et al.* (2012) como um pacote composto por múltiplos arquivos de texto (*core* e extensões) e um arquivo de metadados denominado “meta.xml”, descrevendo a estrutura do conjunto de dados, delimitadores, tipos de registro e chaves de relacionamento.

Essa arquitetura modular permite, segundo os autores, representar relações complexas, como múltiplas medições associadas a uma única ocorrência, e é amplamente utilizada em repositórios de biodiversidade, como o *Global Biodiversity Information Facility* (GBIF). Além disso, essa estrutura viabiliza a extensibilidade arquitetural, capacidade do padrão ser estendido para incorporar termos de domínios complementares ou especializados, conforme apontado por Meyer *et al.* (2021).

Essa característica de extensibilidade do *Darwin Core* é aplicada em domínios científicos como a agrobiodiversidade. A agrobiodiversidade é definida por Soares, Saraiva e Drucker (2020) como a diversidade biológica relevante para a alimentação e a agricultura. O monitoramento e a preservação dessa diversidade desempenham um papel central na produção alimentar global, especialmente considerando que, conforme destacado pela OECD/FAO (2023), os sistemas agrícolas enfrentam atualmente o triplo desafio de garantir a segurança alimentar e nutricional para uma população crescente, apoiar os meios de subsistência e promover a resiliência e a sustentabilidade ambiental frente às mudanças climáticas.

A aplicação do *Darwin Core* em dados de agrobiodiversidade é viável devido ao fato dos dados compartilharem elementos centrais com o padrão, tais como a descrição de táxon, localização e o período de coleta, permitindo ser representados como *Linked Data* e aplicando o modelo de relacionamento de recursos RDF (Soares; Saraiva; Drucker, 2020).

Dessa forma, é possível desenvolver extensões de metadados incorporando conceitos específicos ausentes no *Darwin Core*, como descritores de culturas agrícolas, solidificando o padrão para a integração de dados biológicos e ecológicos, independentemente do ambiente natural ou agrícola onde a ocorrência foi registrada.

Para operacionalizar essa representação em RDF e possibilitar a integração com grafos de conhecimento e ontologias, o *Darwin Core Maintenance Group* (2021) estabelece diretrizes técnicas específicas:

- a) Cada registro deve possuir um identificador global e persistente (IRI);

- b) Deve-se evitar o uso de literais quando houver um identificador de recurso disponível, ou seja, deve-se evitar valores textuais para identificar uma entidade, priorizando o uso de vocabulários controlados e do *namespace* específico para identificadores não literais (*dwciri:*). Como exemplo, um identificador taxonômico do GBIF deve ser usado para uma espécie em vez de um simples texto, criando, assim, conexões semânticas entre diferentes bases de dados;
- c) Toda entidade descrita em RDF deve declarar explicitamente seu tipo para ser possível interpretar corretamente seu papel na estrutura do grafo, como por exemplo, para a entidade *dwc:Occurrence*, o tipo “Ocorrência”.
- d) Devem ser incluídos metadados de autoria, licença (*dcterms:license*) e modificação nos conjuntos de dados publicados, garantindo a transparência, os direitos de uso e a rastreabilidade das informações.

Desse modo, os termos do *Darwin Core* podem ser utilizados como propriedades ou classes em modelos semânticos, permitindo a vinculação a outros vocabulários e padrões consolidados, como o *Dublin Core*. Essa arquitetura consolida o padrão como plenamente compatível com a Web Semântica e com a publicação de dados conectados (Wieczorek *et al.*, 2012). Consequentemente, o *Darwin Core* transcende a sua finalidade original de catalogação biológica tradicional, apresentando-se como uma infraestrutura de dados robusta e extensível, perfeitamente capacitada para harmonizar informações de agrobiodiversidade e variáveis agrometeorológicas em um ecossistema tecnológico integrado.

2.3.3 Padrão WMO Core Metadata Profile e as observações meteorológicas

No escopo da Organização Meteorológica Mundial (OMM), a padronização e o compartilhamento de metadados são regidos pelo *WMO Core Metadata Profile* (WCMP). Inicialmente, em sua versão 1.3, o padrão foi estruturado como um perfil estrito da norma geográfica ISO 19115, exigindo codificação dos registros em formato XML (WMO, 2023). Contudo, visando reduzir as barreiras de descoberta de dados e promover maior interoperabilidade, a WMO (2025) ressalta que a OMM modernizou sua infraestrutura tecnológica para o ecossistema WIS2 (*WMO Information System 2.0*). Essa transição impulsionou o desenvolvimento do WCMP Versão 2, substituindo a antiga arquitetura orientada a serviços (SOA) por uma arquitetura orientada a recursos (ROA), alinhada às melhores práticas da Web, à codificação GeoJSON e aos princípios FAIR.

Para operacionalizar essa nova arquitetura, a WMO (2025) formalizou o WCMP Versão 2 como uma extensão do padrão *OGC API – Records – Part 1: Core*. O seu objetivo principal é descrever, estruturar e uniformizar a descoberta de recursos focando na granularidade de conjuntos de dados (*datasets*) das áreas de meteorologia, clima e hidrologia.

Ao adotar o GeoJSON como formato oficial de codificação, a WMO (2025) destaca que o padrão passou a representar a extensão espacial e os registros de metadados de forma mais leve e passível de rastreamento e indexação nativa via HTTP. Essa característica torna o WCMP compatível com práticas modernas de interoperabilidade por meio da Web Semântica e do *Linked Data*. Além disso, embora a estrutura tecnológica tenha mudado para se alinhar à Web, a documentação da norma assegura aos conceitos de descoberta espacial e temporal, originalmente descritos na ISO 19115, permanecerem mapeados nas propriedades do GeoJSON. Essa característica viabiliza o mapeamento estrutural direto entre o WCMP atual e o padrão geográfico clássico da ISO 19115.

Em sua estrutura geral, a documentação oficial do WCMP organiza os metadados em registros compostos por propriedades obrigatórias e opcionais, abrangendo as dimensões para a descoberta, acesso e rastreabilidade dos dados (WMO, 2025). O Quadro 5 detalha essas propriedades e suas funções:

Quadro 5 – Propriedades e Funções do WCMP

(continua)

Propriedade	Requisito Obrigatório	Descrição e Função
<i>id</i>	Sim	Identificador único do <i>dataset</i> , para consulta e identificação no Catálogo Global de Descoberta (GDC).
<i>type</i>	Sim	Valor fixo (<i>Feature</i>) para conformidade com a estrutura do GeoJSON.
<i>conformsTo</i>	Sim	Indica a versão do padrão WCMP a que o registro está em conformidade.
<i>geometry</i>	Sim	Localização geoespacial delimitadora do <i>dataset</i> , usando o sistema de referência de coordenadas geográficas (WGS 84).
<i>time</i>	Sim	Extensão temporal do <i>dataset</i> , devendo usar o calendário Gregoriano e representações válidas do ISO 8601.
<i>properties.title</i>	Sim	Nome do <i>dataset</i> , legível por humanos.
<i>properties.description</i>	Sim	Resumo em texto livre do <i>dataset</i> .
<i>properties.themes</i>	Sim	Listas de conceitos que classificam o recurso, referenciadas a vocabulários controlados.
<i>properties.contacts</i>	Sim	Informações de contato associadas à parte responsável pelo <i>dataset</i> .
<i>properties.type</i>	Sim	Tipo de recurso descrito (ex.: <i>dataset</i> , serviço ou processo).
<i>properties.created</i>	Sim	Data de criação do registro de metadados WCMP.

(conclusão)

Propriedade	Requisito Obrigatório	Descrição e Função
<i>links</i>	Sim	URLs (<i>Uniform Resource Locator</i>) e APIs para acesso aos dados ou subscrição de notificações, sendo o principal mecanismo de interação
<i>properties.wmo:dataPolicy</i>	Condicional	Classificação do <i>dataset</i> como <i>core</i> ou <i>recommended</i> , sendo obrigatória se o recurso for um <i>dataset</i> .

Fonte: O autor.

Nota: Elaborado com base em WMO (2025).

O Quadro 5 evidencia o WCMP Versão 2 transcendendo a mera catalogação descritiva, estruturando-se como um recurso nativo da Web. A exigência de propriedades como *geometry* e *time* assegura a manutenção do rigor de localização espaço-temporal originalmente herdado da ISO 19115, dimensão vital para a precisão dos dados agrometeorológicos.

Do ponto de vista da integração de dados, o grande diferencial desta arquitetura reside na propriedade *properties.themes* atuando como o ponto de ancoragem semântica ao exigir as classificações do conjunto de dados serem mapeadas para vocabulários controlados e grafos de conhecimento por meio de URIs (WMO, 2025).

Essa característica viabiliza a integração metodológica proposta nesta pesquisa com ferramentas como o vocabulário AGROVOC. Além disso, a presença mandatória da propriedade *links* consolida a transição da OMM para uma Arquitetura Orientada a Recursos (ROA), transformando metadados passivos em pontos de acesso dinâmicos integrados a APIs e protocolos Web, garantindo o acesso automatizado e o cumprimento integral dos princípios FAIR (WMO, 2025).

Essa organização estrutural reflete diretamente a regra de granularidade do padrão, ou seja, o escopo do registro de metadados. O WCMP exige a cada que registro forneça uma descrição no nível primário de um *dataset*.

Essa regra, segundo a WMO (2025), define que o registro de metadados não descreve cada observação individual, os grânulos de dados, mas sim o pacote de dados completo, deixando o acesso detalhado às variáveis e dimensões espaço-temporais a cargo de serviços de dados ou APIs. Ainda de acordo com a WMO (2025), essa delimitação arquitetural evita a poluição do catálogo de buscas e direcionar o usuário de forma mais eficiente aos recursos de seu interesse.

Nessa abordagem, a WMO (2025) evidencia o registro de metadados atuando como um envelope para o conteúdo do conjunto de dados. Na prática, as propriedades obrigatórias

geometry e *time* não mapeiam uma medição isolada, mas definem, respectivamente, a área geográfica máxima coberta e o intervalo de tempo total abrangido por todo o *dataset*.

Para ilustrar essa diretriz, a documentação oficial cita o caso das observações meteorológicas de superfície, mostrando que em vez de exigir a geração de um registro no catálogo de descoberta para cada medição diária realizada por uma estação, o WCMP orienta a criação de um único registro para a coleção completa.

Desse modo, o usuário localiza o *dataset* relevante rapidamente, e o acesso aos dados em uma granularidade mais fina, ou seja, aos registros individuais em coordenadas específicas, é delegado para um momento posterior, através dos *links* e APIs especificados no próprio metadado (WMO, 2025).

A WMO (2025) documenta que o WCMP atua como a base para o chamado Fluxo de Trabalho de Metadados de Descoberta (*Discovery metadata workflow*). Esse processo foi concebido para oferecer mecanismos flexíveis, permitindo aos provedores publicarem seus metadados da maneira fácil e eficiente, reduzindo as barreiras para que os usuários e aplicações descubram, acessem e recuperem essas informações.

Toda essa dinâmica operacionaliza o ecossistema WIS2, cujo Princípio 9 estabelece a obrigatoriedade de fornecer um catálogo de descoberta contendo metadados, descrevendo tanto os conjuntos de dados em si quanto os serviços e as APIs fornecidos para acessá-los.

Nesse cenário, a aplicação do WCMP está centrada em sua publicação no *Global Discovery Catalogue* (GDC), atuando de três formas principais, conforme a WMO (2025):

- a) Busca Funcional: o GDC utiliza a estrutura semântica do WCMP para suportar buscas complexas, como consultas geoespaciais (*geometry*), temporais (*time*) e baseadas em temas (*properties.themes*), além de consultas de texto completo;
- b) Acesso Direto e Automatizado: o mecanismo primário de acesso e de interação com os dados ocorre por meio da propriedade obrigatória *links*, incluindo acesso direto a URLs ou *endpoints* de API. Em caso de dados em tempo real, o WMO (2025) ressalta que o WCMP permite notificações via protocolos de mensagens *publish-subscribe*, como MQTT/MQTTS (*Message Queuing Telemetry Transport*), garantindo acesso imediato à informação atualizada.
- c) Otimização para o Mercado de Massa: a codificação em GeoJSON e a adoção de uma arquitetura baseada em recursos promovem a indexação dos metadados por mecanismos de busca comerciais, como o SEO (*Search Engine Optimization*), derrubando a barreira da descoberta.

Um aspecto do WCMP Versão 2 para a governança em ecossistemas de dados distribuídos é a adoção de uma política de acesso estrita. Conforme detalhado pela WMO (2025), se o recurso catalogado for um *dataset*, o registro deve conter obrigatoriamente a propriedade *properties.wmo:dataPolicy*. Esta propriedade classifica o dado em duas categorias baseadas na Política Unificada de Dados da OMM:

- a) dados do tipo *core*: determinam o acesso totalmente livre, aberto e irrestrito;
- b) dados do tipo *recommended*: permitem a aplicação de condições de uso, licenciamento e controle de acesso.

Para um *framework* de integração agrometeorológica, o suporte nativo a essas diretrizes é indispensável, pois garante o compartilhamento e o consumo automatizado das informações, respeitando os direitos de propriedade intelectual e o pilar de Acessibilidade dos princípios FAIR.

Avançando na estrutura do padrão, a integração semântica no WCMP Versão 2 é viabilizada pela propriedade obrigatória *properties.themes*. A WMO (2025) determina que a propriedade atua como o mecanismo de classificação do *dataset*, exigindo aos temas serem atrelados a vocabulários controlados ou sistemas de organização do conhecimento por meio de URIs, apontando diretamente para ontologias ou registros de códigos.

Para a validação do *framework* proposto nesta pesquisa, essa exigência é o ponto focal da arquitetura, pois a norma obriga todo registro ser categorizado segundo as disciplinas do Sistema Terrestre definidas na *WIS2 Topic Hierarchy*, como meteorologia, clima e hidrologia. É exatamente dentro dessa propriedade estruturada que o *framework* encontra o espaço nativo para promover o mapeamento e a desambiguação terminológica com os domínios agrícola e biológico. Ao aproveitar essa exigência de vocabulários controlados, torna-se possível criar a ponte semântica necessária para integrar as disciplinas da OMM com ferramentas consolidadas da agricultura, como o vocabulário AGROVOC.

Para assegurar que toda essa interoperabilidade estrutural seja mantida em escala global, a WMO (2025) estabelece, em seus requisitos de conformidade, que todo registro deve ser validado sem erros contra o esquema oficial do padrão, utilizando a tecnologia *JSON Schema*. Conforme a documentação oficial, esse mecanismo computacional atua como uma camada de garantia de qualidade para a *OGC API - Records*, certificando os atributos obrigatórios, como extensões espaciais, temporais e temáticas, estejam perfeitamente estruturados antes de serem catalogados.

No contexto desta pesquisa, o suporte nativo a essa validação sintática e estrutural desempenha um papel relevante, pois o uso do *JSON Schema* permitirá ao sistema verificar

automaticamente a integridade dos dados meteorológicos de entrada, rejeitando anomalias de formatação e garantindo a confiabilidade das informações conectadas ao domínio agrícola.

O WCMP Versão 2 transcende a sua função original de mero padrão de documentação estática, consolidando-se como o elemento central de uma arquitetura orientada a recursos (ROA). Ao padronizar o acesso automatizado, a governança e a validação estrutural, a norma garante a operacionalização eficiente dos princípios FAIR na descoberta e na distribuição de dados em escala global (WMO, 2025).

2.3.4 Padrão AgMES e a evolução para o AGROVOC

O Padrão *Agricultural Metadata Element Set* (AgMES) foi um padrão de metadados desenvolvido pela *Food and Agriculture Organization of the United Nations* (FAO) com o objetivo de padronizar a semântica no domínio agrícola para a descrição, descoberta de recursos, interoperabilidade e troca de dados entre diferentes tipos de sistemas de informação (FAO, 2010). Conforme a especificação oficial, o padrão foi criado para estender o *Dublin Core* (DCMI), contendo propriedades insuficientes para atender às demandas específicas do contexto agrícola.

Embora o AgMES tenha fornecido esquemas em XML e mapeamentos em RDF para descrever suas propriedades e esquemas de codificação, a FAO (2010) ressalta sua estrutura limitada em termos de reutilização semântica extensiva. O padrão foi projetado para descrever objetos de informação como documentos, websites e bases de dados, mas carecia da escalabilidade necessária para a Web Semântica moderna. Em função dessa limitação, o padrão não é mais mantido, ou seja, não possui novas atualizações, tornando-o obsoleto para novos projetos e aplicações. Como alternativa, o AGROVOC passou a desempenhar um papel central na padronização e interoperabilidade de dados agrícolas.

Segundo a FAO (2021), o AGROVOC é um vocabulário controlado multilíngue com o objetivo de representar conceitos do domínio agrícola e de suas áreas relacionadas, como alimentos, nutrição, meio ambiente e economia rural. Diferentemente da rigidez estrutural do AgMES, o AGROVOC fornece conceitos organizados e semanticamente consistentes baseados no padrão SKOS (*Simple Knowledge Organization System*), uma arquitetura pertencente à família de padrões da Web Semântica, que permite maior integração com padrões atuais de metadados, garantindo a interoperabilidade entre sistemas de informação e a reutilização de dados vinculados (*Linked Data*) em diferentes contextos semânticos.

Dentro desta arquitetura baseada no modelo SKOS, a FAO (2021) caracteriza como unidade central de informação o Conceito (*skos:Concept*), representando a ideia abstrata ou o recurso identificado de forma inequívoca por um URI dereferenciável, ou seja, um endereço que, quando acessado via navegador ou sistema, retorna informações estruturadas sobre aquele recurso na Web.

No AGROVOC, cada conceito é expresso em diversos idiomas por meio de lexicalizações, que são classificadas em termos preferenciais (*skosxl:prefLabel*), em que há obrigatoriamente apenas um por idioma, e termos alternativos (*skosxl:altLabel*), englobando sinônimos e variantes.

A adoção da Extensão SKOS-XL, segundo a FAO (2021), viabiliza o gerenciamento avançado desses nomes, permitindo os rótulos serem tratados como recursos RDF de primeira ordem e possibilitando ao sistema atribuir propriedades adicionais diretamente aos termos, como notas editoriais, definições e datas de modificação.

A partir dessa unidade central de modelagem, macroestrutura do vocabulário é definida por uma organização temática rigorosa. Conforme detalha a FAO (2021), todos os conceitos do AGROVOC estão organizados hierarquicamente sob 25 conceitos generalistas de nível superior, os *top concepts*, demonstrados no Quadro 6:

Quadro 6 – Os 25 conceitos de nível superior do AGROVOC

(continua)

Conceito de Nível Superior	Descrição / Abrangência Semântica
<i>Activities</i> (Atividades)	Ações conduzidas ao longo da cadeia de suprimento alimentar, gestão, nutrição ou lazer (ex: transporte, planejamento).
<i>Entities</i> (Entidades)	Coisas distintas e separadas, incluindo agências, rótulos, redes e políticas.
<i>Events</i> (Eventos)	Acontecimentos em um ponto no tempo com participação humana (ex: exposições, cursos de treinamento).
<i>Factors</i> (Fatores)	Elementos frequentemente citados em pesquisas, como fatores abióticos, bióticos, ambientais ou de produção.
<i>Features</i> (Características/Feições)	Relaciona-se a conceitos de geociências e genética, como feições fisiográficas ou morfológicas do solo.
<i>Groups</i> (Grupos)	Conjuntos de indivíduos ou itens, abrangendo desde grupos sociais (consumidores) até profissionais (engenheiros).
<i>Location</i> (Localização)	Pontos ou extensões no espaço, como zonas climáticas, áreas protegidas ou urbanas.
<i>Measure</i> (Medida)	Valores numéricos observáveis que possuem uma unidade anexada (ex: altitude, umidade, preço).
<i>Methods</i> (Métodos)	Formas de realizar tarefas na pesquisa ou produção (ex: irrigação, amostragem, métodos estatísticos).
<i>Objects</i> (Objetos)	Itens tangíveis fabricados por seres humanos, como equipamentos e móveis.

(conclusão)

Conceito de Nível Superior	Descrição / Abrangência Semântica
<i>Organisms</i> (Organismos)	Hierarquia taxonômica de seres vivos relevantes (plantas, animais, pragas, predadores, hospedeiros).
<i>Phenomena</i> (Fenômenos)	Eventos observáveis em ciências naturais ou econômicas (ex: perigos, tendências, dinâmica populacional).
<i>Processes</i> (Processos)	Conjunto de atividades que transformam entradas em saídas (ex: processos biológicos, síntese, evolução).
<i>Products</i> (Produtos)	Itens originados das cadeias agrícolas, como alimentos, rações, resinas e produtos sustentáveis.
<i>Properties</i> (Propriedades)	Qualidades que descrevem algo, como toxicidade, direção do vento, idade ou propriedades do solo.
<i>Resources</i> (Recursos)	Coisas usadas na produção ou necessidades humanas (ex: matérias-primas, patrimônio cultural, recursos naturais).
<i>Site</i> (Sítios/Locais)	Instalações humanas criadas para propósitos específicos (ex: estações meteorológicas, laboratórios, hospitais).
<i>Stages</i> (Estágios)	Fases do ciclo de vida ou estágios de desenvolvimento (ex: estágio embrionário ou reprodutivo).
<i>State</i> (Estado)	Qualquer condição de uma substância física ou organismo (ex: emprego, sono, estado coloidal).
<i>Strategies</i> (Estratégias)	Opções de atuação, incluindo abordagens e estratégias de comunicação ou treinamento.
<i>Subjects</i> (Assuntos)	Disciplinas de estudo ou tópicos de interesse, abrangendo ciências e humanidades.
<i>Substances</i> (Substâncias)	Hierarquia de substâncias químicas classificadas por propriedade física, função ou origem.
<i>Systems</i> (Sistemas)	Sistemas de interação humana, tecnológicos ou científicos (ex: sistemas econômicos, fotovoltaicos, de informação).
<i>Technology</i> (Tecnologia)	Invenções e desenvolvimentos técnicos modernos, como biotecnologia e tecnologia de sementes.
<i>Time</i> (Tempo)	Intervalos temporais com funções específicas, como estações, horários de trabalho ou épocas do dia.

Fonte: O autor.

Nota: Elaborado pelo autor com base em FAO (2021).

Como se observa no Quadro 6, os 25 conceitos de nível superior atuam como raízes estruturais para todas as hierarquias do AGROVOC, com cada subconceito contido no sistema representando um "tipo de" conceito subordinado a uma dessas categorias principais (FAO, 2021). Um exemplo prático dessa estrutura, citado pela própria FAO (2021), ocorre com a categoria *Activities*, que possui agrupados subconceitos representando tipos específicos de ações ligadas à cadeia de suprimentos, como melhoramento (*breeding*), produção (*production*) e tratamento de sementes (*seed treatment*).

Essa organização temática detalhada atende diretamente à multidisciplinaridade exigida para a modelagem de dados agrometeorológicos. Enquanto categorias como *Organisms* e *Stages* permitem mapear as culturas agrícolas e os seus respectivos ciclos biológicos e de desenvolvimento, outras raízes fornecem o arcabouço semântico exato para a integração dos

dados climáticos, como as categorias como *Phenomena*, para mapear fenômenos naturais e perigos ambientais, *Measure*, voltado para medidas quantitativas observáveis, como umidade e altitude, e *Location*, que possibilita registrar recortes espaciais, como zonas climáticas. Dessa forma, a macroestrutura do vocabulário comprova a sua capacidade de acomodar e padronizar tanto as variáveis agrícolas quanto as meteorológicas de forma coesa.

Embora a organização hierárquica sob os 25 conceitos de nível superior forneça uma taxonomia rigorosa para classificar as variáveis, a modelagem de dados de domínios complexos frequentemente exige associações semânticas transversais, isto é, relações não-hierárquicas entre conceitos localizados em ramificações distintas. Para suprir essa necessidade e enriquecer a descrição dos conceitos, a FAO (2021) destaca a integração da *Agrontology*, uma ontologia de suporte desenvolvida especificamente para o AGROVOC a fim de acomodar relacionamentos específicos de domínio. Conforme o guia de diretrizes editoriais da FAO (2022), a *Agrontology* atua como um vocabulário de relações não-hierárquicas agrupadas sob a propriedade *skos:related* da Web Semântica.

Na prática, o uso de suas propriedades permite a criação de uma expressiva teia de conexões lógicas. A documentação oficial detalha dezenas dessas ligações, como as relações de causalidade (*causativeRelationship*), incluindo as propriedades *affects* (afeta) e *isAffectedBy* (é afetado por), e as relações quantitativas (*quantitativeRelationship*), englobando a propriedade *isMeasuredBy* (é medido por).

No contexto agrometeorológico, essa arquitetura relacional possibilita ao sistema estabelecer formalmente um evento extremo da raiz *Phenomena* afeta (*affects*) uma cultura agrícola catalogada em *Organisms*, transformando o vocabulário de uma simples árvore de classificação em uma rede de conhecimento profundo e integrado.

Além das conexões internas providas pela *Agrontology*, o AGROVOC atua como um *hub* central de Dados Abertos Vinculados (*Linked Open Data*) no domínio agrícola por meio de seus alinhamentos semânticos. Segundo a FAO (2021), o vocabulário está conectado a mais de 20 outros conjuntos de dados e tesouros globais, permitindo o estabelecimento de mapeamentos entre sistemas, utilizando diferentes padrões de metadados.

Esses alinhamentos são realizados utilizando propriedades de mapeamento do padrão SKOS, como *skos:exactMatch* (correspondência exata) e *skos:closeMatch* (correspondência próxima), indicando o grau de confiança de dois conceitos de vocabulários distintos ser usados de forma intercambiável.

Entre os principais alvos de alinhamento destacados pela FAO (2021) estão o *National Agricultural Library Thesaurus* (NALT), o *EuroVoc* e o *General Multilingual Environmental*

Thesaurus (GEMET), aos quais Takezaki *et al.* (2020) também acrescentam a integração direta com bases de conhecimento abertas, como a DBpedia.

Nesse contexto, Takezaki *et al.* (2020) enfatizam essa estruturação em *Linked Open Data* como garantia de interoperabilidade semântica. Tal característica é justificada, com base nos achados de Drury *et al.* (2019), pelo fato de permitir aos usuários e sistemas acessar esses múltiplos recursos externos como se fossem uma fonte única de dados.

Para o *framework* proposto, essa capacidade de alinhamento garante a interoperabilidade externa, permitindo aos dados meteorológicos e agrícolas padronizados pelo AGROVOC serem vinculados a uma rede mundial de conhecimento.

Para operacionalizar toda essa arquitetura de conhecimento em sistemas e softwares, como o *framework* proposto nesta pesquisa, a infraestrutura do AGROVOC disponibiliza métodos avançados de acesso tecnológico orientados a máquinas. Conforme detalha a FAO (2021), caso o objetivo seja incorporar o vocabulário em aplicações ou extrair subconjuntos específicos de dados, a abordagem ideal consiste na utilização de suas interfaces programáveis. Entre os principais serviços disponibilizados para essa finalidade, destacam-se a API REST provida pela plataforma *Skosmos* e o ponto de extremidade (*endpoint*) SPARQL.

A API REST permite aos métodos do vocabulário serem consumidos de forma direta pelo código da aplicação, retornando os dados em formatos estruturados e leves, como JSON e XML. Já o SPARQL é apontado pela documentação oficial como a via recomendada para acessar e manipular informações complexas. Por meio do *endpoint* SPARQL, o sistema consegue executar consultas na base de dados, permitindo extrair não apenas os conceitos, mas toda a teia de relações hierárquicas, propriedades da *Agrontology* e alinhamentos externos.

Dessa forma, essas interfaces de programação representam os canais técnicos definitivos viabilizadores do consumo automatizado dos Dados Abertos Vinculados, fornecendo o alicerce operacional para a integração de dados no *framework* desta pesquisa.

Portanto, o AGROVOC não atua como um padrão de metadados estrutural, mas é classificado por Takezaki *et al.* (2020) como um vocabulário internacional multilíngue, atuando como um mecanismo para alcançar a interoperabilidade semântica no domínio agrícola.

No contexto deste trabalho, o AGROVOC funcionará como a ponte semântica na arquitetura de integração de dados, pois sua natureza como vocabulário internacional comum permite, seguindo a premissa de desambiguação demonstrada por Takezaki *et al.* (2020), que termos diferentes, como “milho” em um sistema e “*corn*” em outro, sejam mapeados para o mesmo conceito formal através de identificadores URIs únicos.

Ao solucionar essas inconsistências de terminologia, representando um dos maiores gargalos para os sistemas de tecnologia atuais, a adoção do AGROVOC viabiliza aos dados agrometeorológicos de fontes distintas serem recuperados, comparados e consumidos de forma unificada pelo *framework* proposto.

2.3.5 Convergências e Divergências dos Padrões para a Integração Agrometeorológica

A análise das seções anteriores demonstra que, embora os padrões ISO 19115, *Darwin Core* e *WMO Core Metadata Profile* (WCMP) sirvam a domínios primários distintos, todos compartilham o objetivo conceitual comum de descrever um recurso de dados de forma a garantir os princípios FAIR, ocorrendo fundamentalmente a partir de sua identificação, cobertura espacial e cobertura temporal, para facilitar sua descoberta, acesso e reutilização das informações nos ecossistemas ambientais e climáticos.

Essa convergência conceitual possibilita o uso dos padrões de forma interdisciplinar, porém com limitações devido a divergências técnicas e estruturais, representando barreiras à interoperabilidade direta. Além disso, a falta de padronização terminológica impõe uma forte barreira semântica, exigindo a adoção complementar de ferramentas como o vocabulário AGROVOC para desambiguação e o alinhamento de conceitos. Esse cenário evidencia, portanto, a necessidade de uma arquitetura de integração capaz de orquestrar as complementaridades e mitigar as divergências estruturais e semânticas entre esses padrões.

O Quadro 7 apresenta uma análise comparativa dos Padrões de Metadados adotados na Agrometeorologia.

Quadro 7 – Análise Comparativa dos Padrões de Metadados Agrometeorológicos

(continua)

Característica Central	Padrão ISO 19115	Padrão <i>Darwin Core</i>	Padrão <i>WMO Core Metadata Profile</i>
Domínio Primário	Dados e serviços geoespaciais	Ocorrências de biodiversidade	Dados meteorológicos, climáticos e hidrológicos
Formato	XML (baseado em UML)	Texto simples em CSV ou pacote compactado (<i>Darwin Core Archive</i> - DwC-A) e representação estrutural em RDF para a Web Semântica	GeoJSON (baseado em JSON)
Granularidade	Flexível, mas tipicamente focado no <i>dataset</i> , serviços ou série de dados.	Focado na ocorrência ou evento (ex: uma única observação, medição ou espécime)	Estritamente focado no <i>dataset</i> (o "envelope" ou pacote de dados), não na observação ou medição individual

(conclusão)

Característica Central	Padrão ISO 19115	Padrão <i>Darwin Core</i>	Padrão <i>WMO Core Metadata Profile</i>
Cobertura Espacial	Classe complexa (<i>MD_SpatialRepresentation</i>) e propriedades de extensão (<i>extent</i>) dentro de <i>MD_Identification</i>	Termos simples (ex: <i>dwc:decimalLatitude</i> , <i>dwc:decimalLongitude</i>)	Propriedade obrigatória estruturada em GeoJSON (<i>geometry</i>)
Cobertura Temporal	Classe complexa e propriedades de extensão (<i>extent</i>)	Termo simples (ex: <i>dwc:eventDate</i>)	Propriedade simples e obrigatória (<i>time</i> , padrão ISO 8601)
Descrição Temática (Conceito)	<i>MD_Keywords.keyword</i> (Tipicamente texto livre, ambíguo)	Foco em taxonomia (<i>dwc:scientificName</i>), mas com diretriz de priorizar identificadores globais URIs (<i>namespace dwciri:</i>) ligados a vocabulários controlados em vez de texto livre	<i>properties.themes</i> (Exige mapeamento para vocabulários controlados e ontologias via URIs)
Medição Específica	Não é o foco principal	Central (Classe <i>dwc:MeasurementOrFact</i> com termos <i>measurementType</i> , <i>measurementValue</i> , <i>measurementUnit</i>)	Ausente no metadado (Foco no <i>dataset</i>). O acesso aos dados brutos e granulares é delegado para <i>links</i> e APIs externas

Fonte: O autor.

Como se observa no Quadro 7, a principal convergência técnica entre os três padrões estruturais reside no suporte nativo à cobertura espacial e temporal, dimensões vitais para a precisão dos dados agrometeorológicos. Todos operam com a premissa de localizar o dado na Terra e no tempo, seja por meio das propriedades obrigatórias *geometry* e *time* do padrão WCMP, do suporte às classes *Event* e *Location* no *Darwin Core*, ou da classe *MD_SpatialRepresentation* e extensões de identificação do ISO 19115. Essa harmonização base garante que qualquer sistema adote esses padrões consiga atender ao pilar da Encontrabilidade (*Findable*) dos princípios FAIR.

Por outro lado, o Quadro 7 evidencia divergências arquiteturais, especialmente à granularidade e à medição específica. O WCMP e o ISO 19115 atuam primariamente num nível macro, operando como um "envelope" para o conjunto de dados completo (*dataset*), delegando o acesso às medições individuais para *links* e serviços externos.

Conforme apontam Annane *et al.* (2023), os esforços tradicionais de descrição de dados meteorológicos limitados a metadados de alto nível, focados nas características gerais do conjunto de dados, são insuficientes para garantir a interoperabilidade e a reutilização plena por especialistas de outras comunidades científicas.

Essa limitação reflete exatamente o escopo macro adotado pelo padrão WCMP. Em contrapartida, o *Darwin Core* apresenta uma granularidade muito mais fina, estruturando-se no

nível da observação individual através da classe *MeasurementOrFact*, atuando como uma ponte de interoperabilidade central para registrar variáveis específicas e locais da agricultura.

A complexidade prática da divergência de granularidade e de formato é evidenciada na necessidade de integrar dados brutos em grande escala. Conforme demonstrado por Arsenault *et al.* (2020) no desenvolvimento de bases de dados hidrometeorológicas, a modelagem de sistemas ambientais exige a combinação de múltiplas fontes heterogêneas, como observações de estações físicas, produtos interpolados e dados de reanálise, originalmente disponibilizadas em formatos distintos e resoluções espaciais variadas.

Essa incompatibilidade estrutural primária impede a interoperabilidade direta, exigindo às diferentes fontes passarem por um rigoroso processo computacional de conversão de formato e harmonização de escala para viabilizar análises conjuntas.

Trazendo esse desafio para o domínio específico da Agrometeorologia, a tentativa de integrar coleções de dados climáticos, tipicamente envelopados em formatos alinhados ao WCMP e ISO 19115, com ocorrências biológicas locais, padronizadas em *Darwin Core*, esbarra exatamente na mesma barreira operacional, reforçando a necessidade de uma infraestrutura de integração, automatizando o alinhamento dessas estruturas.

Diante dessa dicotomia, uma análise agrometeorológica integrada como, por exemplo, o cruzamento de dados climáticos com medições biológicas granulares sobre o desenvolvimento de culturas ou ocorrência de pragas, exige uma arquitetura de integração capaz de normalizar essa diferença de escopo e representar tanto o agrupamento geral quanto a estrutura interna e a semântica das observações, como o *framework* proposto nesta pesquisa.

Outra divergência crítica apresentada no Quadro 7 ocorre na camada de Descrição Temática. Enquanto o padrão ISO 19115 permite o uso de texto livre e ambíguo nas suas palavras-chave através do elemento *MD_Keywords.keyword*, o *Darwin Core* e o WCMP impulsionam a transição para a Web Semântica ao exigirem ou recomendarem o uso de identificadores globais. O WCMP exige o mapeamento para vocabulários controlados via URIs na propriedade *properties.themes*, e o *Darwin Core* estabelece a diretriz de priorizar o *namespace* “*dwciri:*” para conexões semânticas, evitando valores textuais isolados.

No contexto da integração de dados ambientais e de biodiversidade, conforme alertam Zárate *et al.* (2020), o uso prevalente de vocabulários, não sendo ontologicamente definidos limita a integração à mera agregação de conjuntos de dados, impedindo a interoperabilidade plena ser alcançada para o consumo autônomo por máquinas.

Aplicando essa constatação à dicotomia dos padrões avaliados, um sistema baseado em ISO 19115 poderia preencher o campo *MD_Keywords.keyword* simplesmente com o texto livre "Milho", estando sujeito a ambiguidades e regionalismos.

A fragilidade dessa abordagem baseada em texto livre é atestada por Michel *et al.* (2022), que demonstram que os modelos dependentes de palavras-chave isoladas falham em capturar os relacionamentos semânticos transversais em sistemas agrícolas e de biodiversidade.

Por outro lado, um registro biológico em *Darwin Core* focaria estritamente na taxonomia formal *dwc:scientificName*, que, como Zárate *et al.* (2020) ressaltam, tradicionalmente carece de uma definição ontológica nativa. Já o WCMP exigiria a vinculação a uma ontologia.

Essa heterogeneidade descritiva cria uma barreira de ambiguidade semântica, impedindo a consulta integrada de informações sobre uma mesma cultura ou fenômeno transversal.

Nesta necessidade de padronização temática, o AGROVOC se insere como a quarta tecnologia desse arranjo multidisciplinar. Por não ser um padrão de metadados estrutural, mas sim um vocabulário internacional multilíngue fundamentado no modelo SKOS, o AGROVOC não compõe a comparação técnica estrutural do Quadro 7. Em vez disso, ele atua como a ponte semântica unificando as divergências descritivas dos outros padrões, ao fornecer os identificadores URIs únicos necessários para alimentar a propriedade *properties.themes* do WCMP e as identificações do *Darwin Core*.

Essa abordagem é fortemente corroborada por Xu *et al.* (2023), onde a verdadeira interoperabilidade só é alcançada quando os metadados adotam vocabulários seguindo os princípios FAIR, caracterizados pelo uso de identificadores globais únicos e representação formal compreensível por máquinas. Ao utilizar esse vocabulário controlado para conectar um fenômeno climático a uma cultura agrícola dentro de um mesmo ecossistema de conhecimento, torna-se possível mitigar as ambiguidades terminológicas, viabilizando o consumo automatizado e a verdadeira interoperabilidade entre os domínios.

Além das barreiras de granularidade e de semântica, a análise comparativa do Quadro 7 evidencia uma profunda divergência nos esquemas de codificação sintática adotados por cada padrão. Enquanto o ISO 19115 estrutura-se tradicionalmente em XML, o WCMP moderno opera nativamente em GeoJSON e o *Darwin Core* transita entre arquivos de texto simples (CSV) e pacotes estruturados DwC-A. Conforme demonstrado por Putrama e Martinek (2024), essa diversidade de fontes e formatos torna o campo da integração de dados ambientais complexo.

Essa heterogeneidade impede que um sistema de análise processe de forma nativa e integrada, tornando indispensável a implementação de um processo tecnológico automatizado de tradução e normalização dos metadados.

Desse modo, a análise contundente demonstra que a integração direta entre as fontes agrometeorológicas não é viável. A superação dessas barreiras exige uma arquitetura de integração capaz de atuar em duas frentes complementares, refletindo os níveis de interoperabilidade definidos por Vogt *et al.* (2025).

Primeiramente, exige-se o alcance da interoperabilidade proposicional por meio do cruzamento de esquemas, normalizando as divergências de formato e as assimetrias de granularidade. Além disso, o cenário comprova a necessidade de garantir a interoperabilidade terminológica através de uma "ponte semântica", papel a ser desempenhado pelo vocabulário AGROVOC, para mapear as entidades e harmonizar os conceitos temáticos ambíguos. Somente interligando de forma estrutural e linguística torna-se possível viabilizar a verdadeira interoperabilidade no ecossistema de dados agrometeorológicos, justificando a arquitetura proposta nesta pesquisa.

2.4 ESTADO DA ARTE

A busca pela interoperabilidade de dados agrometeorológicos esbarra em heterogeneidades estruturais e semânticas, refletindo a complexidade de se unificar domínios intrinsecamente distintos.

A análise dos padrões de metadados demonstrou as divergências de granularidade, evidenciadas pelo foco no conjunto de dados (*dataset*) do WCMP e da ISO 19115 em contraposição à granularidade fina do *Darwin Core*, que impedem a integração entre dados climáticos e biológicos ocorrer de forma direta.

Somado a isso, a transição para a Web Semântica revela que o simples uso de ontologias e identificadores globais não são suficientes se os modelos permanecerem isolados, gerando "silos de informação" incapazes de se comunicarem.

Diante desta dicotomia teórica, esta seção do Estado da Arte tem como objetivo investigar como a comunidade científica tem enfrentado esses desafios estruturais e semânticos na prática. A partir dos resultados obtidos por meio de uma revisão sistemática da literatura, serão analisadas as principais tecnologias, *frameworks* e aplicações práticas de metadados desenvolvidas nos últimos anos.

Para acomodar o volume de pesquisas recuperadas e manter a coerência analítica com as convergências e divergências já mapeadas nesta fundamentação, esta seção foi dividida em três abordagens principais. Primeiramente, são analisados os trabalhos que aplicam padrões em dados climáticos e geoespaciais. Em seguida, são exploradas as pesquisas focadas na aplicação de padrões na agrobiodiversidade e na marcação semântica de observações locais. Finalmente, são apresentadas as iniciativas recentes para promover a integração entre padrões de metadados heterogêneos.

2.4.1 Aplicações de Padrões em Dados Climáticos e Geoespaciais

Para investigar as raízes do isolamento tecnológico na Agrometeorologia, é imprescindível analisar inicialmente como a própria comunidade climática estrutura a sua informação. No domínio estritamente meteorológico, as iniciativas revelam um foco histórico na eficiência de armazenamento físico em detrimento da interoperabilidade semântica. O'Grady *et al.* (2021) destacam a modelagem climática e a previsão do tempo para padrões otimizados para o armazenamento de matrizes multidimensionais e dados em grade, sendo os formatos NetCDF, GRIB (*GRIdded Binary*) e HDF as tecnologias proeminentes adotadas globalmente. Embora essas estruturas operem de forma excepcional para o processamento massivo de variáveis físicas, perpetuam uma arquitetura priorizando o desempenho computacional e a conveniência da física atmosférica em detrimento de uma verdadeira representação semântica.

Uma crítica contundente a esse modelo estrutural é amparada por um estudo da iniciativa *Global Open Data for Agriculture and Nutrition* (GODAN), relatado por O'Grady *et al.* (2021). Os padrões de dados climáticos consolidados na comunidade são, paradoxalmente, os menos alinhados aos princípios FAIR. Os autores enfatizam o grave risco de se confundir a adequação a um padrão tecnológico com a conformidade de uso dos dados, apontando que, na prática climática, métricas de adoção e conveniência frequentemente se sobrepõem à capacidade do dado de ser reutilizado por domínios externos.

O impacto severo dessa deficiência semântica materializa-se diretamente quando há a tentativa de integrar essas matrizes físicas brutas a modelos de decisão agrícola baseados na biologia, como os sistemas de previsão de pragas e doenças. O'Grady *et al.* (2021) advertem a incapacidade de interpretar metadados meteorológicos com precisão como a principal causa de falha em modelos fenológicos e mecanicistas de proteção de cultivos.

Ao importar dados de temperatura e umidade sem o uso de vocabulários controlados capazes de desambiguar falhas de medição, como lacunas temporais ou picos anômalos por desvio de calibração, o modelo biológico herda o caos estrutural do clima.

Consequentemente, o limite térmico de uma praga ou o limiar de esporulação de um fungo podem ser ativados erroneamente, resultando na falência do modelo de alerta. Fica comprovado que a entrega do dado climático isolado carece de uma ponte semântica que valide o seu contexto para a agronomia.

Essa limitação nativa dos formatos meteorológicos físicos leva a comunidade científica recorrer aos padrões geoespaciais e de metadados focados no nível *dataset* para tentar organizar a descoberta da informação climática.

A tentativa prática de solucionar a falta da adesão aos princípios FAIR dos formatos estritamente climáticos por meio de seu empacotamento com padrões de metadados geoespaciais é perfeitamente ilustrada pelo desenvolvimento do conjunto de dados MADIA (*Meteorological variables for Agriculture: a Dataset for the Italian Area*).

Neste estudo, Parisse *et al.* (2022) processaram dados horários da reanálise climática ERA5 para derivar variáveis agrometeorológicas para a Itália, como radiação solar de ondas curtas e a evapotranspiração de referência calculada pelo método *FAO Penman-Monteith*.

Para estruturar essa volumosa matriz física, os autores adotaram o formato de cubo de dados *NetCDF*, garantindo a sua descoberta por meio de uma anotação rigorosa de metadados globais e específicos baseada no Padrão ISO 19115, na Convenção *Climate and Forecast* (CF) e no padrão ACDD (*Attribute Convention for Dataset Discovery*).

Apesar do alto rigor técnico na construção deste arcabouço unindo matrizes multidimensionais e envelopes espaciais, o estudo de Parisse *et al.* (2022) revela uma concessão arquitetural expondo a fragilidade pragmática do modelo para integrações de dados heterogêneos.

O Formato *NetCDF* e sua complexa estrutura de metadados embutidos criam uma barreira de usabilidade para pesquisadores de outras áreas, como a agronomia, que não possuem familiaridade com cubos de dados.

Para contornar essa alienação tecnológica e tornar os dados efetivamente consumíveis, Parisse *et al.* (2022) foram forçados a replicar integralmente toda a série histórica climática em um formato de tabelas de texto plano separadas por vírgula, ou seja, em arquivos no formato “.csv”.

Esse recuo metodológico atua como uma comprovação empírica das limitações previamente discutidas. A tentativa de encapsular matrizes físicas, como a *NetCDF*, com

padrões no nível do *dataset*, como o ISO 19115, gera uma estrutura tão rígida e isolada em seu próprio domínio, onde os criadores do modelo precisam recorrer a planilhas de texto plano em “.csv” para garantir o acesso real aos agrônomos.

Ao "achatar" a complexidade do cubo de dados em arquivos de texto bidimensionais, perde-se o potencial de controle ontológico e de navegabilidade inerente à Web Semântica. Fica evidente, portanto, a ponte entre o clima e a biologia não pode depender de formatos físicos alienantes mitigados por arquivos de texto, mas sim de um modelo de representação do conhecimento comum e desambiguado por vocabulários controlados externos, reforçando a pertinência do uso de ontologias de domínio para viabilizar a interoperabilidade agrometeorológica.

No domínio específico dos dados geoespaciais, as iniciativas de interoperabilidade concentram-se fortemente no alinhamento estrutural de padrões focados no nível do *dataset*. Han e Han (2022) propuseram um modelo integrado de metadados para recursos de geociências visando a preservação de dados históricos. A técnica tradicional de mapeamento cruzado (*cross-walking*) torna-se insustentável a partir do aumento do número de esquemas, pois exige mapeamentos bidirecionais complexos entre cada par de padrões.

Para contornar esse problema, a pesquisa adotou a metodologia de *switching-across*, canalizando as transformações por meio de um esquema central. Na prática, Han e Han (2022) definiram o ISO 19115 como o modelo representativo e mapearam para ele os padrões CSDGM, o Perfil de Metadados ANZLIC e o modelo INSPIRE.

O aprofundamento técnico do trabalho consistiu em analisar as similaridades semânticas e as regras de obrigatoriedade, como mandatório, condicional ou opcional, de cada elemento, fundindo redundâncias e criando um modelo com classes comuns, como “Identificação”, “Qualidade de Dados” e “Representação Espacial”.

Além disso, para lidar com dados legados, os autores precisaram expandir o ISO 19115 original, adicionando classes como “Representação Temporal” e elementos de “Informação de Aquisição” para descrever os equipamentos de medição.

Embora o modelo de Han e Han (2022) seja bem-sucedido ao criar regras consistentes de interoperabilidade semântica para o acesso de classes equivalentes entre esses quatro padrões, ele evidencia a fronteira metodológica do domínio geoespacial. A solução é estritamente estrutural, alinhando a sintaxe e harmoniza as regras de conteúdo no nível global do conjunto de dados, como coordenadas de contorno e sistemas de referência.

Esse tipo de arquitetura, focada no "envelope" da informação, confirma a limitação de se tentar realizar integrações diretas sem o apoio de ontologias de domínio. Sem a utilização de

vocabulários controlados externos para atuar na desambiguação terminológica granular das observações, modelos como este permanecem restritos ao isolamento do escopo geoespacial, sendo insuficientes para as demandas de uma integração agrometeorológica completa.

Enquanto as iniciativas focadas no nível do *dataset* buscam harmonizar a estrutura sintática, outros trabalhos recentes evidenciam a ausência de ontologias de domínio gerando um profundo caos terminológico, mesmo dentro de uma mesma área de pesquisa.

O contraponto para essa limitação estrutural é o estudo de Mutlu *et al.* (2025), que investigou a qualidade dos metadados espaciais no sistema EMPODAT, uma grande base de dados de monitoramento ambiental e químico de águas superficiais mantida pela rede europeia NORMAN.

A coleta de dados realizada por centenas de laboratórios ao longo de décadas, sem a adoção de vocabulários semânticos controlados, resultou em uma base severamente degradada. Foram identificados problemas massivos de dados faltantes, inconsistências de idioma, erros de digitação e coordenadas contraditórias em relação ao país da amostragem. Para mitigar essa fragmentação, a pesquisa desenvolveu o *CleanGeoStreamR*, um *framework* automatizado desenvolvido em linguagem R, voltado exclusivamente para a curadoria corretiva dessas inconsistências geográficas.

Mutlu *et al.* (2025) ressaltam o aprofundamento técnico da solução consistindo em normalizar agressivamente os metadados textuais associados aos locais de amostragem, bacias hidrográficas e corpos d'água. Para isso, o algoritmo realizou transformações para remover pontuações, padronizar caracteres e unificar variações de grafia e de idioma. Além disso, a ferramenta empregou técnicas como a geocodificação reversa, ou seja, conversão de coordenadas geográficas em nomes locais legíveis, para preencher metadados ausentes cruzando coordenadas com nomes de corpos d'água, além de calcular médias espaciais para unificar registros do mesmo local possuidor de pequenos desvios de coordenadas. Ao todo, o *framework* conseguiu preencher e corrigir cerca de 107 milhões de lacunas nos metadados espaciais.

A magnitude do esforço relatado por Mutlu *et al.* (2025) atua como uma comprovação empírica da lacuna argumentada pela presente pesquisa. O fato de pesquisadores precisarem desenvolver um complexo algoritmo corretivo *a posteriori* apenas para tentar padronizar nomes de rios, países e bacias hidrográficas inseridos de forma heterogênea, evidencia o colapso gerado por sistemas construídos como silos de texto livre.

Ao lidar apenas com o "envelope" estrutural da informação não resolve as inconsistências no nível granular do conteúdo. A verdadeira interoperabilidade exige regras de

desambiguação terminológica, reforçando de forma incontestável a necessidade da adoção de ontologias globais, como o AGROVOC proposto na presente pesquisa, para atuar como ponte semântica e normalizar os termos descritivos antes dos dados precisarem de curadoria corretiva.

Para consolidar a crítica às limitações semânticas dos padrões focados no nível do *dataset*, o estudo de Roić e Pivac (2024) comprova a insuficiência estrutural dos campos descritivos abertos. Os autores investigaram a aplicabilidade do Padrão ISO 19115-1 em documentos cadastrais analógicos, como mapas e registros de demarcação de limites, originários do estabelecimento sistemático de cadastros territoriais.

A inovação trazida pela pesquisa consistiu em realizar o mapeamento semântico das propriedades desses documentos para os pacotes de metadados do padrão, com forte atuação sobre as classes do pacote *Identification information*, como *MD_DataIdentification*, e *Lineage information*.

No entanto, ao tentarem aplicar o padrão de forma estrita, Roić e Pivac (2024) constataram uma falha estrutural, comprometendo a qualidade semântica da modelagem, ou seja, a maioria dos elementos da ISO 19115-1 permitia a inserção de dados sob o domínio de "texto livre" ou possuía obrigatoriedade apenas opcional.

Para o modelo funcionar sem gerar o caos terminológico e a ambiguidade previamente discutidos, propondo extensões formais à arquitetura do padrão, a intervenção metodológica consistiu em impor regras de obrigatoriedade mais restritivas, como transformar elementos opcionais em mandatórios, e, principalmente, substituir os domínios de texto livre por listas de códigos de valores rigorosas. Um exemplo consiste no elemento *LI_Lineage.statement*, em vez de permitir texto livre para descrever o método de levantamento do mapa, forçaram o uso de um vocabulário restrito com métodos pré-definidos, tais como gráfico, ortogonal, polar, entre outros. Ao todo, os domínios originais do padrão precisaram ser alterados e bloqueados em proporções variando de 17% a 38%, dependendo do tipo de documento mapeado, conforme o estudo.

Esta intervenção metodológica relatada por Roić e Pivac (2024) corrobora de forma contundente a premissa da presente pesquisa. Se a própria comunidade geoespacial, ao lidar estritamente com mapas e documentos cadastrais do mesmo domínio, conclui a existência de uma falha semântica no ISO 19115 e exige extensões para substituir o texto livre por vocabulários controlados restritos, torna-se evidente a impossibilidade de utilizar tais padrões em sua forma nativa para integrações complexas.

Para transcender o alinhamento puramente sintático e conectar o domínio geoespacial à granularidade da observação biológica, é imprescindível estruturar e restringir a entrada de

dados. Assim, consolida-se o argumento da adoção de que um vocabulário controlado, como o AGROVOC proposto neste *framework*, não é apenas uma funcionalidade acessória, mas uma via arquitetural capaz de garantir a qualidade e a verdadeira interoperabilidade terminológica exigida pela integração agrometeorológica.

A transição da aplicação de padrões do domínio puramente geoespacial para o domínio agrícola prático revela lacunas semânticas no nível do *dataset*. O estudo de Dulaney *et al.* (2024), descrevendo o desenvolvimento de um grande arquivo histórico de dados em grade de produtividade agrícola para o Departamento de Agricultura dos Estados Unidos (USDA), exemplifica perfeitamente essa fronteira arquitetural.

Os autores processaram um extenso volume de dados provenientes de monitores de colheita, aplicando rotinas de filtragem e técnicas de interpolação espacial para gerar mapas contínuos de produtividade. Para garantir a padronização e a descoberta desses mapas na base da *National Agricultural Library* (NAL), todos os dados geoespaciais entregues devem aderir ao padrão de metadados ISO 19115.

Contudo, ao analisar o tratamento dado às variáveis agronômicas e de manejo, constata-se a severa limitação imposta por esse alinhamento estritamente espacial. Como a ISO 19115 opera focada no "envelope" cartográfico da informação, ela carece de classes semânticas nativas para representar metadados biológicos e agrícolas específicos.

Consequentemente, para documentar variáveis vitais como tipo e variedade da cultura, população de plantas, datas de plantio e pulverização, Dulaney *et al.* (2024) foram forçados a extrair essas informações dos softwares de manejo agrícola, como o *Farm Works* e o *FarmLogic*, e armazená-las no repositório exclusivamente na forma de arquivos de texto genéricos não estruturados, ou seja, no formato “.txt”.

Esse arranjo arquitetural evidencia de forma irrefutável o problema de interoperabilidade abordado pela presente pesquisa. Quando instituições de referência tentam aplicar padrões no nível do *dataset* em bases de cartografia agrícola, os metadados puramente espaciais, coordenadas, sistemas de referência e limites, ficam bem estruturados, mas as variáveis de manejo, na pesquisa agrometeorológica, acabam desamparadas semanticamente, isoladas em arquivos de texto livre sem qualquer controle ontológico.

Consolida-se, portanto, que a adoção isolada de normas espaciais como a ISO 19115 é estruturalmente insuficiente. Para transcender esse isolamento e garantir a qualidade descritiva das variáveis, é imprescindível cruzar o modelo espacial com os padrões de granularidade biológica e aplicar uma ponte semântica, como o AGROVOC, para normalizar as variáveis de domínio ausentes do escopo geoespacial.

A constatação empírica dos modelos descritivos no nível do *dataset* recorrerem a arquivos de texto plano para registrar variáveis agronômicas encerra a viabilidade de uma integração fundamentada exclusivamente nos padrões das Ciências da Terra.

Uma vez evidenciado que os formatos climáticos e geoespaciais como não conseguem acomodar a granularidade biológica com o rigor semântico necessário, torna-se imprescindível investigar a via metodológica oposta.

A próxima seção analisa, portanto, o estado da arte das aplicações focadas no nível da observação local, explorando como a comunidade agrícola e biológica tem utilizado padrões de granularidade fina, a exemplo do *Darwin Core*, na tentativa de estruturar semanticamente as variáveis em nível de medição, escapando ao escopo global geoespacial.

2.4.2 Aplicações de Padrões em Agrobiodiversidade

A transição da análise dos padrões climáticos e geoespaciais para o domínio da agrobiodiversidade revela que o desafio da interoperabilidade não é exclusividade das Ciências da Terra. Ao focar no nível da observação local, o cenário prático concentra-se em padrões de granularidade fina, tendo o *Darwin Core* (DwC) como o modelo dominante para a estruturação de dados em redes globais de biodiversidade, como o GBIF e o OBIS.

O estudo de Meyer *et al.* (2021) demonstra de forma contundente que o domínio biológico também sofre com a formação de silos de informação estruturais a partir do momento em que a captura de dados se torna mais complexa e detalhada.

Os autores investigam o isolamento gerado pelo advento das tecnologias "ômicas", como a metagenômica, na avaliação da biodiversidade. Para organizar essa nova camada de informações genéticas, a comunidade passou a adotar os *checklists* MIxS (*Minimum Information about any (x) Sequence*), desenvolvidos pelo *Genomic Standard Consortium* (GSC). No entanto, a ausência de um alinhamento nativo entre a estrutura do MIxS e o padrão *Darwin Core* resultou no aprisionamento dos dados, impedindo às observações moleculares e genômicas conversarem de forma fluida com as observações biológicas tradicionais.

Para romper esses silos operacionais, o aprofundamento técnico relatado por Meyer *et al.* (2021) revela que a solução encontrada pela força-tarefa não se limita a um alinhamento puramente estrutural (*cross-walking*).

Para garantir uma interoperabilidade sustentável, os pesquisadores precisaram gerar um mapeamento semântico rigoroso entre as terminologias de ambas as especificações. Esse alinhamento foi formalizado utilizando o padrão SSSOM (*Simple Standard for Sharing*

Ontology Mappings), permitindo o mapeamento sintático e semântico dos valores e a posterior criação de uma extensão MIxS-DwC. Com essa extensão, tornou-se possível incorporar termos do MIxS não mapeados diretamente para dentro do formato de empacotamento *Darwin Core Archive* (DwC-A).

O esforço de integração detalhado por Meyer *et al.* (2021) atua como uma validação inquestionável para a arquitetura proposta na presente pesquisa. O fato da própria comunidade biológica e genômica reconhecer que a integração de dados granulares heterogêneos exige a adoção de mapeamentos semânticos explícitos, como o SSSOM, e de extensões de vocabulário, comprova a insuficiência das abordagens estritamente sintáticas.

Fica evidente a complexidade do nível de observação biológica exigir o apoio de ontologias de domínio. Portanto, se a unificação de variáveis dentro da própria biologia já demanda essas pontes semânticas, o desafio de integrar essa base biológica granular com as matrizes físicas do domínio climático requer, de forma ainda mais crítica, o emprego de um vocabulário centralizado e robusto, consolidando a necessidade de se utilizar o AGROVOC como o alicerce metodológico do *framework* proposto.

A limitação estrutural do *Darwin Core* (DwC) para acomodar variáveis complexas agrava-se substancialmente quando o foco transita da biodiversidade silvestre estrita para a dinâmica dos agroecossistemas. Soares, Saraiva e Drucker (2020) corroboram nessa restrição ao constatar que, embora o DwC seja a linguagem padronizada dominante para o compartilhamento de dados de ocorrência de espécies na natureza e o suporte da maioria dos mais de 1,4 bilhão de registros do GBIF, o padrão nativo carece das classes semânticas para representar a agrobiodiversidade.

As informações para a pesquisa agrícola, como os descritores de cultivo mantidos pela *Bioversity International*, dados de germoplasma e registros do conhecimento de manejo dos agricultores, simplesmente não encontram correspondência na taxonomia do padrão biológico tradicional.

Para mitigar essa profunda incompatibilidade, a pesquisa de Soares, Saraiva e Drucker (2020) propôs o desenvolvimento de uma extensão de metadados para o *Darwin Core*, buscando mapear e incorporar os conceitos agronômicos ausentes por meio da integração de padrões complementares, como as diretrizes da *Global Genome Biodiversity Network* (GGBN) e a extensão *DwC-germplasm*.

O aprofundamento metodológico dessa solução reside na constatação dessa extensão não poder ser apenas estrutural, pois defendem a representação da agrobiodiversidade exigir os

princípios da Web Semântica, projetando o compartilhamento dessas informações na forma de *Linked Data* e aplicando o RDF como modelo de relacionamento de recursos.

Esse esforço de modelagem atua como um endosso direto à arquitetura desenvolvida. A necessidade imperativa de se recorrer ao RDF e ao modelo de *Linked Data* demonstra que os padrões biológicos, isoladamente, são incapazes de representar as interações do ambiente agrícola.

Se a própria comunidade voltada à agrobiodiversidade reconhece a urgência de estabelecer extensões semânticas relacionais para forçar o *Darwin Core* a compreender descritores de cultivo e manejo, torna-se incontestável à integração completa, almejando cruzar essa biologia granular com as matrizes físicas do clima abordadas anteriormente, demanda uma ontologia global e consolidada. A adoção do AGROVOC não é uma escolha arbitrária, mas sim o estado da arte indispensável para atuar como a ponte semântica capaz de normalizar essas variáveis heterogêneas no domínio agrometeorológico.

A transição da necessidade teórica de expansão para a aplicação prática na integração entre ocorrências biológicas e variáveis ambientais expõe de forma ainda mais explícita as limitações semânticas do *Darwin Core*. O estudo de Zárate *et al.* (2020) aborda exatamente essa barreira estrutural ao investigar o cruzamento de dados de biodiversidade do *Global Biodiversity Information Facility* (GBIF) com medições biogeográficas e ambientais do *Ocean Biogeographic Information System* (OBIS). Ao delinearem o escopo da pesquisa, a terminologia nativa do DwC torna-se inadequada para descrever os dados de forma semântica e computacionalmente útil em cenários de integração complexa.

Para viabilizar a convergência entre a observação biológica e o ambiente físico, o aprofundamento metodológico de Zárate *et al.* (2020) consistiu no desenvolvimento do *BiGe-Onto*, um sistema baseado em ontologias transcendendo o encapsulamento em arquivos de texto tradicionais, como os *Darwin Core Archives* (DwC-A).

Os registros textuais foram convertidos para o modelo de grafos da Web Semântica. Contudo, a conversão sintática não foi suficiente. Para o modelo ser processado, cruzamentos espaciais e físicos são necessários para ancorar as classes do *Darwin Core* em ontologias de domínio especializadas externas.

Para isso, adotaram a *Environment Ontology* (EnvO) visando a descrição de biomas, características e materiais do habitat, e implementaram o padrão GeoSPARQL para habilitar a capacidade de raciocínio espacial, permitindo consultas complexas sobre topologias e relacionamentos geométricos.

A arquitetura semântica relatada por Zárate *et al.* (2020) espelha com exatidão o obstáculo central abordado pela presente pesquisa e endossa sua rota tecnológica. A comprovação prática de que a união entre a ocorrência de espécies biológicas e as medições físicas do ambiente, como as variáveis de um corpo d'água ou de clima, exige a conversão dos dados para RDF e a mediação de ontologias externas, validando inquestionavelmente o *framework* proposto nesta pesquisa.

O DwC não se sustenta de forma isolada em integrações multidomínio. Para cruzar a granularidade da biologia agrícola com a estrutura física agrometeorológica, a adoção de uma estrutura relacional fundamentada em uma “ponte”, papel delineado para o AGROVOC, reflete a principal diretriz metodológica do estado da arte. A interoperabilidade granular requer pontes semânticas robustas e direciona a análise para como a comunidade científica vem lidando com desafios adicionais de modelagem.

O esgotamento das abordagens isoladas no nível da observação biológica consolida o diagnóstico central desta pesquisa. A constatação de que o padrão dominante de agrobiodiversidade, o *Darwin Core*, exige mapeamentos semânticos explícitos, modelagem em RDF e a ancoragem em ontologias de domínio externas para conseguir representar o manejo de culturas e as variáveis do ambiente físico comprova que o ecossistema biológico sofre do mesmo engessamento verificado nas Ciências da Terra e do Clima.

Se tanto os padrões espaciais quanto os modelos biológicos de granularidade fina falham estruturalmente ao tentarem cruzar as fronteiras de seus domínios nativos, fica evidente que a verdadeira interoperabilidade agrometeorológica não emerge da simples justaposição de arquivos e formatos.

Diante desse cenário de isolamento, torna-se imprescindível investigar como a literatura recente tem superado essas barreiras. A próxima subseção avança, portanto, para a análise das iniciativas de integração multidomínio, explorando como o estado da arte tem estruturado *frameworks* semânticos complexos e vocabulários controlados para alinhar fluxos de dados radicalmente heterogêneos sob os princípios FAIR.

2.4.3 Iniciativas e Frameworks para Integração de Padrões de Metadados

A superação dos silos de dados e das ilhas de informação descritos nas subseções anteriores exige o desenvolvimento de arquiteturas e *frameworks* modulares capazes de traduzir a semântica de ecossistemas heterogêneos (Meyer *et al.*, 2021; Roussaki *et al.*, 2023). Nos últimos anos, as pesquisas têm propostas diversas abordagens para mitigar a fragmentação da

informação ambiental e agrícola, variando desde a construção de conectores de programação baseados em APIs (Romani *et al.*, 2023) e infraestruturas em nuvem (Brizuela *et al.*, 2024; Knapen *et al.*, 2023), até o emprego avançado de Grafos de Conhecimento e ontologias de domínio (Wu *et al.*, 2024; Alharbi *et al.*, 2024). Apesar disso, a transição da teoria da Web Semântica para a implementação prática e para a verdadeira integração frequentemente esbarra em limitações operacionais, metodológicas ou de escala (Thalhath; Nagamori; Sakaguchi, 2025); Sadeghi *et al.*, 2024).

Para sistematizar a análise dessas abordagens contemporâneas, o Quadro 8 apresenta uma síntese das principais iniciativas e *frameworks* de integração, destacando o foco principal de cada trabalho, a sua respectiva contribuição tecnológica e a lacuna técnica, que impede a plena interoperabilidade no contexto da Agrometeorologia. Na sequência, a partir das lacunas apontadas no quadro, essas iniciativas são analisadas detalhadamente, partindo das abordagens puramente sintáticas e estruturais até chegar às soluções semânticas mais complexas.

Quadro 8 – Síntese das Iniciativas e Frameworks de Integração Analisados

(continua)

Iniciativa / Autores	Foco Principal	Contribuição Tecnológica	Lacuna Específica (Limitação)
Motta <i>et al.</i> (2023)	Alinhamento de telemetria robótica com dados oceanográficos.	Conversão de telemetria bruta para NetCDF e encapsulamento com ISO 19115.	Evidência que a integração com o ambiente físico não ocorre sem forçar a ancoragem das variáveis a um vocabulário externo.
AgroAPI (Romani <i>et al.</i> , 2023)	Integração de dados agrícolas heterogêneos via APIs (Embrapa).	Disponibilização ágil de dados encapsulados no formato textual JSON.	Falta de base semântica nativa e ontologias, dependendo de formatos textuais que não desambigam os conceitos.
BASIN-3D (Varadharajan <i>et al.</i> , 2022)	Plataforma de corretagem (<i>brokering</i>) de dados ambientais.	Síntese de dados heterogêneos orientada por padrões do consórcio OGC	Altíssimo esforço de engenharia, exigindo a criação manual de <i>plugins</i> conectores para quase toda nova fonte inserida.
D4Science (Knapen <i>et al.</i> , 2023)	Ambiente Virtual de Pesquisa (VRE) baseado em nuvem.	Processamento distribuído de matrizes espaço-temporais para simulações de crescimento agrícola.	Falhou na anotação semântica automatizada, gerando perda de dados ao forçar matrizes a se alinharem a formatos sintáticos predefinidos.

(continua)

Iniciativa / Autores	Foco Principal	Contribuição Tecnológica	Lacuna Específica (Limitação)
<i>FAIRagro</i> (Brizuela <i>et al.</i> , 2024)	Federação em nuvem de repositórios agrícolas (Alemanha).	Uso de Objetos Digitais FAIR (FDOs) e JSON-LD para conectar dezenas de repositórios autônomos.	Dependência crítica de curadoria humana exaustiva para o cruzamento de metadados e alinhamento das terminologias divergentes.
BDAKG (Nath <i>et al.</i> , 2024)	Grafo de Conhecimento Agrícola de Bangladesh.	Automação do mapeamento conceitual usando a ontologia QB4OLAP em cubos analíticos.	O processo OLAP (<i>Online Analytical Processing</i>) agrega as informações estatisticamente, sacrificando a resolução fina para medições contínuas de sensores.
Martínez <i>et al.</i> (2025)	Modelo conceitual em 4 camadas para dados ambientais.	Isola a modelagem semântica (catálogo externo) do armazenamento físico das matrizes de dados.	Alto custo de design manual por especialistas e escassez de mecanismos para lidar com o versionamento contínuo das bases.
EDAM (Samourkasidis; Athanasiadis, 2020)	Fusão de séries temporais para modelagem de culturas.	Usa motor de inferência lógica para identificar equivalências observacionais em tempo de execução.	Opera com ontologias locais customizadas, exigindo que analistas redijam manualmente anotações textuais para cada novo formato.
STSDaMaS (San Emeterio <i>et al.</i> , 2023)	Processamento de alto volume de dados dinâmicos de Sensores/IoT.	Separa medições em banco de séries temporais (TSDB) da semântica em repositório de triplas (SOSA).	Depende do módulo tradutor IDI (<i>wrapper</i>), que exige o mapeamento prévio de conversão para cada plataforma externa que adentra o sistema.
RESEO (Aldana-Martín <i>et al.</i> , 2022)	Arquitetura de dados para observação da Terra (Satélite/Clima).	Emprega a Web Semântica (RDF, SWRL) para cruzar índices de vegetação com condições climáticas	O uso de ontologias genéricas (SWEET, SSN) obriga o acoplamento de motores de tradução forçada entre as matrizes e o modelo de grafos.
<i>LinkedGeoClimate</i> (Wu <i>et al.</i> , 2024)	Plataforma de orquestração de matrizes de dados climáticos e locais.	Resolve a interoperabilidade sem migração física unindo RDF e GeoSPARQL.	Exige que analistas escrevam consultas lógicas complexas e utiliza ontologias genéricas sem aprofundamento agrônomo.
Beretta <i>et al.</i> (2021)	Modelo de metadados semântico centrado no usuário.	Isola a amostra física da característica final de interesse usando SOSA e o AGROVOC.	Alto custo operacional de anotação manual e obrigatoriedade contínua de manter mapeamentos entre várias ontologias fragmentadas.

(conclusão)

Iniciativa / Autores	Foco Principal	Contribuição Tecnológica	Lacuna Específica (Limitação)
AgrODSS (Alharbi <i>et al.</i> , 2024)	Suporte à decisão para identificação de pragas e doenças.	Integra dados de pragas, clima e manejo usando a linguagem OWL e AGROVOC com inferência lógica.	Aquisição de conhecimento e instanciação no grafo feitas de forma inteiramente manual, o que limitou a pesquisa a uma única cultura.
Zeginis <i>et al.</i> (2024)	Meta-modelo para agricultura e pecuária de precisão.	Unifica DCAT, RDF Data Cube (QB), PROV-O, e as ontologias AGROVOC e QUDT.	Arquitetura rígida com alto nível de acoplamento técnico, dependente de restrições complexas baseadas em formas SHACL.
Goldstein <i>et al.</i> (2022)	Framework (DSS) em nuvem para controle de pragas agrícolas.	Uso do vocabulário AGROVOC no repositório de biologia (<i>triplestore</i> consultável via SPARQL).	Retrocesso metodológico e fragmentação ao lidar com o clima por meio de bancos relacionais (SQL) e rotinas de ETL inflexíveis.
Link-Climate (Wu <i>et al.</i> , 2022)	Integração de grafos climáticos e ferramentas de Machine Learning.	Fluxo de trabalho que conecta a base semântica aos ecossistemas de programação em Python.	“Achata” os grafos de volta para matrizes tabulares 2D para a análise, destruindo a topologia e as regras de inferência originais.

Fonte: O Autor.

A partir das limitações evidenciadas no Quadro 8, a análise inicia-se pelas abordagens fundamentadas na integração sintática e em formatos físicos. Nesse escopo inicial, destacam-se as iniciativas e *frameworks* projetados para harmonizar os registros brutos gerados por maquinário autônomo com as observações do ambiente físico onde operam. O estudo de Motta *et al.* (2023) ilustra perfeitamente essa transição tecnológica ao propor um *framework* para alinhar a telemetria bruta de veículos robóticos marinhos com os padrões tradicionais da oceanografia observacional.

Convencionalmente, os dados das missões robóticas eram exportados em formatos textuais simples, como arquivos em formato “.csv”, muitas vezes sem diretrizes suficientes e sem vocabulário compartilhado. A ausência de metadados padronizados impossibilitava o cruzamento automatizado do comportamento da máquina com as variáveis climáticas e ambientais coletadas simultaneamente.

Para renderizar esses conjuntos de dados acoplados em estrita conformidade com os princípios FAIR, a arquitetura desenvolvida por Motta *et al.* (2023) estabeleceu uma esteira de processamento abandonando o isolamento do formato tabular e textual bruto. O *framework* converte automaticamente a telemetria para o formato de cubo de dados *NetCDF*,

encapsulando-o com os padrões de metadados globais do ISO 19115 e da convenção de descoberta ACDD (*Attribute Convention for Data Discovery*).

Para garantir a interoperabilidade das variáveis de controle do robô, os autores precisaram ancorá-las estruturalmente em um vocabulário controlado externo. Para isso, adotaram a Convenção *Climate and Forecast* (CF) e realizaram um rigoroso mapeamento semântico, substituindo os nomes de *log* proprietários, como as variáveis de guinada e inclinação da máquina, por nomes de variáveis padronizados, anotados semanticamente de forma a respeitar a dicotomia robótica-ambiental.

A estratégia de modelagem detalhada por Motta *et al.* (2023) corrobora frontalmente a abordagem proposta na presente pesquisa. O sucesso da integração entre a Robótica e a Oceanografia dependeu intrinsecamente do uso de uma convenção de vocabulário comum para normalizar as variáveis de ambos os escopos heterogêneos.

Transpondo essa validação para o contexto agrícola, consolida-se a premissa da integração dos dados de operação de maquinário, telemetria e manejo com as matrizes agrometeorológicas não pode prescindir de um esforço semelhante de ancoragem e alinhamento de esquemas.

Os *frameworks* de integração semântica modernos exigem o acoplamento de vocabulários controlados robustos, justificando metodologicamente a adoção do vocabulário AGROVOC como o eixo central capaz de traduzir a informação da observação agrícola para a realidade climática.

Apesar da comprovada necessidade de ancoragem em vocabulários controlados, grande parte das iniciativas institucionais ainda tenta superar os silos de dados concentrando esforços estritamente na camada de acesso sintático. Essa transição da governança de dados isolados para ecossistemas de inovação centralizados baseados em serviços web encontra forte respaldo na realidade agrícola brasileira.

Romani *et al.* (2023) corrobora essa tendência ao apresentar a plataforma AgroAPI, desenvolvida pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), com o objetivo de fornecer informações heterogêneas para o setor do agronegócio. O sistema disponibiliza uma série de interfaces de programação de aplicações (APIs) voltadas para o cruzamento de variáveis climáticas, classificação de perfis de solo, monitoramento de índices de vegetação e catálogos de insumos biológicos.

A arquitetura da plataforma, segundo os autores, fundamenta-se na extração de dados de bancos internos estruturados e na posterior entrega dessas respostas encapsuladas no formato

textual JSON, operando como uma camada intermediária, um *middleware*, para facilitar o desenvolvimento de soluções digitais.

Embora a adoção de plataformas baseadas em APIs represente um salto significativo na disponibilidade sintática da informação, a concepção do AgroAPI evidencia as barreiras tecnológicas inerentes a abordagens, as quais não adotam a interoperabilidade semântica de forma nativa.

A entrega de dados por meio de estruturas JSON resolve a comunicação física entre o servidor e os sistemas consumidores, mas não desambigua os conceitos científicos tampouco incorpora o significado relacional das variáveis.

Romani *et al.* (2023) reconhecem essa fronteira metodológica estrutural ao analisarem os desafios de lidar com a ampla diversidade de conceitos embutidos nos diferentes serviços agrícolas, onde a arquitetura atual carece de uma base de representação, além de afirmar a necessidade futura de incorporar ontologias e taxonomias agrícolas para prover o real conhecimento de domínio à plataforma.

Essa constatação empírica por parte de uma das maiores instituições de pesquisa agrícola do mundo atua como uma chancela inequívoca para o problema investigado na presente pesquisa.

A admissão de que a simples disponibilização de APIs modernas e formatos como o JSON não substitui a necessidade de ontologias comprova que o ecossistema agrícola demanda uma nova etapa de maturidade na engenharia de dados.

A dependência contínua de integrações puramente textuais perpetua o gargalo de exigência de interpretação humana na ponta do usuário. Diante de um cenário onde os pesquisadores e os consórcios institucionais reconhecem essa limitação latente, a comunidade científica tem intensificado a proposição de arquiteturas alternativas, buscando suprir essa lacuna e garantir uma representação orgânica da informação, explorando novos modelos computacionais capazes de integrar nativamente a semântica de domínio ao ciclo de vida dos dados.

A ausência dessa integração semântica nativa e a consequente dependência de esforços contínuos de engenharia de software evidenciam-se também em grandes infraestruturas de arquiteturas descentralizadas voltadas à integração ambiental e climática.

Varadharajan *et al.* (2022) exemplificam essa barreira estrutural ao apresentar o BASIN-3D, uma plataforma projetada para recuperar e sintetizar dados heterogêneos de maneira dinâmica, abrangendo desde observações hidrológicas até séries temporais meteorológicas de estações de campo.

Para orquestrar fontes com linguagens e estruturas distintas, a arquitetura afasta-se de um banco de dados centralizado e adota um modelo conceitual baseado nos padrões abertos do consórcio OGC, como o *Observations and Measurements*. Contudo, a comunicação entre os repositórios originais e a camada de síntese do sistema demanda a implementação de conectores dedicados, operando através de uma arquitetura baseada em *plugins* customizados.

Embora o modelo de arquitetura descentralizada confira flexibilidade à integração, a análise técnica desta abordagem expõe o esgotamento prático da interoperabilidade baseada em adaptadores externos. Varadharajan *et al.* (2022) atestam que a diversidade semântica e sintática dos provedores de dados exige a construção de um *plugin* de acesso exclusivo para praticamente cada nova fonte inserida no ecossistema.

O desenvolvimento e a manutenção desses conectores são processos morosos, exigindo tanto conhecimentos avançados em engenharia de software quanto o julgamento empírico de especialistas para mapear manualmente os termos e variáveis das planilhas para a estrutura central.

Essa constatação demonstra que as plataformas de arquiteturas descentralizadas de larga escala esbarram em um teto de escalabilidade operacional imposto pelo contínuo esforço de tradução algorítmica, evidenciando a necessidade de abordagens centralizadas da semântica de forma nativa para contornar o desenvolvimento contínuo de novos conectores.

Na tentativa de contornar as limitações de escalabilidade dos conectores locais, a necessidade de orquestrar fluxos contínuos de dados tem impulsionado a adoção de Ambientes Virtuais de Pesquisa (VREs) baseados em nuvem, como a plataforma D4Science proposta por Knapen *et al.* (2023). Essa infraestrutura visa executar fluxos de modelagem agrometeorológica para lidar com a escala massiva de informações e superar a limitação de análises puramente estatísticas.

Para isso, os autores implementaram casos de uso cruzando dados climáticos diários, imagens de satélite e características de solo para calcular a fenologia das culturas e processar simulações de crescimento em nível de parcela agrícola. Knapen *et al.* (2023) detalham a arquitetura se apoiando em processamento distribuído e interfaces de programação (APIs) para extrair matrizes de dados espaço-temporais de repositórios dinâmicos e alimentar algoritmos analíticos complexos de forma colaborativa.

A avaliação prática desta robusta infraestrutura revela que a mera provisão de poder computacional e a centralização em nuvem não resolve, por si só, o gargalo da heterogeneidade da informação.

Knapen *et al.* (2023) relatam que, embora a plataforma ofereça um conjunto de ferramentas semânticas, a integração desses recursos no fluxo de trabalho demonstrou-se insuficiente, tendo em vista que a arquitetura falhou em permitir a anotação semântica automatizada dos dados gerados utilizando vocabulários e ontologias, comprometendo severamente o alcance da interoperabilidade e, também, a *FAIRness*.

Além disso, os autores admitem que a tentativa de forçar a integração de matrizes heterogêneas, como observações da terra e registros biológicos agrícolas em processos operacionais genéricos, frequentemente resulta em perda de dados, pois o sistema só consegue ingerir de forma confiável as informações que se alinham aos formatos sintáticos predefinidos.

Essa constatação empírica de falha na camada de interoperabilidade, oriunda de uma plataforma em nuvem de alto nível, atesta que a justaposição de ferramentas de processamento não substitui uma governança de dados estruturada desde a sua concepção.

O reconhecimento documentado da ligação semântica de dados agrícolas permanece como um desafio complexo e laborioso, mesmo em ambientes virtuais avançados dotados de alto investimento tecnológico, solidificando a tese dos vocabulários de domínio não atuarem como acessórios periféricos do sistema.

Tendo em vista esse diagnóstico contundente do estado da arte, aponta-se para a urgência de superar a dependência de plataformas integradoras genéricas, negligenciando o significado nativo das variáveis e demandando a construção de arquiteturas especializadas onde a semântica atue de ponta a ponta como o motor central da interoperabilidade dos fluxos.

Diante disso, a busca por soluções automatizadoras da interoperabilidade de dados heterogêneos em larga escala tem mobilizado consórcios de pesquisa. O trabalho de Brizuela *et al.* (2024) documenta essa complexidade ao apresentar o desenvolvimento do *middleware* federado do *FAIRagro*, uma infraestrutura projetada para unificar dezenas de repositórios autônomos de dados do ecossistema agrícola na Alemanha.

Para integrar bases de estudos físicos de solo até fenótipos e dados ambientais, os autores propõem uma arquitetura baseada na extração de metadados no formato de grafos JSON-LD, utilizando vocabulários estruturais flexíveis como *Schema.org* e *Bioschemas*.

Brizuela *et al.* (2024) evidenciam a estratégia metodológica avançando para o encapsulamento dessas informações em Objetos Digitais FAIR (FDOs) e operando sobre repositórios de versionamento no GitLab para rastrear a proveniência dos dados, visando facilitar a federação entre as diferentes infraestruturas de pesquisa sem violar sua autonomia.

Apesar de representar uma das mais robustas iniciativas contemporâneas de governança de dados agrícolas, a concepção prática deste *middleware* apresenta as limitações semânticas

que ainda afligem o estado da arte. Brizuela *et al.* (2024) reconhecem que uma integração estreita e não supervisionada de bases de dados radicalmente heterogêneas permanece inviável. Para alcançar a harmonização estrutural, o *framework* FAIRagro depende fortemente de processos semiautomatizados e de curadoria manual detalhada.

O estudo relata o maior desafio arquitetural do projeto ser a obrigatoriedade de construir e manter exaustivos cruzamentos de metadados guiados por especialistas de domínio, a fim de forçar o alinhamento entre as terminologias divergentes de cada repositório.

Essa dependência contínua de mapeamentos manuais e de curadoria humana para traduzir variáveis, mesmo em arquiteturas de federação de ponta, ratifica a dificuldade intrínseca da integração de dados heterogêneos.

A constatação dos consórcios de grande escala e com alto investimento tecnológico ainda esbarrarem na pesada barreira da tradução semântica supervisionada reforça a necessidade de buscar abordagens minimizadoras dessa intervenção humana e centralizadoras do conhecimento em ontologias de domínio sólidas.

A persistência desse gargalo operacional impulsiona a análise de outras arquiteturas recentes, buscando investigar como modelos emergentes tentam otimizar a governança e a automação desses ecossistemas complexos.

Nesse contexto, Nath *et al.* (2024) propõem a busca por métodos automatizados para a integração de variáveis de campo ocorrendo com o desenvolvimento de processos semânticos de Extração, Transformação e Carga através de um *framework* para gerar o Grafo de Conhecimento Agrícola de Bangladesh (BDACKG).

Para isso, os autores desenvolveram uma tubulação automatizadora do mapeamento conceitual de planilhas e relatórios governamentais, trazendo como a grande inovação metodológica a fusão de tecnologias da Web Semântica com a análise multidimensional, em que o sistema emprega a ontologia QB4OLAP para estruturar os dados da biologia das culturas, da silvicultura e de variáveis geográficas na forma de cubos analíticos.

Para Nath *et al.* (2024), essa modelagem garante as instâncias geradas automaticamente herdarem uma estrutura dimensional hierárquica, possibilitando a execução de consultas transversais complexas e a ligação automatizada a repositórios externos, como a *CropOntology*.

Entretanto, uma análise rigorosa do *framework* mostra que a automação estrutural alcançada ainda apresenta restrições incompatíveis com o dinamismo tecnológico da agricultura de precisão.

Nath *et al.* (2024) evidenciam um severo gargalo de pré-processamento, no qual a fase de extração e limpeza das informações consumiu mais de 70% do tempo total da arquitetura,

atestando que a dependência de fontes textuais estáticas, como arquivos PDF ou planilhas desconexas, continuam cobrando um alto custo de intervenção prévia.

Ao adotar a modelagem em cubos de dados, o sistema direciona-se intrinsecamente para a agregação estatística em larga escala, limitando a análise ao somatório anual e ao agrupamento distrital das safras. Conforme os autores, essa consolidação histórica suprime o nível de detalhe exigido para a interoperabilidade contínua das lavouras modernas, apresentando limitações estruturais para acomodar as medições intradiárias de sensores agrometeorológicos e a telemetria em frequência do maquinário.

Desse modo, a constatação de que a automação semântica atual ainda sacrifica a granularidade da informação em favor da totalização estatística ratifica a demanda por modelos atuantes na raiz da observação isolada, impulsionando a busca por arquiteturas modulares capazes de isolar a modelagem semântica do armazenamento físico, a fim de preservar a resolução espacial e temporal dos dados.

Como uma resposta direta a essa demanda por arquiteturas modulares, o estudo de Martínez *et al.* (2025) desenvolve um modelo conceitual em quatro camadas para o tratamento de informações ambientais. Para solucionar a ineficiência de modelos universais de armazenamento e o isolamento crônico de modelos criados *ad hoc*, os autores propõem isolar estruturalmente a modelagem semântica da persistência física do dado. Com isso, em vez de embutir a semântica das observações de forma rígida nas tabelas físicas do banco de dados, o sistema ancora os metadados e as propriedades em vocabulários controlados padronizados, como a convenção CF, passando a ser gerenciados externamente pelo catálogo de alto nível.

Dessa forma, Martínez *et al.* (2025) comprovam que, nesta arquitetura, a verdadeira interoperabilidade ocorre na governança semântica e não no mero armazenamento do dado bruto. Todavia, identificam limitações referente à integração entre áreas heterogêneas. Essa abordagem cobra um alto custo operacional na prática, exigindo um esforço de design manual extenso por parte dos especialistas para instanciar as aplicações finais.

Além disso, a arquitetura apresenta dificuldades para lidar com o ciclo de vida da informação, carecendo de mecanismos maduros para suportar mudanças de esquema e o versionamento contínuo dos bancos de dados.

A literatura atual ainda sofre com a rigidez hierárquica e com a dependência de instanciação manual, reforçando a necessidade de evoluir para ontologias mais unificadas. Diante desse cenário metodológico, pesquisas têm direcionado esforços para arquiteturas contornando essas limitações de automação e aplicando soluções de interoperabilidade focadas

em fluxos de dados ainda mais complexos e diversos, como as séries temporais voltadas à modelagem de culturas agrícolas.

Assim, o estudo de Samourkasidis e Athanasiadis (2020) apresenta o módulo EDAM, uma arquitetura projetada para a fusão e transformação de séries temporais ambientais destinadas a suprir ecossistemas agrícolas consolidados, como os modelos APSIM (*Agricultural Production Systems sIMulator*) e DSSAT (*Decision Support System for Agrotechnology Transfer*).

Para orquestrar a heterogeneidade sintática e semântica das fontes climáticas, a plataforma proposta pelos autores emprega um motor de inferência lógica, processando arquivos de metadados e gabaritos textuais, viabilizando a identificação de equivalências observacionais e a conversão matemática de unidades de medida em tempo de execução.

No campo das limitações, Samourkasidis e Athanasiadis (2020) reforçam que o sistema opera fundamentado exclusivamente em uma ontologia local e customizada. Ao prescindir da adoção de um vocabulário de domínio de alcance global, a arquitetura transfere o esforço de integração para a curadoria humana, exigindo que os analistas redijam manualmente as anotações e os arquivos de estruturação para cada novo formato de dado a ser ingerido. Os próprios autores da plataforma afirmam que garantir a coerência e a exatidão dessas anotações empíricas constitui um desafio prático significativo, ressaltando que a ferramenta não foi projetada para importar ou reutilizar ontologias externas consolidadas de forma nativa.

A constatação de que as plataformas concebidas especificamente para a modelagem agrícola permanecem reféns de dicionários locais isolados e intervenções manuais exaustivas reforça que o esgotamento desse modelo de engenharia instiga a busca por abordagens arquiteturais com padrão ontológico universal e atua como o elo de interoperabilidade definitivo, suprimindo o uso de traduções locais.

Nesse cenário, o desafio se torna ainda mais crítico quando a integração evolui para o monitoramento contínuo e o processamento em tempo real de sensores e da Internet das Coisas (IoT). Para atender a essa demanda, San Emeterio de la Parte *et al.* (2023) apresentam o *framework* STSDaMaS, uma arquitetura concebida para gerenciar o imenso volume de dados dinâmicos gerados por sensores e telemetria no campo.

Para assegurar a semântica atuando como motor central sem comprometer o desempenho de leitura e orquestração dos dados, foi proposta uma divisão inteligente do repositório, com as medições ambientais de frequência sendo inseridas por um banco de dados não relacional orientado a séries temporais, enquanto a inteligência do domínio e as descrições dos dispositivos operam nativamente em um repositório de triplas consultável via SPARQL.

Com essa modelagem, San Emeterio de la Parte *et al.* (2023) comprovam a possibilidade de integrar dados heterogêneos de granularidade espaço-temporal mantendo o rigor das tecnologias de conhecimento em todo o ciclo de vida da informação.

No entanto, para viabilizar a interoperabilidade com infraestruturas externas, sem compartilhamento do seu modelo ontológico nativo, fundamentado nas ontologias genéricas de sensoriamento SOSA e SSN, San Emeterio de la Parte *et al.* (2023) precisaram desenvolver um componente tradutor denominado *Interoperability through Data Integration* (IDI).

Esse módulo atua como um invólucro, dependendo do estabelecimento prévio de grafos de mapeamento e regras exaustivas de correlação de propriedades para converter e forçar a adequação dos dados recebidos ao modelo interno do sistema. Essa dependência de módulos de tradução expõe que a arquitetura não resolve o problema raiz da fragmentação terminológica.

A obrigatoriedade de projetar e manter chaves de conversão para cada nova plataforma que adentra no ecossistema perpetua o esforço tecnológico e a intervenção manual, reafirmando a necessidade premente de uma ontologia unificadora de domínio atuante como eixo global e elimine a urgência de ferramentas de tradução intermediária nas camadas de interoperabilidade.

A dependência de módulos de tradução também se manifesta no estudo de Aldana-Martín *et al.* (2022), os quais propuseram a ontologia RESEO, uma arquitetura desenhada para consolidar dados de observação da Terra, cruzando imagens e produtos do satélite Sentinel-2 com variáveis meteorológicas extraídas da agência espanhola AEMET (*Agencia Estatal de Meteorología*).

O modelo fundamenta-se inteiramente em tecnologias da Web Semântica, aplicando repositórios RDF e regras lógicas, como a linguagem SWRL, para relacionar índices de vegetação estruturados a partir das matrizes de *pixels* com as condições climáticas de uma determinada região.

Contudo, para orquestrar essas diferentes fontes espaciais, Aldana-Martín *et al.* (2022) recorrem ao reuso de ontologias de propósito geral, adotando a SSN (*Semantic Sensor Network*) para a descrição técnica dos sensores e a SWEET (*Semantic Web for Earth and Environmental Terminology*) para modelar a terminologia ambiental ampla. Ao operar sem um vocabulário restrito e nativo ao ecossistema agrônomo, o *framework* esbarra no mesmo gargalo operacional daquele proposto por San Emeterio de la Parte *et al.* (2023), com uma arquitetura exigindo o acoplamento de um motor de funções de mapeamento.

Esse módulo atua de forma dedicada na tradução forçada das matrizes e pacotes originais para o modelo de grafos. A recorrência estrutural desses motores de conversão intermediária nas diferentes camadas da observação agrícola comprova que o emprego de

ontologias genéricas perpetua o esforço computacional e impõe um teto de escalabilidade técnica, impulsionando a comunidade científica a buscar modelos arquiteturais onde uma ontologia central elimine a necessidade estrutural dessas pontes de tradução e viabilize a orquestração dinâmica da informação diretamente nos grafos.

A materialização dessa orquestração dinâmica da informação diretamente nos grafos de conhecimento é encontrada no estudo de Wu *et al.* (2024), os quais propuseram a plataforma *LinkedGeoClimate*, projetada para superar o isolamento estrutural imposto pelos sistemas de gerenciamento de bancos de dados relacionais tradicionais na disponibilização de dados climáticos.

Para viabilizar a análise conjunta entre a meteorologia e o contexto ambiental e espacial, os autores desenvolveram uma arquitetura utilizando o modelo RDF para transcender esquemas rígidos, orquestrando as matrizes de dados por meio de consultas em SPARQL. O trunfo analítico desta plataforma reside, segundo Wu *et al.* (2024), na sua capacidade de cruzar, em tempo de execução, medições de estações climáticas com características de repositórios externos, como perfis de solo, mapeamento de cobertura da terra e a localização de propriedades rurais, apoiando-se nos padrões de representação da norma GeoSPARQL. Essa modelagem permite ao contexto geográfico atuar como uma âncora natural para a observação climática, resolvendo a interoperabilidade de dados heterogêneos na camada semântica sem a obrigatoriedade de migração física e centralizada das bases originais.

Embora a aplicação de consultas SPARQL em grafos de conhecimento solucione o isolamento sintático, o *framework* avaliado impõe barreiras metodológicas e operacionais significativas para o ecossistema agrícola de precisão. Wu *et al.* (2024) ressaltam que a arquitetura possui severas limitações técnicas para ingerir dados no formato matricial, como as imagens de satélite, recursos imprescindíveis para o monitoramento contínuo das lavouras.

Somado a isso, a plataforma delega o ônus do cruzamento da informação à interface de consumo, exigindo que os analistas e pesquisadores dominem a elaboração de roteiros textuais e lógicos complexos, como consultas nativas SPARQL, para acessar as observações.

Sob a ótica da representação do conhecimento, constata-se que o sistema se baseia em ontologias genéricas de sensoriamento, prescindindo de vocabulários terminológicos aprofundados para as variáveis agronômicas e biológicas, fragilizando a descrição do manejo e da fenologia.

A constatação de que as arquiteturas semânticas contemporâneas ainda esbarram na barreira da usabilidade e na exclusão de formatos sinaliza que a literatura continua em busca de *frameworks* que unam a governança robusta dos metadados a uma experiência de integração

transparente para o domínio final, impulsionando a investigação de novas arquiteturas centradas no usuário e ancoradas em vocabulários de domínio específicos, como o AGROVOC, para aproximar a observação física do real interesse agrônomo.

Para contornar a rigidez arquitetural e o alto acoplamento gerado pela tentativa de fundir diferentes catálogos estruturais, a integração mais fluida de dados heterogêneos ocorre quando a modelagem desce ao nível da informação, o paradigma da observação.

Beretta *et al.* (2021) demonstram essa mudança de perspectiva ao proporem um modelo de metadados centrado no usuário, fundamentado na ontologia leve SOSA (*Sensor, Observation, Sample, and Actuator*). Ao diagnosticarem o estado da arte, os autores corroboram diretamente as críticas levantadas ao longo desta pesquisa, atestando que, padrões dependentes de domínio, como o Padrão ISO 19115 nas Ciências da Terra e o EML (*Ecological Metadata Language*) na Biologia, adotam estritamente a visão técnica do produtor do dado.

Essa abordagem restringe a descoberta e o reuso das informações por pesquisadores de outras áreas, onde a semântica permanece encapsulada em jargões do domínio de origem. Para romper esse isolamento disciplinar, a arquitetura de Beretta *et al.* (2021) modela as matrizes de dados separando a entidade física medida da característica temática final de interesse.

Essa distinção permite propriedades complexas serem desmembradas em conceitos atômicos, possibilitando um único conjunto de dados ser descoberto e consumido por múltiplos campos científicos simultaneamente.

A convergência desta abordagem com a metodologia da presente pesquisa torna-se inquestionável na aplicação prática do modelo. Para viabilizar a interoperabilidade real entre variáveis díspares, cruzando, em seu caso de estudo, dados da biologia de ecossistemas com medições do ambiente físico e hidrológico, o *framework* delega a representação semântica dos temas observados a ontologias externas, utilizando nominalmente o AGROVOC como uma de suas bases terminológicas para harmonizar as anotações.

A despeito de sua sofisticação semântica, Beretta *et al.* (2021) admitem limitações pragmáticas substanciais, afirmando que a arquitetura cobra um custo operacional altíssimo e exige um esforço expressivo de anotação e curadoria manual por parte dos usuários para descrever os conjuntos de dados em conformidade com as ontologias do sistema. Como diferentes comunidades científicas preferem adotar vocabulários distintos, o *framework* permanece refém da necessidade constante de construir e manter mapeamentos complexos entre essas ontologias fragmentadas para assegurar a interoperabilidade contínua.

O obstáculo de anotação manual também compromete a escalabilidade em sistemas avançados de suporte à decisão na agricultura, como o AgrODSS, arquitetura proposta por

Alharbi *et al.* (2024), projetada para gerenciar dados heterogêneos e orientar a identificação de pragas e doenças agronômicas.

Para processar informações não estruturadas e convertê-las em recomendações rastreáveis, o *framework* acopla um modelo de explicação fundamentado em evidências a uma ontologia de domínio estruturada em OWL. O mérito tecnológico dessa abordagem reside, segundo os autores, na capacidade de utilizar motores de inferência lógica para cruzar variáveis do ambiente físico, como temperatura e umidade, dados da biologia e práticas de manejo dentro de um Grafo de Conhecimento unificado, superando o cruzamento sintático tradicional.

Ao fundamentar a concepção do modelo, Alharbi *et al.* (2024) atestam o estado da arte referenciando explicitamente o AGROVOC como o vocabulário controlado mais abrangente do setor, ratificando a consolidação desta ferramenta global para a estruturação do domínio agrícola.

Entretanto, a aplicação prática desta arquitetura de suporte à decisão evidencia o persistente desafio da automação na engenharia de dados semânticos. Alharbi *et al.* (2024) diagnosticam as estruturas terminológicas puramente hierárquicas exigindo adaptações sintáticas complexas para alimentar regras restritivas e permitir inferências lógicas pelas máquinas.

A arquitetura possui limitação em relação à aquisição do conhecimento biológico e climático, bem como à instanciação das variáveis no grafo, executadas de forma inteiramente manual a partir da extração de documentos e textos não estruturados. Essa dependência contínua da curadoria humana restringiu o *framework* à análise isolada de uma única cultura, inviabilizando a sua escalabilidade orgânica para outros cenários agrometeorológicos sem a futura adoção de extratores baseados em processamento de linguagem natural.

Os sistemas semânticos mais modernos ainda sofrem com a inflexibilidade da alimentação manual dos dados evidenciada por uma lacuna estrutural persistente, exigindo inovações arquiteturais e estabelecendo métodos mais fluidos e automatizados para a integração contínua das variáveis de campo.

Nesse sentido, Zeginis *et al.* (2024) propõem um meta-modelo semântico projetado para unificar fluxos de dados radicalmente heterogêneos. Ao analisarem casos de uso, exigiam o cruzamento de variáveis climáticas, dados de maquinário e previsões biológicas fenológicas. Houve constatação da barreira estrutural crônica da área, a inexistência de um modelo padronizado único capaz de endereçar todas as facetas do dado agronômico.

Para solucionar essa lacuna, o *framework* proposto por Zeginis *et al.* (2024) abandona a dependência de um formato isolado e implementa um perfil de aplicação fundamentado em Padrões W3C.

A arquitetura funde o catálogo de metadados DCAT com o vocabulário *RDF Data Cube* (QB) para definir as dimensões e as medidas estruturais das matrizes, acoplando a ontologia PROV-O para rastrear a proveniência e o ciclo de vida das informações de origem.

O grande trunfo metodológico deste modelo, converge de forma decisiva para os objetivos da presente pesquisa, residindo na adoção rigorosa de vocabulários controlados globais para garantir a interoperabilidade.

Zeginis *et al.* (2024) aplicam diretamente a ontologia AGROVOC para harmonizar a semântica dos domínios agrícolas investigados e recorrem à ontologia QUDT (*Quantities, Units, Dimensions and Data Types*) para normalizar fisicamente as métricas observadas nos *datasets*. Ainda assim, a fusão arquitetural de múltiplos padrões macroscópicos impõe restrições técnicas.

Para forçar o DCAT a aceitar a estrutura de cubos de dados do QB sem quebrar as regras sintáticas originais dos modelos, o *framework* exige a construção de complexas restrições baseadas em formas SHACL (*Shapes Constraint Language*), gerando um alto nível de acoplamento técnico.

A formulação e as limitações deste meta-modelo mostram a integração de dados na agrometeorologia requisitando o apoio em vocabulários controlados como o AGROVOC, mas a simples tentativa de "costurar" padrões distintos por meio de "gambiarras sintáticas" ou restrições complexas perpetua um modelo de difícil manutenção e escalabilidade.

Consolida-se, portanto, a necessidade de uma abordagem de integração coesa, com o vocabulário de domínio atuando diretamente como a espinha dorsal da interoperabilidade, mitigando a rigidez e a complexidade arquitetural verificada nas soluções atuais.

Com o intuito de tentar sanar essa complexidade arquitetural, Goldstein *et al.* (2022) propõem um *framework* estruturado em quatro camadas para a integração e análise de dados agrícolas heterogêneos. Para tal, desenvolveram uma arquitetura abrangendo desde a aquisição da informação física até a camada de aplicações para o usuário final.

O grande avanço metodológico desse estudo é evidenciado, conforme Goldstein *et al.* (2022), na estruturação do Sistema de Suporte à Decisão (DSS) para o controle de pragas, para contornar a incompatibilidade terminológica entre diferentes legislações e jargões agrícolas, o sistema implementa repositório de triplas (*triplestore*) consultável via SPARQL, adotando

vocabulários como o AGROVOC, como base de conhecimento para integrar de forma automatizada conceitos de culturas e defensivos biológicos distribuídos na web.

Goldstein *et al.* (2022) revelam uma fragmentação metodológica expondo as limitações operacionais dos *frameworks* contemporâneos. Embora a arquitetura tenha demonstrado sucesso ao empregar o AGROVOC e o modelo RDF para solucionar o isolamento da observação biológica, o *framework* falha em sustentar esse rigor semântico ao tratar as demais matrizes ambientais.

Para o processamento de variáveis espaciais, como a distribuição geográfica de pragas, e agrometeorológicas, como dados de clima e sensores de solo para recomendação de irrigação, Goldstein *et al.* (2022) retrocedem para a adoção de bancos de dados relacionais em SQL e dependem de processos de extração e transformação (ETL) desenvolvidos por meio de scripts proprietários.

Essa abordagem dicotômica indica as ontologias aplicadas como uma ferramenta utilitária para uma aplicação específica, e não como o alicerce unificador de todo o sistema. O uso de scripts rígidos de conversão para forçar o cruzamento entre os registros biológicos e as variáveis físicas perpetua a fragilidade do modelo diante de mudanças na estrutura da informação observada.

Essa vulnerabilidade se agrava ainda mais quando essas arquiteturas tentam se integrar a fluxos analíticos avançados e de aprendizado de máquina, como no estudo de Wu *et al.* (2022), que investiga a desconexão estrutural entre os Grafos de Conhecimento climáticos e as ferramentas tradicionais de ciência de dados. Para viabilizar a orquestração entre esses ambientes, Wu *et al.* (2022) propuseram um fluxo de trabalho conectando a plataforma semântica *Link-Climate* aos ecossistemas de programação em Python.

Ao estruturar os registros meteorológicos sob a ontologia genérica de sensoriamento SOSA, o sistema permite que consultas federadas enriqueçam bases locais estáticas com novas variáveis, resultando em um ganho preditivo de até 10% em modelos de classificação climática.

No entanto, uma análise aprofundada desta infraestrutura de consumo revela uma regressão metodológica na manutenção da representatividade do conhecimento. Para garantir a compatibilidade com os algoritmos de predição convencionais, a arquitetura de Wu *et al.* (2022) implementa módulos de conversão, como a biblioteca *kgextension*, projetados com o propósito estrito de traduzir e “achatar” as respostas em grafos de volta para matrizes bidimensionais e formatos tabulares.

Essa conversão forçada impõe a perda imediata de toda a topologia relacional e das regras de inferência embutidas na rede semântica original. Além de operar sob vocabulários de

propósito geral, o modelo evidencia o ecossistema analítico contemporâneo tratando a representação semântica como um mero repositório de extração periférico, e não como o núcleo interpretativo da análise.

O modelo de grafos em favor de tabelas planas para viabilizar o processamento expõe uma grave descontinuidade metodológica, indicando que a verdadeira integração do setor agrícola demanda arquiteturas a quais preservem a riqueza das ontologias de domínio de ponta a ponta, sem rebaixar a estrutura da informação em sua camada analítica.

A análise das iniciativas existentes e exemplificadas no Quadro 8 demonstra as soluções atuais de integração de dados agrometeorológicos concentrando-se primordialmente na criação de catálogos descritivos, na provisão de infraestruturas virtuais em nuvem, na federação de repositórios por meio de contêineres de Objetos Digitais FAIR (FDO) e no acesso sintático a dados processados via Interfaces de Programação de Aplicações (APIs).

A constatação empírica de esgotamento das abordagens baseadas em traduções locais evidencia que nenhuma das arquiteturas analisadas oferece um *framework* desenhado para a tradução automática e direta entre as estruturas dos padrões de metadados centrais na Agrometeorologia, tais como o *Darwin Core* (DwC), o *WMO Core Metadata Profile* (WCMP) e a ISO 19115.

Para superar a fragmentação desse ecossistema e eliminar a dependência de conectores intermediários genéricos, torna-se imprescindível a adoção do AGROVOC como a ponte semântica primária capaz de orquestrar a interoperabilidade nativa entre esses padrões. A persistência desse cenário de isolamento consolida um conjunto de barreiras tecnológicas, cujas lacunas são sistematizadas na próxima seção.

2.5 DESAFIOS E LACUNAS NA LITERATURA

A fundamentação teórica e a análise crítica do estado da arte confirmam a existência de uma lacuna crônica na integração de dados agrometeorológicos. Embora a literatura defenda a necessidade de interoperabilidade e as tecnologias da Web Semântica forneçam o arcabouço teórico para alcançá-la, a aplicação prática em arquiteturas de dados enfrenta barreiras impedindo a comunicação direta entre o clima e a biologia. Conforme evidenciado, essas barreiras consolidam-se em três lacunas específicas.

Primeiramente, constata-se uma lacuna estrutural devido à ausência de um modelo de integração comum para a normalização sintática. Embora os padrões ISO 19115, DwC e WCMP possuam convergências conceituais, como a necessidade de localização no tempo e no

espaço, suas implementações técnicas operam em formatos nativamente incompatíveis. A literatura demonstra a tentativa de orquestrar dados em XML, do Padrão ISO 19115; GeoJSON, do Padrão WCMP; e arquivos de texto plano ou DwC-A, do Padrão *Darwin Core*; resultando em alto esforço computacional e dependência contínua de ferramentas de conversão isoladas.

Em seguida, nota-se uma lacuna de granularidade em função da divergência de escopo, inviabilizando o cruzamento direto das variáveis, uma vez que padrões voltados ao clima e à geoinformação, como o WCMP e o ISO 19115, atuam sob a perspectiva macro, descrevendo o “envelope” do conjunto de dados (*dataset*). Em contrapartida, os padrões da agrobiodiversidade, como o *Darwin Core*, operam em um nível micro, focados na ocorrência ou na medição individual. Desse modo, as arquiteturas analisadas falham em oferecer um modelo unificado acomodando simultaneamente o agrupamento geral da informação climática e o detalhamento granular da observação biológica.

Finalmente, verifica-se uma lacuna semântica, visto que há uma carência severa de validação semântica ativa e direta. O estudo de Beretta *et al.* (2021), por exemplo, corrobora essa afirmação ao diagnosticar os padrões dependentes de domínio que adotam estritamente a visão técnica do produtor do dado, fazendo com que a semântica permaneça encapsulada em jargões de origem.

Essa característica demonstra a integração de dados heterogêneos ser inviável sem a ancoragem estrutural em ontologias externas. Embora iniciativas recentes reconheçam o AGROVOC como o vocabulário de referência do setor para romper esse isolamento, carece de ferramentas para automatizar o mapeamento, convertendo campos propensos à ambiguidade, como o texto livre, em identificadores formais (URIs) de forma direta durante a integração.

Portanto, identificou-se a necessidade de um mecanismo de tradução automática entre os padrões agrometeorológicos, uma vez que a análise de *frameworks*, APIs e iniciativas existentes demonstrou que as soluções atuais mitigam superficialmente a sintaxe, mas não contemplam a interoperabilidade ontológica e estrutural completa.

Diante do exposto, a lacuna central para o desenvolvimento da presente pesquisa está relacionada à ausência de um *framework* focado na tradução automática utilizando um modelo de integração comum para normalizar as estruturas e granularidades divergentes dos Padrões ISO 19115, DwC e WCMP, e, simultaneamente, a implementação da validação semântica ativa, usando o AGROVOC como ponte de harmonização.

A superação desses desafios define os objetivos deste trabalho. O Capítulo 3, a seguir, descreve o material e os métodos adotados para a concepção, o desenvolvimento e a validação do *framework*.

3 MATERIAL E MÉTODOS

Este capítulo apresenta os procedimentos metodológicos adotados para a condução desta pesquisa, detalhando as etapas, os métodos e os instrumentos empregados para o alcance dos objetivos propostos. Visando uma melhor organização lógica e cronológica, a metodologia desenvolvida neste trabalho foi estruturada em duas fases interdependentes.

A primeira fase descreve o rigoroso processo da Revisão Sistemática da Literatura (RSL), pautada nas diretrizes do protocolo internacional PRISMA 2020, utilizada para mapear o estado da arte e embasar tecnicamente o estudo.

A segunda fase detalha as abordagens metodológicas, as referências teóricas e a estratégia de modelagem selecionadas para viabilizar a concepção arquitetural do *framework* agrometeorológico proposto. Sustentada pelas diretrizes da Metodologia *NeOn*, esta etapa estabelece a especificação dos requisitos ontológicos, o paradigma de integração e os critérios de amostragem do cenário de testes, preparando o embasamento necessário para a implementação computacional do sistema.

3.1 REVISÃO SISTEMÁTICA DA LITERATURA

O percurso metodológico delineado para este estudo fundamenta-se nos preceitos da pesquisa aplicada, caracterizada por Marconi e Lakatos (2017) como a investigação de um problema relativo à aplicabilidade prática do conhecimento científico, focando, neste contexto, em solucionar as lacunas de interoperabilidade entre os padrões de metadados agrometeorológicos.

A investigação assume, concomitantemente, uma natureza descritiva, tipologia a qual tem a finalidade principal de delinear e analisar minuciosamente as características de estruturas ou fenômenos (Marconi; Lakatos, 2017), materializada na etapa de análise e comparação dos padrões de metadados ISO 19115, *Darwin Core* e WCMP. Apoando-se, assim, em um procedimento reflexivo, sistemático e controlado, o trabalho evolui da análise teórica para a proposição de um produto tecnológico, culminando na criação e validação de um *framework* para integração estrutural e semântica de padrões de metadados agrometeorológicos.

Para identificar as lacunas existentes, os desafios de integração e analisar as iniciativas e *frameworks* existentes, foi realizada uma Revisão Sistemática da Literatura (RSL). Conforme as diretrizes do protocolo PRISMA, a RSL caracteriza-se como uma revisão que emprega

métodos explícitos e sistemáticos para agrupar e sintetizar as descobertas de estudos voltados a responder a uma questão claramente formulada (Page *et al.*, 2021).

As diretrizes ressaltam que, diferentemente de uma revisão de conveniência, na qual o pesquisador agrupa trabalhos de forma subjetiva, a RSL exige a preparação de um relato transparente, completo e preciso de todas as etapas conduzidas. Page *et al.* (2021) especificam essa exigência protocolar e auditável permitindo aos leitores avaliar a adequação dos métodos empregados, atestando a confiabilidade dos achados e viabilizando a total reprodutibilidade da pesquisa.

Para guiar a condução desta RSL e fundamentar a busca na literatura, definiu-se a seguinte questão principal de pesquisa: “Como a literatura tem abordado a integração semântica e a resolução de lacunas de interoperabilidade entre padrões de metadados agrometeorológicos?”. Em consonância com as diretrizes do protocolo PRISMA proposto por Page *et al.* (2021), é a partir dessa problemática que se estruturam as demais etapas da revisão.

Em seguida, Page *et al.* (2021) estabelecem a necessidade de definir critérios de elegibilidade rigorosos para a seleção dos estudos. Desse modo, os critérios de inclusão adotados foram:

- a) Estudos publicados no recorte temporal entre 2020 e 2025;
- b) Estudos publicados como artigos de revisão e artigos de dados;
- c) Publicações nos idiomas inglês, português e chinês;
- d) Publicações disponíveis em acesso aberto (*open access*) nas bases consultadas.

Em contrapartida, os critérios de exclusão abrangeram:

- a) Estudos publicados fora do período estipulado;
- b) Livros, capítulos de livros e resumos;
- c) Publicações em idiomas distintos dos definidos no escopo;
- d) Publicações com acesso restrito.

A adoção desses critérios pautou-se na necessidade de garantir o rigor e o alinhamento direto com a questão de pesquisa formulada, permitindo a avaliação da adequação dos métodos empregados, conforme as diretrizes de Page *et al.* (2021). O recorte temporal justifica-se pela rápida evolução tecnológica do setor, buscando capturar o estado da arte mais recente das propostas de integração semântica. A inclusão dos idiomas inglês, português e chinês visa abranger a vanguarda das publicações científicas e dos polos de desenvolvimento agropecuário e tecnológico.

A restrição a artigos de revisão e artigos de dados permite focar especificamente em estudos sintetizadores de *frameworks* estruturais ou detalhados por conjuntos de metadados,

isolando essas evidências de pesquisas experimentais primárias, fugindo ao escopo da interoperabilidade.

A busca por evidências foi conduzida nas bases de dados *Scopus*, *Web of Science*, *AGRIS*, *CAB Abstracts*, *IEEE Xplore*, *ScienceDirect*, *SpringerLink* e *ACM Digital Library*, durante o período de abril de 2025 a maio de 2025. Para delinear a estratégia de recuperação dos estudos, os termos foram estruturados em cinco categorias temáticas: Padrões de Metadados, Dados Agrometeorológicos e Climáticos, Integração de Metadados e Interoperabilidade, Tecnologias e Metodologias, e Vocabulário Controlados e Tesouros (AGROVOC e outros). As *strings* de busca detalhadas, combinadas por meio de operadores booleanos e adaptadas às sintaxes específicas de cada base consultada, encontram-se sistematizadas no Quadro 9:

Quadro 9 – Estratégias de Busca

(continua)

Categoria	Busca 1	Busca 2	Busca 3	Busca 4
Padrões de Metadados	("metadata standards" OR "metadata schema" OR "metadata integration" OR "AgMES" OR "Agricultural Metadata Element Set") AND ("WMO Core" OR "ISO 19115" OR "Darwin Core" OR "Dublin Core" OR "AgMES")	("semantic metadata" OR "semantic mapping" OR "metadata transformation" OR "metadata harmonization") AND ("ISO 19115" OR "WMO Core" OR "Darwin Core" OR "AgMES")	("metadata harmonization" OR "metadata conversion" OR "semantic alignment" OR "metadata reconciliation") AND ("climate data" OR "agricultural data" OR "agrometeorological data")	-
Dados Agrometeorológicos e Climáticos	("climate data" OR "meteorological data" OR "agrometeorological data") AND ("metadata" OR "data standardization" OR "AgMES")	("agricultural data" OR "biodiversity data" OR "soil data") AND ("metadata interoperability" OR "data integration" OR "AgMES")	("environmental data" OR "geospatial metadata") AND ("ISO 19115" OR "Darwin Core" OR "AgMES")	-
Integração de Metadados e Interoperabilidade	("semantic interoperability" OR "metadata interoperability") AND ("climate data" OR "agriculture data" OR "agrometeorological data")	("ontology-based integration" OR "semantic metadata mapping") AND ("WMO Core" OR "ISO 19115" OR "AgMES")	("metadata reconciliation" OR "metadata alignment" OR "metadata schema alignment") AND ("agrometeorological data" OR "geospatial data" OR "climate data")	("automatic metadata transformation" OR "metadata conversion framework") AND ("climate data" OR "agriculture" OR "AgMES")

(conclusão)

Categoria	Busca 1	Busca 2	Busca 3	Busca 4
Tecnologias e Metodologias	("RDF" OR "OWL ontology" OR "SPARQL") AND ("metadata integration" OR "data interoperability" OR "AgMES")	("linked data" OR "semantic web") AND ("metadata standardization" OR "ontology mapping" OR "AgMES")	("data harmonization" OR "metadata standardization" OR "semantic mediation") AND ("agricultural metadata" OR "climate metadata" OR "AgMES")	("machine learning for metadata integration" OR "AI for metadata conversion") AND ("climate data" OR "agriculture data" OR "AgMES")
Vocabulários Controlados e Tesouros (AGROVOC e outros)	("AGROVOC" OR "controlled vocabulary" OR "agricultural thesaurus") AND ("metadata standardization" OR "semantic mapping" OR "AgMES")	("semantic annotation" OR "ontology alignment") AND ("AGROVOC" OR "agricultural thesaurus") AND ("climate data" OR "agricultural metadata" OR "AgMES")	-	-

Fonte: O Autor.

Nesse contexto, o processo de seleção dos estudos contou com o auxílio de ferramentas de automação e de gerenciamento bibliográfico, desenvolvendo-se em fases sucessivas de triagem. A busca inicial nas bases de dados resultou na recuperação de 2664 registros, os quais foram exportados e organizados por meio do gerenciador de referências *Zotero*. Na sequência, as referências foram importadas para o sistema web *Rayyan*, plataforma na qual procedeu-se à identificação e exclusão das publicações em duplicidade, restando 2075 artigos únicos para análise.

A etapa de triagem baseou-se na leitura de títulos e resumos e ocorreu em duas etapas complementares. Primeiramente, na Etapa 1 empregou-se a ferramenta de inteligência artificial (IA) *Google Gemini* como método de automação para auxiliar na exclusão preliminar de registros que não atendiam ao escopo da pesquisa, reduzindo o volume para 275 estudos. Em seguida, a avaliação dos títulos e resumos desses 275 artigos remanescentes foi conduzida de forma independente por um único pesquisador e resultou em 75 estudos pré-selecionados.

Cabe ressaltar que a utilização de ferramentas de inteligência artificial (IA) na Etapa 1 de triagem encontra respaldo metodológico nas diretrizes do protocolo PRISMA, que reconhece e incorpora os avanços das ferramentas de automação e de *machine learning* para a seleção de grandes volumes de evidências científicas (Page *et al.*, 2021). Sob a ótica ética e procedimental, a adoção desses recursos justifica-se pela capacidade da IA de otimizar o processamento de

grandes bases de dados, reduzindo o tempo despendido em processos repetitivos (Almeida *et al.*, 2025).

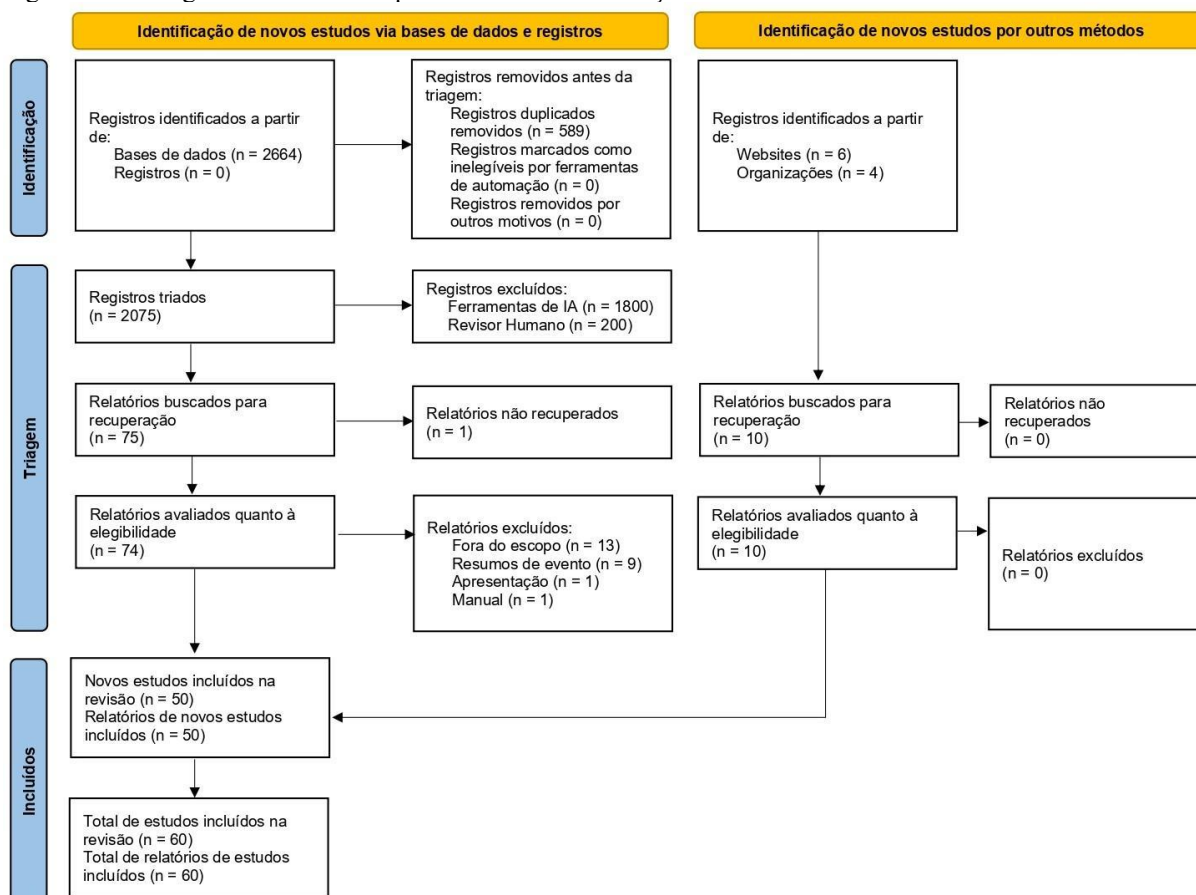
Entretanto, em estrito alinhamento com as recomendações para o uso responsável da tecnologia na pesquisa, assegurou-se que a IA operasse exclusivamente como um recurso assistivo. A centralidade do processo de tomada de decisão foi mantida sob a responsabilidade humana, sendo a triagem final e a aplicação dos critérios de elegibilidade geridas e supervisionadas de maneira autônoma pelo pesquisador, garantindo o rigor metodológico e mitigando eventuais vieses algorítmicos.

A Etapa 2 consistiu na busca pela recuperação e leitura na íntegra desses 75 estudos. Deste montante, um relatório não pôde ser recuperado, por possuir acesso restrito, restando 74 documentos para a aplicação definitiva dos critérios de elegibilidade. Após a leitura completa, 24 relatórios foram excluídos, sendo 13 por estarem fora do escopo, 9 resumos de evento, 1 apresentação e 1 manual, culminando em uma amostra de 50 estudos incluídos provenientes das bases de dados.

Em complemento, a identificação de evidências incluiu a prospecção de novos estudos por meio de outros métodos, englobando a busca em websites e organizações. Foram identificados 10 registros adicionais por essas vias, sendo 6 websites com documentação oficial e 4 documentações de organizações. Todos os 10 relatórios foram recuperados e avaliados na íntegra, sendo todos aprovados e incluídos na revisão.

Dessa forma, somando-se as evidências das bases de dados com as evidências advindas de outros métodos, a amostra final consolidou-se em 60 estudos. Todo o fluxo de identificação, triagem e inclusão, com seus respectivos quantitativos, encontra-se detalhado no fluxograma da Figura 2.

Figura 2 – Fluxograma PRISMA do processo de busca e seleção de estudos



Fonte: Elaborado pelo autor (2024), com base em Page *et al.* (2021).

Com base nos resultados da RSL, os padrões de metadados ISO 19115, *Darwin Core* (DwC) e *WMO Core Metadata Profile* (WCMP) foram vistos como os mais relevantes para o domínio agrometeorológico e foram selecionados para uma análise aprofundada. Nesta etapa, buscou-se por documentações oficiais e conceitos de cada estrutura, independente do ano de publicação, para dar respaldo para uma análise descritiva e comparativa. Paralelamente, o vocabulário controlado AGROVOC foi analisado como a principal “ponte semântica” para harmonizar os conceitos agrícolas. Cabe ressaltar que a busca técnica por esses manuais, diretrizes e especificações compõe os 10 registros identificados por outros métodos (websites e organizações) previamente detalhados no processo de seleção e visto na Figura 2.

Para a etapa de coleta e extração de dados, os 60 documentos incluídos na amostra final foram submetidos a um processo de fichamento técnico. A extração das informações foi conduzida por um único revisor e contou com o suporte da ferramenta baseada em Inteligência Artificial (IA) *Google Gemini* para auxiliar na sumarização e estruturação inicial do conteúdo. Todo o conteúdo extraído e fichado pela ferramenta foi sistematicamente revisado, corrigido e

validado de forma autônoma pelo pesquisador, assegurando a precisão dos dados e a fidelidade aos documentos originais.

De acordo com Page *et al.* (2021), as diretrizes do protocolo PRISMA requerem a definição dos “Itens de dados”, ou seja, das variáveis extraídas de cada estudo e tabuladas, visando transparência e reprodutibilidade do método. Desse modo, a ficha de leitura contemplou três categorias principais:

- 1) Dados de Identificação e Bibliometria: compreendem o título do artigo, autores, nome do periódico e ano de publicação;
- 2) Características e Escopo da Pesquisa: extraíndo informações como o tipo de artigo (artigo de revisão ou artigo de dados), objeto de estudo, objetivos, questões ou hipóteses levantadas e a metodologia empregada;
- 3) Principais achados empíricos ou teóricos: com informações sobre a apresentação dos resultados, discussão do autor com trabalhos correlatos, conclusões alcançadas e a indicação de trabalhos futuros.

Para a etapa de síntese dos resultados, as diretrizes do protocolo PRISMA 2020 reforçam a necessidade de explicitar os chamados Métodos de Síntese (Page *et al.*, 2021). Para isso, adotou-se uma abordagem de síntese qualitativa e narrativa, ou seja, os dados extraídos das 60 fichas de leitura foram analisados, tabulados e agrupados tematicamente de acordo com as seções estruturantes da fundamentação teórica desta pesquisa. Desse modo, os estudos foram divididos em quatro eixos principais de análise:

- 1) Estudos sobre metadados e interoperabilidade;
- 2) Estudos sobre tecnologias e abordagens para integração semântica;
- 3) Padrões de metadados utilizados na agrometeorologia;
- 4) O estado da arte referente à aplicação prática dos padrões, bem como iniciativas e *frameworks* para a sua integração.

Essa estruturação por categorias temáticas viabilizou uma análise comparativa transversal das principais tecnologias, limitações e soluções relatadas na literatura. A partir deste mapeamento sistemático dos agrupamentos, foi possível consolidar as abordagens mais eficientes e extrair os requisitos técnicos necessários para embasar conceitual e estruturalmente a modelagem e a proposição final do *framework* desenvolvido nesta pesquisa. Consequentemente, os resultados detalhados desta RSL fundamentaram a análise crítica da literatura, a identificação das lacunas de pesquisa e a justificativa deste trabalho, conforme apresentado no Capítulo 2.

3.2 ESPECIFICAÇÃO DE REQUISITOS DA ONTOLOGIA

Com a consolidação do mapeamento da literatura e a identificação dos padrões de metadados ISO 19115, *Darwin Core*, WCMP e do vocabulário AGROVOC como pilares estruturais, a segunda fase desta pesquisa concentrou-se na concepção tecnológica da solução. Para guiar este processo, adotou-se a Metodologia *NeOn* proposta por Suárez-Figueroa *et al.* (2012), que se caracteriza por ser um *framework* de referência internacional para a engenharia de redes de ontologias. Diferente das abordagens tradicionais, a Metodologia *NeOn* destaca-se por não prescrever um fluxo de trabalho rígido, baseando-se em um conjunto de cenários flexíveis que priorizam a reutilização, o alinhamento e a reengenharia de recursos ontológicos e não ontológicos preexistentes (Suárez-Figueroa *et al.*, 2012).

Para o escopo deste trabalho, que visa superar as heterogeneidades semânticas entre os domínios do clima e da biologia agrícola, essa metodologia mostra-se ideal pois fornece o embasamento prático para harmonizar os padrões isolados mapeados na revisão sistemática dentro de um ecossistema semântico conectado.

Estabelecido o modelo metodológico, a etapa inicial de concepção do *framework* consistiu na especificação dos seus requisitos, fase que a Metodologia *NeOn* define, segundo Suárez-Figueroa *et al.* (2012), como Especificação de Requisitos da Ontologia, cujo objetivo é delimitar o escopo da solução, identificando seus usos pretendidos e as necessidades técnicas que deve atender.

Desse modo, o propósito e escopo da ontologia consistem em atuar como uma ponte semântica interdomínios para a integração de dados biológicos, meteorológicos e de proveniência. Como linguagem de implementação formal, definiu-se a utilização da OWL, recomendada pelo W3C, conforme abordado no Capítulo 2 de Fundamentação Teórica desta pesquisa.

A escolha desta tecnologia justifica-se por sua capacidade de elevar a expressividade semântica do modelo, fornecendo as entidades e os axiomas necessários para descrever formalmente as restrições rigorosas, cardinalidades e relações do domínio (W3C, 2012). Ademais, de acordo com Fensel *et al.* (2003), é a formalidade rigorosa da OWL que possibilita aos motores de inferência avaliar o esquema, verificar a ausência de contradições lógicas e realizar o raciocínio automatizado.

Para atender às diretrizes subsequentes da fase de Especificação de Requisitos da Ontologia, Suárez-Figueroa *et al.* (2012) enfatizam que a Metodologia *NeOn* define os usuários finais e os usos pretendidos da ontologia. Assim, para elucidar esses aspectos e delimitar com

exatidão o escopo tecnológico desta pesquisa, a cadeia de valor da informação do *framework* proposto foi estruturada conceitualmente em três camadas lógicas distintas: Ingestão, Integração e Aplicação.

A Camada de Ingestão não constitui um módulo de software desenvolvido nesta pesquisa, mas sim a representação conceitual do cenário de dados brutos encontrado na web. Ela delimita o ambiente onde residem os arquivos originais exportados pelos repositórios, os quais se apresentam em formatos sintáticos e estruturais estritamente heterogêneos, como arquivos XML, GeoJSON e CSV.

A caracterização desta camada ilustra de forma prática a problemática mapeada por Putrama e Martinek (2024), os quais advertem que a crescente diversidade de fontes e a falta de padrões de metadados uniformes formam a principal barreira tecnológica da atualidade, impedindo que as informações ambientais originais sejam cruzadas de forma nativa por sistemas comerciais.

A Camada de Integração, por sua vez, representa o núcleo de atuação desta pesquisa. Nesta etapa, o *framework* proposto opera como um *middleware* semântico que consome os formatos heterogêneos e processa as regras de mapeamento.

Para realizar a desambiguação terminológica necessária, a arquitetura utiliza o vocabulário AGROVOC, que atua como o mecanismo para alcançar a interoperabilidade semântica no domínio agrícola, conforme estabelecido por Takezaki *et al.* (2020). O produto final desta camada é uma infraestrutura de dados limpa e unificada como um Grafo do Conhecimento.

Para sustentar essa arquitetura, adotou-se o modelo de dados abstrato RDF, sendo projetado para ser consultável de forma nativa via linguagem SPARQL. Essa abordagem materializa o enriquecimento semântico preconizado por Meckler *et al.* (2023), transformando dados isolados em conhecimento interligado e interpretável por máquinas.

Por último, a Camada de Aplicação delineia os usos pretendidos e define os usuários finais da solução. O público-alvo direto da ontologia não é o produtor rural na ponta da cadeia, mas sim desenvolvedores de software, meteorologistas, cientistas de dados e sistemas automatizados.

Esses usuários técnicos utilizarão o Grafo de Conhecimento consolidado pelo *middleware* como infraestrutura de dados subjacente. A interação entre os sistemas não ocorre pela transferência do grafo em sua totalidade, mas sim por meio de requisições na linguagem SPARQL, que atua como a interface unificada para recuperar as informações. A partir dessas consultas, a Camada de Aplicação extrai respostas estruturadas e direcionadas para alimentar

soluções de mercado, como painéis de inteligência de negócios (*dashboards*) ou aplicativos móveis de alerta agrometeorológico.

Essa separação de escopo reflete a visão de Roussaki *et al.* (2023), que destacam que a interoperabilidade não se limita à troca de dados, mas deve garantir a capacidade de os sistemas interagirem de forma automatizada, transformando a abundância de informações em conhecimento produtivo para apoiar a tomada de decisão humana.

Após a definição das diretrizes iniciais de propósito, linguagem, usuários e arquitetura, a etapa seguinte da Especificação de Requisitos da Ontologia exige a definição das necessidades funcionais do modelo (Suárez-Figueroa *et al.*, 2012).

Para formalizar esses requisitos, a Metodologia *NeOn* orienta a extração e o uso de Questões de Competência (do inglês, *Competency Questions - CQs*), conceituadas pelos autores como perguntas estruturadas em linguagem natural que representam os problemas e cenários reais que a rede de ontologias e a solução devem ser capazes de responder após a sua implementação.

Para assegurar a validade e a relevância prática do *framework* proposto, a formulação das Questões de Competência desta pesquisa foi embasada na literatura técnica especializada do domínio agrometeorológico e instanciada com base num cenário real e de elevada importância econômica, o complexo praga-cultura da *Spodoptera frugiperda* e do Milho (*Zea mays*).

A escolha desta espécie justifica-se pela sua extrema severidade e pelo seu alto impacto econômico destrutivo na cultura do milho, pois, conforme atestado em estudo da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) conduzido por Oliveira *et al.* (2023), a *Spodoptera frugiperda* constitui a principal praga-chave do milho em toda a América do Sul, caracterizando-se como um inseto capaz de se alimentar e atacar diferentes famílias de plantas, sendo extremamente destrutivo e com rápida adaptação às principais estratégias de manejo.

Assim, ao invés de elaborar perguntas puramente abstratas, a adoção desse cenário de gravidade agrônômica permite demonstrar de forma inequívoca a capacidade do sistema em resolver ambiguidades textuais, inferir relações semânticas e cruzar dimensões temporais extremas, como o estresse hídrico, com o impacto produtivo direto. Em estrito alinhamento com a RSL apresentada no Capítulo 2, as Questões de Competência (CQs) deste estudo foram definidas para solucionar as principais lacunas de integração agrometeorológicas, conforme mostra o Quadro 10:

Quadro 10 – Questões de Competência e os Padrões de Metadados Agrometeorológicos Envolvidos

Código	Questão de Competência	Objetivo Técnico	Padrões e Vocabulários Envolvidos
CQ01	Quais ocorrências da praga <i>Spodoptera frugiperda</i> (registadas individualmente no padrão <i>Darwin Core</i>) foram detectadas dentro da extensão geográfica e temporal de um conjunto de dados meteorológicos (WCMP / ISO 19115) que reportou precipitação abaixo de 10mm?	Resolver o conflito de granularidade (<i>Dataset</i> vs. Ocorrência)	<i>Darwin Core</i> , WCMP e ISO 19115
CQ02	Quais registros agrometeorológicos descrevem a cultura do milho, independentemente de a fonte original ter usado texto livre (ISO 19115) ou uma taxonomia estrita (<i>Darwin Core</i>)?	Padronizar termos e eliminar o texto livre ambíguo	ISO 19115, <i>Darwin Core</i> e AGROVOC
CQ03	Qual é a linhagem (<i>lineage</i> da ISO 19115) e a política de dados (<i>dataPolicy</i> do WCMP) do recurso climático que serviu de entrada para a monitorização em tempo real da <i>Spodoptera frugiperda</i> ?	Garantir a reprodutibilidade através da proveniência	ISO 19115 e WCMP
CQ04	Quais medições específicas de sensores (<i>MeasurementOrFact</i> do <i>Darwin Core</i>) indicam fatores climáticos extremos que afetam (<i>propriedade affects</i> da <i>Agrontology</i>) o desenvolvimento da cultura do Milho?	Correlacionar variáveis de sensores em tempo real	<i>Darwin Core</i> e AGROVOC (<i>Agrontology</i>)

Fonte: O Autor.

Conforme sintetizado no Quadro 10, a formulação de cada Questão de Competência (CQ) foi delineada para transpor barreiras específicas de interoperabilidade mapeadas na literatura. A CQ01 trabalha com o desafio da divergência de escopo, lidando com a Lacuna de Granularidade. Enquanto padrões como o WCMP e a ISO 19115 atuam em um nível macro, funcionando como um "envelope" para o conjunto de dados completo (*dataset*), o padrão *Darwin Core* foca na granularidade fina da ocorrência e do evento biológico. Ao cruzar essas estruturas, a CQ01 visa garantir o que Vogt *et al.* (2025) definem como interoperabilidade proposicional, alinhando as regras lógicas dos esquemas.

A CQ02, por sua vez, foi desenvolvida para responder a Lacuna Semântica, uma vez que a literatura evidencia que o uso de descrições genéricas baseadas em texto livre, como permitido pela ISO 19115, gera ambiguidades e leva à criação de tecnologias isoladas, as chamadas "ilhas de informação". Para solucionar esse problema, a CQ02 propõe o alcance da interoperabilidade terminológica, como ressaltado por Vogt *et al.*, 2025, utilizando o vocabulário multilíngue AGROVOC como ponte semântica.

Já a CQ03 concentra-se na camada de *inputs* da pilha analítica, assegurando os princípios da Pesquisa Computacional Reprodutível (PCR) preconizados por Leipzig *et al.* (2021). A extração de dados sobre a linhagem da informação (*lineage* da ISO 19115) e suas

políticas de acesso (*dataPolicy* do WCMP) atua diretamente no pilar de Reutilização dos princípios FAIR. Com isso, o *framework* previne o declínio e a perda de dados a longo prazo, fenômeno classificado por Habermann (2020) como a "Entropia da Informação".

A CQ04 é voltada para a Análise Holística e o Monitoramento Preditivo. Para responder a perguntas analíticas de causa e efeito, o modelo explora a classe *MeasurementOrFact* do *Darwin Core*, cruzando os dados dinâmicos de sensores com as relações não-hierárquicas oferecidas pela ontologia de suporte do AGROVOC, à *Agrontology*, como as propriedades de causalidade *affects* e *isAffectedBy*. Dessa forma, assegura-se que a interação entre as variáveis climáticas e o desenvolvimento agrícola seja compreendida de forma automatizada.

Com as Questões de Competência consolidadas e fundamentadas, em seguida a Metodologia *NeOn* orienta a extração de um pré-glossário de termos (Suárez-Figueroa *et al.*, 2012). A partir das CQs formuladas, extraiu-se a terminologia que o modelo deverá representar formalmente. Termos-chave como *dataset*, ocorrência, observação, temperatura, precipitação, praga e coordenadas espaciais foram mapeados para compor a base vocabular. Este pré-glossário atua como o principal insumo para a próxima etapa metodológica, norteando a busca, a seleção e o reuso de recursos ontológicos preexistentes capazes de suprir essas necessidades terminológicas.

3.3 ESPECIFICAÇÃO DO FRAMEWORK E ESTRATÉGIA DE MODELAGEM

Tendo como base o pré-glossário estabelecido na seção anterior, a seleção de ferramentas e a estratégia de modelagem para a viabilização técnica da arquitetura planejada foram delineadas visando garantir uma integração escalável e de alto desempenho.

Conforme as diretrizes da Metodologia *NeOn* abordadas por Suárez-Figueroa *et al.* (2012), a criação ou importação massiva de grandes ontologias monolíticas gera severos problemas de reusabilidade, escalabilidade e manutenção. Para mitigar esse fenômeno, evidenciado pelos autores e amplamente conhecido na literatura técnica como “inchaço do modelo” (*ontology bloating*), a estratégia metodológica adotada baseou-se no Alinhamento por Referência de URI.

Para viabilizar tecnicamente a representação do pré-glossário estabelecido e garantir uma integração escalável, a estratégia de modelagem foi delineada evitando o desenvolvimento de classes redundantes. Dessa forma, o ambiente *Protégé* e a linguagem OWL foram selecionados para construir uma Ontologia de Alinhamento enxuta, que atua exclusivamente como uma ponte semântica, ou seja, em vez de importar vocabulários inteiros, a camada de

terminologias (*TBox*) foi estruturada declarando classes de ancoragem, como *dwc:Occurrence*, *iso19115:MD_Metadata* e *wcmp2:Record*, que conectam os recursos brutos aos padrões internacionais e ao vocabulário AGROVOC.

Com o intuito de complementar e enriquecer esse modelo, a arquitetura semântica da Camada de Integração foi delineada internamente em duas subcamadas interdependentes, projetadas para atender aos diferentes tipos de termos mapeados no pré-glossário.

A Subcamada de Domínio opera os padrões agrometeorológicos mapeados para cobrir os termos específicos do nicho, enquanto a Subcamada de Fundamentação incorpora o reuso das subclasses universais *SOSA*, *GeoSPARQL* e *OWL-Time* para abranger os termos transversais do pré-glossário, como espaço, tempo e observações. Este reuso é recomendado pelas diretrizes da Metodologia *NeOn*, que traz que o uso das chamadas ontologias fundacionais ou de núcleo (*core ontologies*) previne o retrabalho e garante uma estruturação lógica universal validada pela comunidade científica (Suárez-Figueroa *et al.*, 2012).

A adoção dessa arquitetura em duas camadas resolve o dilema da integração evidenciado no Capítulo 2 desta pesquisa, uma vez que espelha as duas dimensões de interoperabilidade propostas por Vogt *et al.* (2025). Enquanto a Subcamada de Domínio, atrelada ao AGROVOC, garante a interoperabilidade terminológica, padronizando os nomes específicos do nicho agroambiental, como discutido por Takezaki *et al.* (2020) e Putrama e Martinek (2024), a Subcamada de Fundamentação provê a estrutura para a interoperabilidade proposicional trazida pelos próprios Vogt *et al.* (2025). Esta subcamada fornece as regras lógicas estritas e universais de como os elementos do mundo real interagem, garantindo que o modelo seja processável em ambientes computacionais.

Na prática, a aplicação destas *core ontologies* na Subcamada de Fundamentação transcende o alinhamento sintático, representando a principal contribuição tecnológica para solucionar o engessamento das bases de dados agrícolas. O reuso da *GeoSPARQL* confere ao modelo a capacidade de executar consultas topológicas automatizadas para o cruzamento espacial entre pragas e bacias hidrográficas, por exemplo. Simultaneamente, a adoção da *OWL-Time* transforma cadeias de texto estáticas em entidades temporais computáveis, habilitando o uso de filtros complexos de proximidade e sucessão cronológica necessários para o monitoramento agrometeorológico.

Já a ancoragem dos metadados de observação à ontologia *SOSA* soluciona a grave lacuna de granularidade mapeada na literatura, ao permitir a descrição da informação na escala atômica do sensor, atuando na quebra do que Prudhomme *et al.* (2025) chamaram de "silos de

ontologias", e viabilizando que as variáveis de chuva e temperatura sejam integradas organicamente a ecossistemas de Internet das Coisas (IoT) e inteligência artificial.

Em seguida, para assegurar a integridade estrutural e a ausência de contradições lógicas no modelo de alinhamento proposto, estabeleceu-se a aplicação de uma etapa de validação automatizada baseada em Lógica de Descrição, como sugerido por Fensel *et al.* (2003). Para tanto, elegeu-se o motor de inferência (*reasoner*) *HermiT*, nativamente integrado ao *Protégé*, que executou os testes de Verificação de Consistência (*Consistency*) e Classificação Automática (*Classification*), inspecionando rigorosamente as restrições de domínio e alcance das propriedades entre um domínio e outro.

Este critério metodológico atesta a robustez do esquema semântico concebido, garantindo a ausência do que Suárez-Figueroa *et al.* (2012) chamaram de classes insatisfatíveis (*unsatisfiable class*), ou seja, classes que provocam uma contradição lógica tão grave que é matematicamente impossível que sejam classes com instâncias dentro do modelo.

A operacionalização desta arquitetura exigiu a definição das ferramentas de processamento e consulta, estabelecendo uma rigorosa separação de preocupações (*separation of concerns*). Essa estratégia metodológica encontra forte respaldo nas soluções arquiteturais contemporâneas mapeadas na literatura. Conforme preconizado pelo modelo de Martínez *et al.* (2025), a verdadeira interoperabilidade exige o isolamento estrutural entre a modelagem semântica e a persistência física dos dados. Adicionalmente, essa divisão reflete a arquitetura proposta por San Emeterio de la Parte *et al.* (2023), que comprova a eficiência de delegar a inteligência de domínio e o raciocínio ontológico para repositórios de triplas especializados independentes.

Apoiado nessas diretrizes e atuando exclusivamente na Camada de Integração, o *framework* foi concebido como uma Interface de Programação de Aplicações (API) de design *RESTful*, desenvolvida na linguagem Python sob a infraestrutura do *FastAPI*. Essa API atua como o motor responsável por receber os arquivos originais, que compõem a abstração conceitual da Camada de Ingestão, e processar as regras de conversão integralmente em memória, utilizando a biblioteca *RDFLib*. Nesta etapa, a linguagem SPARQL opera de forma encapsulada nos bastidores da API apenas para interrogar o vocabulário global AGROVOC em tempo real.

O processamento finaliza serializando as instâncias geradas (*ABox*) para o padrão estrutural *Turtle* (.ttl). Para assegurar a eficiência da comunicação *RESTful*, o formato JSON é utilizado como o envelope de transporte, garantindo que a transação de dados seja auditável e contenha metadados de controle de execução. A serialização *Turtle* em RDF é mantida

internamente nesse *payload*, preservando a interoperabilidade semântica e permitindo que o sistema cliente trate a resposta como um objeto de software antes de persistir o grafo bruto em um banco de dados.

Em contrapartida, a validação prática do *framework*, materializada pela resposta às Questões de Competência (CQs), é delegada à Camada de Aplicação. Para viabilizar essa validação de conceito, o grafo gerado foi consumido pelo banco de dados orientado a grafos *GraphDB*, ambiente onde as requisições em linguagem SPARQL foram executadas de forma explícita pelo pesquisador para testar as inferências lógicas e extrair o conhecimento orquestrado. Esse desacoplamento assegura que o *framework* atue de forma ágil e sem estado, entregando o Grafo de Conhecimento empacotado e transferindo a responsabilidade de manipulação e armazenamento final para os sistemas que o consumirem.

3.4 AMOSTRAGEM E COLETA DE DADOS

Com a arquitetura semântica do *framework* estruturada e logicamente validada, a etapa seguinte da pesquisa concentrou-se na obtenção dos insumos necessários para a instanciação e a execução da prova de conceito do modelo. Seguindo as diretrizes do Cenário 3 da Metodologia *NeOn*, Suárez-Figueroa *et al.* (2012) orientam a necessidade de identificação e reutilização de recursos não ontológicos preexistentes, procedendo à captura de dados e metadados reais de produção.

Para afastar qualquer viés de arbitrariedade na coleta, o critério de amostragem baseou-se estritamente na representatividade dos três domínios que compõem a tríade agrometeorológica: o biológico, focado na ocorrência de pragas, o geoespacial de proveniência, focado no mapeamento hídrico regional, e o meteorológico, focado nas redes de sensores. Além disso, cada recurso foi selecionado exclusivamente a partir de plataformas oficiais e vinculado de forma direta à capacidade de validar e responder às Questões de Competência (CQs) estabelecidas na fase de requisitos.

Para representar o domínio biológico da tríade agrometeorológica, a extração de dados concentrou-se no monitoramento de vetores nocivos à agricultura. O conjunto de dados foi obtido a partir do portal *Global Biodiversity Information Facility* (GBIF), que abriga o maior repositório de dados de ocorrência de espécies do mundo.

O critério de amostragem delimitou-se à taxonomia da espécie *Spodoptera frugiperda* (lagarta-do-cartucho), restringindo o escopo espacial às observações realizadas no território brasileiro. Como justificado anteriormente na formulação das Questões de Competência, a

extração de dados biológicos concentrou-se nos registros da espécie *Spodoptera frugiperda*, principal praga-chave do milho em toda a América do Sul, caracterizando-se como um inseto polífago extremamente destrutivo e de rápida adaptação às principais estratégias de manejo. Esse cenário de gravidade e complexidade agrônômica constitui, assim, um caso de uso real e representativo para validar as capacidades analíticas do *framework*.

Em alinhamento direto com as diretrizes mapeadas na revisão da literatura, o recurso foi exportado de sua base original estruturado no formato *Darwin Core Archive* (DwC-A), visto que, conforme evidenciado por Wiczorek *et al.* (2012) e pelo *Darwin Core Maintenance Group* (2021), a utilização desse padrão garante uma representação de granularidade fina da informação.

Diferentemente dos padrões climáticos e espaciais, como o WCMP v2 e a ISO 19115, que operam no nível macro do conjunto de dados atuando como um "envelope", o padrão DwC desce ao nível atômico, assegurando que cada linha do arquivo atue como uma instância única e independente da classe biológica *dwc:Occurrence*.

A seleção estrutural desse recurso não ontológico fundamenta a viabilidade técnica do modelo e supre, de forma direta, os insumos exigidos para responder a duas Questões de Competência (CQs). Primeiramente, o arquivo provê os atributos precisos de geolocalização e tempo para a CQ01.

O padrão documenta as coordenadas exatas através dos termos *dwc:decimalLatitude* e *dwc:decimalLongitude*, bem como o carimbo temporal no campo *dwc:eventDate*. Como demonstrado no Capítulo 2 desta pesquisa, a extração dessas coordenadas espaciais e temporais no nível da observação é a premissa tecnológica indispensável para permitir que a detecção da praga seja correlacionada de forma cruzada com a ocorrência das matrizes físicas e meteorológicas do ambiente.

Em segundo lugar, a adoção deste arquivo suporta diretamente a resolução da CQ02, uma vez que ao garantir o fornecimento de uma identificação estrita e padronizada no campo *dwc:scientificName*, o *Darwin Core* elimina o uso das descrições textuais informais que tipicamente geram "silos de informação" e ambiguidades.

Essa taxonomia rigorosa fornece a estrutura ideal para que o *framework* construa os mapeamentos semânticos em direção ao vocabulário AGROVOC, garantindo a desambiguação terminológica necessária para a plena interoperabilidade no ecossistema de dados agrometeorológicos.

Nesse contexto, para representar o domínio geoespacial e de proveniência, a coleta de metadados foi realizada no catálogo GeoInfo, mantido pela Empresa Brasileira de Pesquisa

Agropecuária (Embrapa). O recurso não ontológico selecionado constitui o conjunto de dados intitulado "Excedentes hídricos decendiais na Bacia Hidrográfica do Rio Paraná", com o escopo geográfico delimitado ao estado do Paraná. Para garantir a preservação da estrutura original da instituição, o arquivo foi exportado em sua sintaxe XML nativa, orientada pela norma de codificação ISO 19139.

Diferentemente da granularidade atômica exigida no domínio biológico, a estruturação deste recurso apoia-se no padrão ISO 19115, que, conforme discutido no Capítulo 2, opera tipicamente no nível macro do conjunto de dados (*dataset*). Essa arquitetura atua como um "envelope" descritivo para a informação, fornecendo o contexto espacial e administrativo necessário para que as matrizes físicas de excedente hídrico sejam processadas de forma padronizada pelo *framework*.

A escolha deste recurso específico justifica-se pela necessidade estrita de fornecer os insumos metodológicos para a validação da CQ03, uma vez que o mapeamento estrutural do padrão ISO 19115 viabiliza a extração normalizada de pacotes de metadados para a governança de dados, como as informações do histórico de processamento (*LI_Lineage*) e a identificação formal da responsabilidade institucional sobre o dado (*contact/responsible party*).

O resgate dessas propriedades atua diretamente na camada de *inputs* da pilha analítica da Pesquisa Computacional Reprodutível (PCR) abordada por Leipzig *et al.* (2021). Ao garantir a documentação rigorosa da origem da informação agroambiental, o uso deste padrão no *framework* sustenta o pilar da Reutilização dos princípios FAIR e atua como o "antídoto crítico" preconizado por Habermann (2020) para frear a Entropia da Informação, evitando o declínio e a perda de dados a longo prazo.

Finalmente, para consolidar a tríade agrometeorológica, a extração de dados do domínio meteorológico foi conduzida junto ao nó oficial do Instituto Nacional de Meteorologia (INMET), o qual encontra-se plenamente integrado ao ecossistema moderno WIS 2.0 da Organização Meteorológica Mundial (OMM). O recurso não ontológico selecionado corresponde ao conjunto de dados de valores diários de estações de superfície, denominado DAYCLI, identificado unicamente pelo URI “*urn:wmo:md:br-inmet:daycli*”. Para garantir a interoperabilidade nativa com as redes globais da Web, o arquivo foi capturado no formato estruturado GeoJSON, em estrita conformidade com o padrão internacional *WMO Core Metadata Profile version 2* (WCMP v2).

Conforme evidenciado na revisão da literatura, o padrão WCMP v2 opera estruturalmente no nível macro do conjunto de dados (*dataset*), semelhante ao ISO 19115,

atuando como um "envelope" que descreve as características globais da coleção de medições, e delegando o acesso aos grânulos individuais para serviços e APIs de dados externos.

A escolha desta arquitetura justifica-se por fornecer as propriedades mandatórias de abrangência geográfica (*geometry*) e delimitação temporal (*time*). Adicionalmente, a classificação do recurso sob a política de dados abertos da OMM (*properties.wmo:dataPolicy* do tipo *core*) garante o acesso livre e irrestrito à informação, atendendo diretamente ao pilar de Acessibilidade dos princípios FAIR.

A adoção deste padrão meteorológico atende de forma abrangente aos requisitos do *framework*, fornecendo os insumos transversais para a validação das demais Questões de Competência (CQs). O fornecimento do envelope espaço-temporal atua como o alicerce para o filtro cruzado exigido pela CQ01, enquanto a governança e identificação providas pelo metadado complementam as exigências da CQ03.

O grande diferencial tecnológico deste recurso reside na sua propriedade *properties.themes*, que exige a vinculação dos temas meteorológicos a vocabulários controlados externos por meio de URIs. Como fundamentado na pesquisa, essa exigência estrutural fornece a âncora semântica exata para que o *framework* conecte as variáveis do clima ao vocabulário AGROVOC.

Esse arranjo viabiliza o alinhamento de sensores necessário para solucionar a CQ04, integrando, em um único grafo de conhecimento, o impacto dos fenômenos meteorológicos sobre o desenvolvimento e o monitoramento no domínio biológico.

Durante a composição do cenário de testes do domínio institucional e de proveniência, identificou-se que o arquivo original em formato XML, aderente ao padrão ISO 19115, apresentava omissões de preenchimento estruturais em *tags* críticas, indicadas nativamente pelo atributo *gco:nilReason="missing"*.

Visto que a ausência dessas *strings* textuais inviabilizaria a validação técnica da CQ03, cujo objetivo é atestar a capacidade mecânica do *framework* em extrair a autoria institucional e a rastreabilidade histórica, optou-se por criar uma simulação estrita em ambiente controlado.

Para preservar a integridade e a proveniência bruta do arquivo original, foi gerada uma cópia de testes na qual se realizou o povoamento manual desses campos ausentes em nível local. Desse modo, as *tags* *<gmd:organisationName>* e *<gmd:statement>* foram populadas com a instância legítima “Embrapa Milho e Sorgo” e com o detalhamento descritivo da linhagem de coleta das armadilhas, respectivamente.

Essa estratégia de injeção de dados simulados no arquivo cópia garantiu que o *pipeline* de triplificação em Python pudesse ser validado sobre um conjunto de dados semanticamente

íntegro, comprovando a eficácia extrativa do modelo para atuar sobre o esquema canônico da ISO sem incorrer em adulteração primária dos dados originais da instituição

Ainda no escopo deste cenário de testes, e como estabelecido pelas diretrizes do padrão WCMP v2, o arquivo GeoJSON atua estritamente como um envelope macroscópico, delegando o acesso às medições individuais para serviços externos. Desse modo, para viabilizar a execução prática do *framework* e o cruzamento granular das variáveis meteorológicas e biológicas, procedeu-se à extração de um arquivo tabular CSV contendo as observações reais de superfície vinculadas a este metadado. O recorte espacial da amostra concentrou-se no município de Morretes, localizado na região ecológica da Serra do Mar, no estado do Paraná.

A escolha desta área experimental justifica-se rigorosamente por critérios de intersecção espaço-temporal e relevância climatológica. Primeiramente, o município encontra-se na mesma unidade federativa delimitada pelo conjunto de dados geoespaciais da Embrapa (Bacia Hidrográfica do Rio Paraná), fornecendo a sobreposição geográfica exigida para a validação do filtro cruzado da CQ01. Além disso, por se tratar de uma região caracterizada por dinâmicas atmosféricas complexas e elevados índices pluviométricos decorrentes do relevo orográfico, Morretes oferece o ambiente ideal para a avaliação de limiares meteorológicos críticos.

A inserção dessas matrizes de clima extremo no Grafo de Conhecimento permite avaliar, de forma contundente, a capacidade do modelo em responder à CQ04, emitindo alertas de estresse ambiental e correlacionando as variações climáticas ao desenvolvimento e monitoramento biológico.

Desse modo, o município apresentou a convergência ideal de repositórios públicos heterogêneos devido à presença de registros de ocorrências biológicas georreferenciadas no *Global Biodiversity Information Facility* (GBIF) concomitante à operação contínua da estação meteorológica automática A873 do Instituto Nacional de Meteorologia (INMET).

A delimitação da amostragem para o ano de 2025 justifica-se pela necessidade de garantir a extração de dados reais recentes e viabilizar o emparelhamento cronológico entre os registros. A sobreposição exata de coordenadas espaciais e carimbos temporais permitiu testar a robustez das regras de inferência do *framework*, garantindo que o entrelaçamento semântico gerasse um Grafo de Conhecimento conexo e auditável face às Questões de Competência estabelecidas.

Com a consolidação desta amostra de dados agrometeorológicos e a aplicação dos devidos ajustes ao cenário de testes, encerra-se a etapa de identificação e coleta de recursos não ontológicos estabelecida pela Metodologia *NeOn*. Tendo os insumos brutos e as diretrizes

arquiteturais plenamente definidas, o Capítulo 4 será dedicado ao desenvolvimento prático e à implementação computacional do *framework*.

A próxima etapa detalhará a construção da Ontologia de Alinhamento no ambiente *Protégé*, a operacionalização do pipeline de triplificação em linguagem Python e a instânciação e validação do Grafo de Conhecimento unificado, demonstrando a resolução prática das Questões de Competência (CQs) propostas para o ecossistema agrometeorológico.

4 DESENVOLVIMENTO E IMPLEMENTAÇÃO

Este capítulo apresenta o desenvolvimento prático e a implementação computacional do *framework* AgroMeta, proposto para a integração semântica de metadados agrometeorológicos. Fundamentado nas diretrizes arquiteturais e na coleta de recursos não ontológicos definidos no capítulo anterior, o texto detalha a transição do planejamento metodológico para a execução tecnológica. Nas seções a seguir, serão descritas as etapas de construção da Ontologia de Alinhamento no ambiente *Protégé*, os processos de desambiguação terminológica, a operacionalização do *pipeline* de triplificação utilizando a linguagem Python e a consolidação do Grafo de Conhecimento unificado. Desse modo, o capítulo culmina na validação prática da arquitetura, demonstrando a resolução das Questões de Competência (CQs) propostas e a identificação de lacunas em bases globais, revelando o potencial do modelo como uma ferramenta de auditoria para o ecossistema agrometeorológico.

4.1 CONSTRUÇÃO DA ONTOLOGIA DE ALINHAMENTO

A primeira etapa para a materialização da Camada de Integração consistiu na modelagem estrutural da Ontologia de Alinhamento, concebida para atuar como o alicerce terminológico (*TBox*) do *framework* focado na integração de metadados heterogêneos de biodiversidade, meteorologia e proveniência institucional.

A engenharia desta ontologia foi conduzida no ambiente *Protégé*, fundamentando-se estritamente na diretriz da Metodologia *NeOn*, que recomenda a reutilização de vocabulários internacionais consolidados e recursos não ontológicos para evitar a redundância de conceitos e garantir a interoperabilidade global do modelo.

Desse modo, em vez de criar classes redundantes ou importar ontologias monolíticas, a estratégia de modelagem priorizou o estabelecimento de uma ponte semântica enxuta em linguagem OWL. Essa abordagem permitiu conectar os padrões de metadados originais às ontologias fundacionais GeoSPARQL, SOSA e *OWL-Time* de forma declarativa.

Para viabilizar a estruturação dessa ponte semântica e evitar ambiguidades ou problemas de renderização e *cache* gráfico no ecossistema de desenvolvimento, a ação inicial no ambiente *Protégé* consistiu na configuração rigorosa dos *namespaces* e de seus respectivos prefixos. Conforme as boas práticas da Web Semântica, abandonou-se o uso do prefixo vazio “:” em favor da adoção de um prefixo explícito e oficial derivado do nome concebido para a solução:

o *framework* AgroMeta. Dessa forma, a adoção do prefixo *agrometa:* garante a identidade global e única das propriedades modeladas nativamente.

Adicionalmente, em estrita conformidade com a diretriz de reuso metodológico abordada, os URIs das *core ontologies*, dos padrões de metadados e do vocabulário controlado AGROVOC foram formalmente declarados no cabeçalho do documento, conforme mostra o Quadro 11:

Quadro 11 – Prefixos e Namespaces configurados na Ontologia de Alinhamento

Prefixo	Domínio/Padrão Representado	Namespace (URI)
agrometa	Prefixo <i>Core</i> do <i>Framework</i>	http://agrometa-framework.org/ontology/agrometa#
wcmp2	<i>WMO Core Metadata Profile 2</i>	http://wmo-im.github.io/wcmp2/
dwc	<i>Darwin Core</i>	http://rs.tdwg.org/dwc/terms/
iso19115	ISO 19115 (<i>Geographic Information</i>)	http://www.isotc211.org/2005/gmd/
geo	OGC GeoSPARQL	http://www.opengis.net/ont/geosparql#
sosa	SOSA (<i>Sensor, Observation, Sample...</i>)	http://www.w3.org/ns/sosa/
time	W3C <i>OWL-Time</i>	http://www.w3.org/2006/time#
agro	AGROVOC/FAO <i>Agrontology</i>	http://aims.fao.org/aos/agrontology#

Fonte: O Autor.

Conforme detalhado no Quadro 11, a configuração de *namespaces* da ontologia de alinhamento reflete diretamente a divisão arquitetural planejada para o *framework*. O prefixo “agrometa:” foi instituído como o identificador central para encapsular as propriedades relacionais concebidas nativamente no modelo.

Para viabilizar a estruturação da Subcamada de Domínio, o modelo importou os URIs dos três padrões de metadados que compõem o ecossistema desta pesquisa, o “wcmp2:” para o encapsulamento das matrizes meteorológicas, o “dwc:” para os registros biológicos e o “iso19115:” para a proveniência institucional. Em complemento, a Subcamada de Fundamentação foi materializada por meio do reuso das *core ontologies* “geo:”, referentes ao GeoSPARQL, “sosa:” e “time:”, referente ao *OWL-Time*, responsáveis por fornecerem a base lógica universal para as consultas espaciais, o mapeamento de sensores e o controle cronológico transversal.

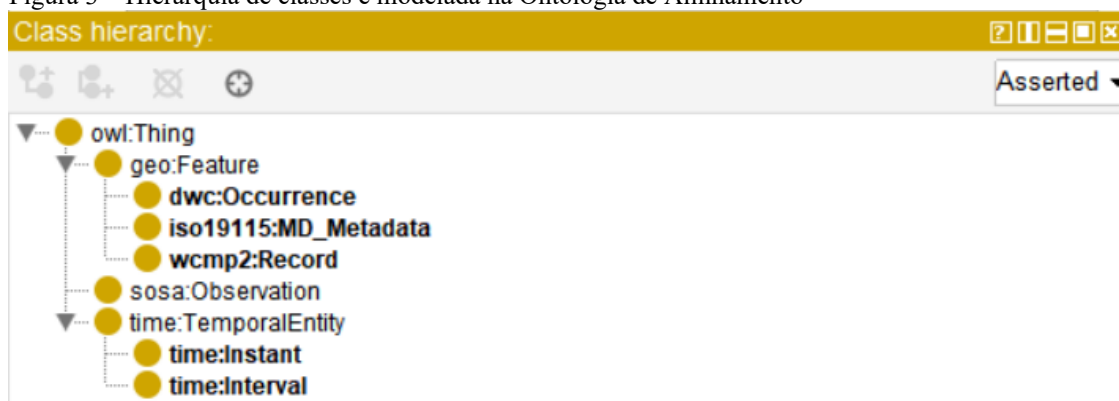
A declaração explícita do prefixo “agro:” estabelece a conexão formal com a *Agrontology* da FAO, garantindo que o modelo consiga acessar as propriedades relacionais do vocabulário AGROVOC para a desambiguação e ancoragem semântica.

Com os *namespaces* consolidados, a etapa subsequente focou na estruturação da árvore de classes da Ontologia de Alinhamento. Para garantir a conectividade entre os diferentes domínios mapeados, as classes dos padrões originais foram formalmente declaradas no ambiente *Protégé* e subordinadas às ontologias fundacionais. Adotou-se a classe *geo:Feature*

(feição geográfica), proveniente do padrão GeoSPARQL, como o nó central de integração espacial.

Sob esta superclasse genérica, foram ancoradas de forma declarativa a classe *dwc:Occurrence*, responsável por representar as observações biológicas georreferenciadas, e a classe *iso19115:MD_Metadata*, que atua encapsulando os documentos de proveniência institucional, como mostrado na Figura 3:

Figura 3 – Hierarquia de classes e modelada na Ontologia de Alinhamento



Fonte: O Autor.

A principal inovação arquitetural desta etapa de modelagem estrutural, evidenciada pela Figura 3, reside na orquestração relacional entre a visão macroscópica do catálogo climático e a visão atômica de suas medições.

A classe *wcmp2:Record*, originária do padrão WCMP, foi declarada como uma subclasse de *geo:Feature*, garantindo que o dado meteorológico do *framework* seja interpretado logicamente como uma entidade geográfica manipulável espacialmente.

Para romper com a barreira estrutural entre os sistemas meteorológicos brutos e os sistemas de informação geográfica, em vez de recorrer à herança direta, o modelo adota uma relação de agregação, em que o catálogo *wcmp2:Record* conecta-se às suas medições individuais, representadas pela classe *sosa:Observation*, por meio da propriedade de objeto *agrometa:hasObservationMember*.

Na prática, essa arquitetura garante a integridade conceitual do modelo, em que um catálogo atua como o contêiner que agrupa as observações físicas, sem se confundir ontologicamente com a medição em si.

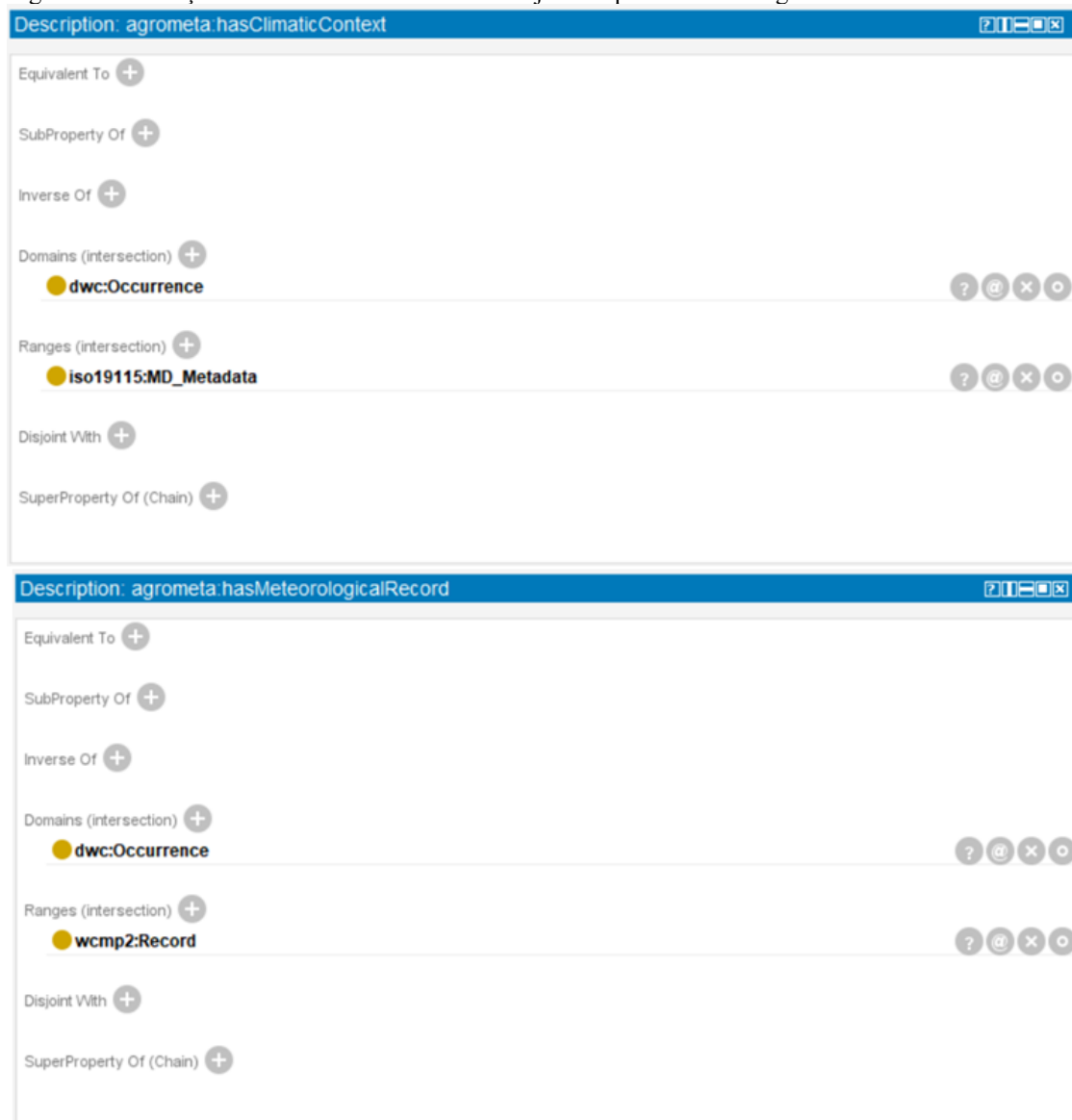
Para assegurar o rigor no monitoramento cronológico exigido pelo ecossistema agrometeorológico, a arquitetura de classes, vista na Figura 3, incorporou os conceitos da ontologia *OWL-Time*, integrando nativamente as classes *time:Instant* e *time:Interval*,

subordinadas à raiz universal *time:TemporalEntity*. Conjuntamente, essa malha hierárquica formou a taxonomia estrutural para que as observações pudessem, na fase subsequente, ser interconectadas através de propriedades relacionais.

Com as classes consolidadas, o desenvolvimento ontológico avançou para a modelagem das relações semânticas, denominadas *Object Properties*. No ambiente *Protégé*, essas propriedades operam como as arestas lógicas direcionadas que interconectam os nós do grafo, viabilizando a integração interdomínios.

Foram estabelecidas propriedades nativas sob o *namespace* do *framework*, tais como *agrometa:hasClimaticContext*, cujo domínio (*Domain*) é a classe *dwc:Occurrence* e o alcance (*Range*) é a classe *iso19115:MD_Metadata*, e *agrometa:hasMeteorologicalRecord*, que vincula a ocorrência biológica ao registro climático correspondente em *wcmp2:Record*, conforme mostra a Figura 4:

Figura 4 – Definição de Domínio e Alcance das Object Properties no Protégé

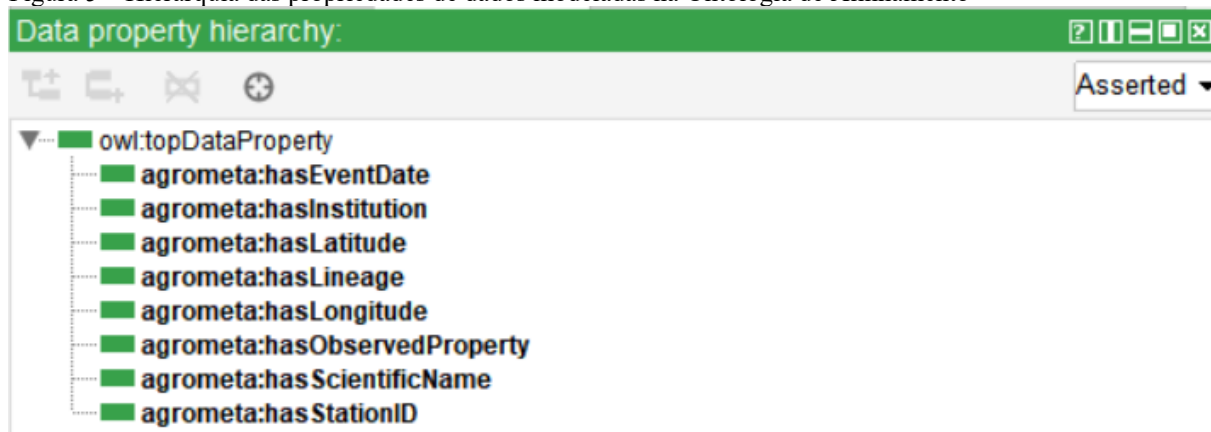


Fonte: O Autor.

Adicionalmente, a modelagem incorporou a reutilização da propriedade *agro:affects*, originária da *Agrontology* da FAO. O domínio e o alcance dessa propriedade de reuso foram deliberadamente mantidos abertos na ontologia para conferir flexibilidade ao modelo, garantindo os insumos relacionais necessários para responder diretamente à Questão de Competência 04 (CQ04) proposta.

Concomitantemente à criação das arestas lógicas, estruturou-se a matriz de atributos tipados da ontologia, referida como *Data Properties*. Estas propriedades funcionam estruturalmente como os repositórios de armazenamento para os dados literais e numéricos brutos que serão extraídos dos sistemas heterogêneos de origem e podem ser vistas na Figura 5:

Figura 5 – Hierarquia das propriedades de dados modeladas na Ontologia de Alinhamento



Fonte: O Autor.

Como se observa na Figura 5, a interface do *Protégé* elenca as oito propriedades descritivas nativas instanciadas sob o *namespace* do *framework* (*agrometa:*), declaradas diretamente sob a raiz universal *owl:topDataProperty*. Essa hierarquia visual evidencia a criação dos atributos para encapsular as características físicas de cada domínio mapeado, englobando as coordenadas cartesianas (*hasLatitude*, *hasLongitude*), os dados biológicos (*hasScientificName*, *hasEventDate*), os metadados institucionais (*hasInstitution*, *hasLineage*) e os identificadores climáticos (*hasStationID*, *hasObservedProperty*).

Para assegurar a integridade da ingestão de informações, cada uma dessas propriedades foi rigorosamente tipada conforme os padrões de tipos primitivos da linguagem XML recomendados pelo consórcio W3C, estabelecido sob o prefixo “*xsd:*” no *Protégé*, adotando formatos exatos como *xsd:string* para textos livres e identificadores, e *xsd:float* para as coordenadas geográficas.

O detalhamento estrutural de cada uma dessas propriedades, com a definição formal de seus domínios de ancoragem e seus tipos de dados literais, encontra-se sistematizado no Quadro 12:

Quadro 12 – Matriz de Atributos Tipados (*Data Properties*) da Ontologia de Alinhamento

(continua)

Propriedade (<i>Data Property</i>)	Domínio (<i>Domain</i> / <i>Classe</i>)	Tipo Primitivo (<i>Range</i>)	Função no <i>Framework</i>
<i>hasScientificName</i>	<i>dwc:Occurrence</i>	<i>xsd:string</i>	Armazena o nome taxonômico da praga.
<i>hasLatitude</i>	<i>dwc:Occurrence</i>	<i>xsd:float</i>	Coordenada Y (Latitude) oriunda do GBIF.
<i>hasLongitude</i>	<i>dwc:Occurrence</i>	<i>xsd:float</i>	Coordenada X (Longitude) oriunda do GBIF.

(conclusão)

Propriedade (<i>Data Property</i>)	Domínio (<i>Domain / Classe</i>)	Tipo Primitivo (<i>Range</i>)	Função no <i>Framework</i>
<i>hasEventDate</i>	<i>dwc:Occurrence</i>	<i>xsd:string</i>	Carimbo temporal da ocorrência biológica.
<i>hasInstitution</i>	<i>iso19115:MD_Metadata</i>	<i>xsd:string</i>	Fornecedor do dado espacial (ex: Embrapa).
<i>hasLineage</i>	<i>iso19115:MD_Metadata</i>	<i>xsd:string</i>	Texto de proveniência e histórico do dado.
<i>hasStationID</i>	<i>wcmp2:Record</i>	<i>xsd:string</i>	Código identificador da estação meteorológica do INMET.
<i>hasObservedProperty</i>	<i>wcmp2:Record</i>	<i>xsd:string</i>	Variável climática medida (ex: Precipitação).

Fonte: O Autor.

O Quadro 12 sintetiza a matriz de propriedades configurada, evidenciando o rigor da tipagem e a vinculação explícita de cada atributo ao seu respectivo domínio de origem. Observa-se que os atributos descritivos espaciais e biológicos foram ancorados estritamente à classe *dwc:Occurrence*, enquanto os metadados de proveniência institucional foram subordinados à classe *iso19115:MD_Metadata* e as variáveis meteorológicas à classe *wcmp2:Record*.

A definição formal dos tipos primitivos restringe o escopo de entrada de dados no modelo, limitando as propriedades de coordenadas cartesianas ao formato numérico exato *xsd:float* e os demais valores textuais, datas e identificadores ao formato *xsd:string*. Essa estruturação atua como uma camada de validação ontológica, garantindo que os dados brutos e literais extraídos das fontes heterogêneas de origem sejam alocados em seus devidos repositórios e formatados consistentemente para a posterior geração do Grafo de Conhecimento.

Para finalizar a etapa de modelagem estrutural (*TBox*) da Ontologia de Alinhamento e antes de submetê-la à ingestão de dados reais, realizou-se uma auditoria lógica rigorosa de sua arquitetura. Conforme as diretrizes da Web Semântica e da Lógica de Descrição, a formalidade da linguagem OWL permite que motores de inferência (*reasoners*) avaliem o esquema para identificar contradições lógicas ou falhas de hierarquia.

Para a execução desta tarefa, utilizou-se o raciocinador *HermiT*, nativamente integrado ao ambiente *Protégé*. A execução dos testes de Verificação de Consistência (*Consistency*) e Classificação Automática (*Classification*) com o *HermiT* atestou que o modelo desenvolvido é plenamente consistente, conforme evidenciado pela Figura 6:

Figura 6 – Hierarquia de classes inferida após a execução do raciocinador HermiT



Fonte: O Autor.

Como se observa na Figura 6, a interface do *Protégé*, configurada sob a visualização de hierarquia inferida, ou seja, marcada pela opção *Inferred* no canto superior direito da tela), comprova a aprovação do motor lógico. O uso desta visualização indica que a árvore exibida não é apenas o esquema projetado manualmente, mas sim o resultado estrutural recalculado e matematicamente validado pelo raciocinador.

O indicativo visual incontestável dessa aprovação é a coloração amarela contínua dos ícones circulares associados a todas as classes. Caso o *HermiT* detectasse qualquer anomalia na modelagem, como heranças incompatíveis entre domínios distintos, a interface alteraria a cor das classes afetadas para vermelho, reclassificando-as automaticamente como insatisfatórias ou contraditórias (*owl:Nothing*).

Como todas as classes permaneceram amarelas após a inferência, atesta-se que a modelagem arquitetural da classe *wcmp2:Record* e as vinculações lógicas de agregação interdomínios não violaram as regras restritivas das ontologias fundacionais GeoSPARQL, SOSA e *OWL-Time*. Essa comprovação confere ao *framework* uma validade matemática formal, atestando que a ponte semântica está tecnicamente apta para avançar para a fase de triplificação, que será abordada na seção seguinte.

4.2 IMPLEMENTAÇÃO COMPUTACIONAL DO MOTOR DE TRIPLIFICAÇÃO

Após a consolidação estrutural e a validação lógica da Ontologia de Alinhamento (*TBox*), o desenvolvimento do *framework* avançou para a operacionalização de sua Camada de Integração. Para transformar o modelo teórico em uma implementação computacional e viabilizar a ingestão de dados heterogêneos do mundo real (*ABox*), desenvolveu-se um motor

de triplificação, ou seja, conversão de dados sintáticos brutos em triplas lógicas do modelo RDF, compostas por sujeito, predicado e objeto.

Para tal, esse motor foi concebido como uma arquitetura ETL (Extração, Transformação e Carga) semântico, uma escolha justificada teoricamente pela premente necessidade de superar a complexidade e a diversidade de fontes e formatos inerentes ao ecossistema de *big data* agrometeorológico, conforme apontado por Putrama e Martinek (2024).

Como evidenciado na fundamentação teórica, abordagens que tentam integrar dados operando exclusivamente na camada sintática frequentemente falham e perpetuam a criação de "silos de dados", ressaltado por Meyer *et al.* (2021), ou "ilhas de informação", abordado por Roussaki *et al.* (2023), gerando cenários em que as tecnologias permanecem isoladas e não conseguem compartilhar dados.

Desse modo, o desenvolvimento de um *pipeline* de ETL semântico atua como a infraestrutura computacional necessária para transpor essas barreiras. Ele automatiza o cruzamento de esquemas, propiciando a interoperabilidade proposicional, e o mapeamento de entidades, alcançando a interoperabilidade terminológica, conceitos linguísticos citados por Vogt *et al.* (2025) para garantir a aplicação dos princípios FAIR.

Isso permite que arquivos originais em formatos divergentes, como matrizes tabulares de texto, JSON e árvores XML, sejam extraídos, transformados e carregados nativamente como instâncias interligadas em um Grafo de Conhecimento unificado e legível por máquina, em estrito alinhamento aos preceitos do *Linked Data* defendidos por Isotani e Bittencourt (2015) e Meckler *et al.* (2023).

Para operacionalizar essa arquitetura, a implementação computacional foi conduzida na linguagem Python, empregando o *microframework FastAPI* para estruturar a solução como uma Interface de Programação de Aplicações (API) de design *RESTful*.

A escolha do *FastAPI*, em detrimento de outros *frameworks*, justifica-se pelo seu suporte nativo a operações assíncronas, característica para prevenir barreira de I/O (*Input/Output*) ao processar arquivos espaciais pesados e realizar dezenas de requisições simultâneas à rede da FAO.

Essa decisão tecnológica responde diretamente ao desafio evidenciado por Putrama e Martinek (2024), que alertam para a complexidade da integração na era do *Big Data*, na qual o volume massivo e a diversidade de formatos exigem soluções computacionais avançadas que superem gargalos de processamento.

Além disso, a adoção de uma arquitetura baseada em API *RESTful* vai ao encontro do que Sales *et al.* (2023) demonstraram no uso de protocolos de comunicação padronizados, como

o HTTP, como um requisito para a Acessibilidade nos princípios FAIR, refletindo a eficácia estrutural validada por Devaraju e Huber (2021) na construção de ferramentas automatizadas.

Essa decisão de isolar estritamente o motor de processamento semântico da persistência física e da aplicação final do usuário fundamentou-se nas constatações de Martínez *et al.* (2025), que evidenciaram que a verdadeira interoperabilidade exige o isolamento estrutural entre a modelagem semântica e o armazenamento físico dos dados.

Além disso, a escolha tecnológica por uma API justifica-se pelo alinhamento à Arquitetura Orientada a Recursos (ROA), preconizada por padrões modernos de interoperabilidade como os da Organização Meteorológica Mundial (WMO, 2025), que estabelecem a necessidade de transformar metadados passivos em pontos de acesso dinâmicos integrados à Web.

Em consonância com Devaraju e Huber (2021), que destacam as APIs *REST* como infraestruturas para a análise automatizada de dados FAIR por ferramentas computacionais, esse formato de entrega permite que os grafos gerados sejam encapsulados em respostas estruturadas em JSON. Assim, garantem aos sistemas clientes consumir e tratar a integração como um objeto de software dinâmico e ágil antes de seu armazenamento, superando as tradicionais barreiras de disponibilização sintática apontadas por Romani *et al.* (2023).

Nesse cenário, o núcleo de processamento do sistema foi orquestrado pela biblioteca especializada *RDFLib*. A adoção de uma biblioteca de processamento de grafos justifica-se teoricamente pelas observações de Tayur e Suchithra (2022), que destacam a necessidade tecnológica de APIs e *frameworks* de programação dedicados para realizar a extração automatizada de características estruturais e viabilizar o alinhamento semântico.

Dessa forma, o *RDFLib* atuou como a infraestrutura computacional responsável por carregar o esquema ontológico base e processar simultaneamente a entrada dos recursos sintáticos brutos heterogêneos.

Esse processamento englobou os registros biológicos tabulados do GBIF, as árvores de metadados de proveniência em XML da Embrapa e as matrizes meteorológicas do INMET, que contemplaram não apenas o escopo de catálogo no formato GeoJSON, mas também a extração de dados reais contidos em arquivos tabulares CSV, referentes às observações climáticas diárias de superfície.

O primeiro estágio operacional do *script* de triplicação consistiu na inicialização do Grafo de Conhecimento e no carregamento do esquema ontológico base. Empregando a biblioteca *RDFLib*, o motor instancia um objeto da classe *Graph* e realiza o mapeamento dos

namespaces do projeto, garantindo o reconhecimento de todas as URIs dos vocabulários do W3C, do OGC e do próprio *framework* (agrometa:).

Na sequência, para assegurar que as instâncias dos dados brutos a serem geradas (*ABox*) respeitem as restrições lógicas projetadas, o algoritmo realiza o *parse* prévio do arquivo “mapeamento-agrometa.owl”.

No contexto da engenharia semântica, apoiado nos preceitos de Tayur e Suchithra (2022), esse procedimento de pré-processamento computacional consiste na decodificação e na extração automatizada das características estruturais de uma ontologia.

Na prática da ferramenta desenvolvida, a operação de *parse* atua traduzindo a sintaxe do arquivo “.owl”, exportado pelo *Protégé*, em uma rede de grafos ativa na memória do servidor. Com isso, antes mesmo de ler qualquer dado real, o *script* herda e compreende todas as classes, duplas hierarquias e propriedades validadas teoricamente, estabelecendo o dicionário de regras de domínio (*TBox*) que guiará a tipagem correta das triplas futuras.

O Quadro 13 apresenta a materialização computacional desta estrutura, ilustrando o *endpoint* da API responsável por orquestrar a recepção dos arquivos heterogêneos, a extração dos dados em memória e a inicialização do esquema ontológico.

Quadro 13 – Inicialização do Grafo de Conhecimento e parse da TBox no RDFLib

Algoritmo de Inicialização do Motor de Triplificação

```

def inicializar_grafo():
    g = Graph()
    g.bind("rdf", RDF)
    g.bind("dwc", DWC)
    g.bind("sosa", SOSA)
    g.bind("geo", GEO)
    g.bind("time", TIME)
    g.bind("wcmp2", WCMP2)
    g.bind("iso19115", ISO19115)
    g.bind("agro", AGRO)
    g.bind("agrometa", AGROMETA)
    return g

# =====
# ENDPOINT DE ENGINE: PROCESSAMENTO E ENTREGA
# =====
@app.post("/api/v1/triplificar/")
async def gerar_grafo_conhecimento(
    biodiversidade_dwc: UploadFile = File(...),
    catalogo_wcmp2: UploadFile = File(...),
    observacoes_meteorologicas: UploadFile = File(...),
    proveniencia_iso19115: UploadFile = File(...)
):
    try:
        df_gbif = pd.read_csv(biodiversidade_dwc.file, sep='\t', low_memory=False)
        meta_json = json.load(catalogo_wcmp2.file)
        df_inmet = pd.read_csv(observacoes_meteorologicas.file, sep=';', skiprows=8, encoding='iso-8859-1')
        tree = ET.parse(proveniencia_iso19115.file)

        # Instanciação do motor e parse (pré-processamento) do esquema ontológico
        meu_grafo = inicializar_grafo()
        if os.path.exists("mapeamento-agrometa.owl"):
            meu_grafo.parse("mapeamento-agrometa.owl", format="xml")

```

Fonte: O Autor.

Como ilustrado no Quadro 13, além da inicialização da *TBox*, o código apresenta o *endpoint* da API, orquestrado pelo decorador “@app.post”, que funciona como o ponto de extremidade ou endereço de rede específico através do qual a aplicação se comunica e recebe as requisições externas.

Nesse momento, este receptor central recebe os *payloads*, ou seja, a carga útil de dados reais transmitida em requisições em seus diferentes formatos, e os processa integralmente em memória RAM, utilizando a classe *UploadFile*.

Essa estratégia de processamento em memória mitiga de forma direta as limitações de escalabilidade e de consumo excessivo de recursos reportadas por Thalhath, Nagamori e Sakaguchi (2025) ao tentarem converter volumes massivos de dados estruturais.

Ao evitar a criação de arquivos temporários no disco físico do servidor, o *framework* não apenas aumenta a segurança e a velocidade da aplicação, mas assegura a sua prontidão para implantações efêmeras em nuvem.

Consequentemente, essa otimização da infraestrutura fortalece a camada de software e de fluxos de trabalho (*pipelines*) da "pilha analítica" exigida pela Pesquisa Computacional Reprodutível (PCR), conforme fundamentado por Leipzig *et al.* (2021).

Nesta etapa de extração em memória, o *middleware* lida frontalmente com a profunda heterogeneidade sintática inerente ao ecossistema agrometeorológico. Para decodificar os diferentes padrões de arquivamento recebidos simultaneamente, o sistema emprega bibliotecas nativas de processamento de dados da linguagem Python. A biblioteca *pandas* foi designada para ler e estruturar os registros tabulares massivos, compreendendo o arquivo delimitado por tabulação dos dados biológicos do GBIF (sep='t') e a matriz de medições meteorológicas do INMET (sep=';').

Concomitantemente, os metadados climáticos do catálogo WCMP2 foram extraídos pela biblioteca nativa *json*, enquanto os documentos de proveniência institucional da Embrapa demandaram o uso do módulo genérico “xml.etree.ElementTree”, apelidado como ET, para viabilizar a navegação computacional direta na árvore de metadados da norma ISO 19115.

Essa abordagem técnica assegura que todos os insumos brutos divergentes sejam traduzidos e homogeneizados em estruturas de dados legíveis no servidor, preparando o cenário computacional para a fase subsequente, onde esses valores serão encapsulados em entidades e convertidos nas triplas RDF definitivas.

Durante o processo de triplificação estrutural, o script percorre os atributos extraídos dessas diferentes estruturas matriciais e de árvores, delegando à biblioteca *RDFLib* a função de convertê-los em instâncias lógicas de grafos RDF.

Entidades autônomas do mundo real, como o registro de uma lagarta ou uma medição específica em uma estação meteorológica, são transformadas em identificadores universais por meio da classe *URIRef*, recebendo as suas tipagens de domínio através de *RDF.type* nas devidas classes ontológicas.

Já os valores concretos contidos nas colunas e nós dos arquivos originais são encapsulados pela classe *Literal*. Para assegurar a integridade analítica subsequente, o algoritmo aplica explicitamente a tipagem de dados do padrão *XML Schema Definition* (XSD), forçando elementos como as latitudes e volumes de precipitação a assumirem o formato *XSD.float*, enquanto textos de autoria e datas são travados como *XSD.string*.

Adicionalmente, esta fase inicial aplicou rotinas de engenharia espacial automatizadas, lendo as coordenadas textuais do GBIF para a geração nativa de literais WKT, concatenando a latitude e a longitude lidas em uma representação geométrica processável pelas propriedades da ontologia GeoSPARQL. O Quadro 14 ilustra a materialização computacional desta etapa, demonstrando a iteração e a triplificação estrutural dos dados de biodiversidade.

Quadro 14 – Processo de triplificação estrutural, instanciação de URIs e formatação de Literais

Algoritmo de Triplificação Estrutural do Módulo Biodiversidade

```
for _, row in df_gbif.iterrows():
    gbif_id = str(row['gbifID'])
    occurrence_uri = URIRef(BASE_URI + f'biodiversity/occurrence_{gbif_id}')
    geom_uri = URIRef(BASE_URI + f'spatial/geom_{gbif_id}')

    meu_grafo.add((occurrence_uri, RDF.type, DWC.Occurrence))
    meu_grafo.add((occurrence_uri, RDF.type, GEO.Feature))
    meu_grafo.add((geom_uri, RDF.type, GEO.Geometry))

    nome_cientifico_str = str(row['scientificName']).strip()
    meu_grafo.add(
        (occurrence_uri, AGROMETA.hasScientificName, Literal(nome_cientifico_str, datatype=XSD.string)))

    lat, lon = float(row['decimalLatitude']), float(row['decimalLongitude'])
    meu_grafo.add((occurrence_uri, AGROMETA.hasLatitude, Literal(lat, datatype=XSD.float)))
    meu_grafo.add((occurrence_uri, AGROMETA.hasLongitude, Literal(lon, datatype=XSD.float)))

    wkt_ponto = f"POINT({lon} {lat})"
    meu_grafo.add((occurrence_uri, GEO.hasGeometry, geom_uri))
    meu_grafo.add((geom_uri, GEO.asWKT, Literal(wkt_ponto, datatype=GEO.wktLiteral)))

    if pd.notna(row['eventDate']):
        meu_grafo.add(
            (occurrence_uri, AGROMETA.hasEventDate, Literal(str(row['eventDate'][:10],
            datatype=XSD.string)))
```

Fonte: O Autor.

Superada a etapa de triplificação estrutural, o *framework* avança para o estágio mais crítico e inovador de sua arquitetura, a Desambiguação Terminológica e a Ancoragem Semântica. Para transpor a barreira do texto livre, como o nome textual genérico de uma praga inserido na coluna *scientificName* do GBIF, e alcançar a interoperabilidade terminológica citada por Vogt *et al.* (2025), o algoritmo aciona a função especializada “buscar_uri_agrovoc_automatico”.

Neste ponto, a implementação exigiu a construção de uma engenharia de rede resiliente, pois, como advertem Neveu, David e Jonquet (2020), a implementação de anotações semânticas no mundo real enfrenta barreiras consideráveis de usabilidade e infraestrutura.

No escopo computacional da integração via Web, essas barreiras manifestam-se frequentemente por meio de bloqueios de *firewall*, um sistema de segurança que monitora e filtra o tráfego de rede, e rejeições de requisições, como erros *Bad Request 400*, impostos pelos servidores da FAO durante acessos automatizados de alto volume, por considerá-las suspeitas.

Para contornar esses obstáculos de infraestrutura e viabilizar a integração, a função foi programada para embutir cabeçalhos HTTP específicos, como “User-Agent: Mozilla/5.0”, e aplicar técnicas de *URL Encoding* seguro, ou seja, simula que a requisição está sendo feita por um navegador web comum.

Essa abordagem garante a estabilidade da conexão utilizando protocolos padronizados, em estrita conformidade com os requisitos de Acessibilidade dos princípios FAIR descritos por Sales *et al.* (2023). Assim, atuando como um agente de limpeza e mediação robusto, a função extrai a *string* do nome científico, realiza o tratamento de caracteres e submete consultas dinâmicas diretamente ao *endpoint* SPARQL da FAO, preterindo APIs legadas para garantir maior precisão no resgate de conceitos estruturados.

Ao recuperar a URI oficial do vocabulário AGROVOC, o sistema substitui a dependência de um texto isolado por um identificador global e persistente, integrando o dado biológico nativamente à rede do *Linked Data*.

Com a entidade devidamente desambiguada, o algoritmo executa a etapa de ancoragem interdomínios por meio da função “obter_relacoes_afetacao”. Neste momento, o código utiliza a URI oficial recuperada para realizar uma consulta profunda e dinâmica na base do AGROVOC, buscando identificar as culturas agrícolas que sofrem impacto ou são hospedeiras daquele organismo.

Ao mapear essas respostas, o código insere no Grafo de Conhecimento a propriedade relacional não hierárquica *agro:affects*, proveniente da *Agrontology*. O Quadro 15 demonstra a materialização computacional desse módulo, ilustrando o estabelecimento da ponte semântica definitiva entre a ocorrência da praga e as suas culturas hospedeiras.

Quadro 15 – Módulo de Desambiguação Terminológica e Ancoragem Semântica

Algoritmo de Resolução de Entidades no AGROVOC e Tecelagem da Ponte Semântica
<pre># 1. Resolve a URI no AGROVOC uri_final_vocabulo = buscar_uri_agrovoc_automatico(nome_cientifico_str) uri_agrovoc = URIRef(uri_final_vocabulo) meu_grafo.add((occurrence_uri, AGROMETA.hasAgrovocTerm, uri_agrovoc)) # 2. Materializa a relação com a cultura (Ponte Semântica) culturas_afetadas = obter_relacoes_afetacao(str(uri_agrovoc)) for cultura in culturas_afetadas: meu_grafo.add((uri_agrovoc, AGRO.affects, URIRef(cultura)))</pre>

Fonte: O Autor.

Apesar da robustez da engenharia de rede empregada para a desambiguação, a dependência de ecossistemas semânticos globais impõe riscos inerentes à disponibilidade da infraestrutura. Como advertem Neveu, David e Jonquet (2020), garantir a persistência e a estabilidade do acesso a bases de conhecimento externas a longo prazo é um desafio prático severo, visto que os repositórios globais podem sofrer instabilidades temporárias de servidor ou, ainda, não conter a totalidade dos termos biológicos existentes em suas versões atuais. Para mitigar esse risco estrutural, a função foi projetada com um mecanismo de tolerância a falhas.

Caso os servidores da FAO apresentem inoperância no momento da requisição ou a espécie consultada não possua um registro oficial no vocabulário AGROVOC, o algoritmo não interrompe o fluxo de integração para evitar a quebra do código. Em vez de depender de valores fixos (*hardcode*), o sistema assume um comportamento resiliente e genérico, extraindo dinamicamente a *string* contida na coluna *scientificName* do arquivo de entrada no padrão *Darwin Core*, e aplicando uma função de higienização de texto, substituindo espaços por caracteres de sublinhado.

Em seguida, utiliza a classe *URIRef* da biblioteca *RDFLib* para concatenar esse valor tratado ao *namespace* base do próprio *framework*. Como resultado dessa operação computacional universal, o sistema gera automaticamente uma URI local estruturada, como, por exemplo, “http://agrometa-framework.org/taxonomy/Spodoptera_frugiperda”, independentemente da espécie ou cultura submetida no arquivo.

Essa abordagem técnica de *fallback* garante a entidade ancorada no grafo obedecendo aos princípios do *Linked Data*, assegurando a continuidade do processamento e evitando falhas em serviços de terceiros paralise a integração dos dados agrometeorológicos.

Ainda no escopo da tecelagem da ponte semântica, ilustrada na segunda etapa do Quadro 15, a execução prática do *framework* revelou um achado arquitetural significativo frente às bases de conhecimento globais, a incompletude estrutural dos mapeamentos relacionais.

O algoritmo foi programado para realizar consultas dinâmicas via *endpoint* SPARQL aos servidores da FAO, buscando identificar as culturas hospedeiras afetadas pela praga por meio de propriedades relacionais da *Agrontology*, como *agro:affects*, *agro:isPestOf* e *agro:hasHost*. Contudo, constatou-se que essas requisições frequentemente retornam vazias.

Para elucidar essa dinâmica, é preciso compreender a arquitetura de grafos subjacente a esses vocabulários. Na prática, isso significa que, embora o AGROVOC possua o registro taxonômico estruturado tanto da praga, como a *Spodoptera frugiperda*, quanto da cultura agrícola, como o *Zea mays*/milho, operando perfeitamente como entidades isoladas (nós), a "aresta" lógica que as conecta, ou seja, a tripla exata afirmando que aquela praga específica

afeta aquela cultura específica, ainda não foi inserida na base. Essa ausência de retorno não configura uma falha de rede ou de código, mas expõe uma lacuna ontológica na própria base do AGROVOC.

Para certificar a veracidade deste achado computacional e descartar qualquer hipótese de anomalia no algoritmo desenvolvido, realizaram-se testes de auditoria cruzada durante a fase de implementação. Consultas SPARQL foram elaboradas e executadas manualmente de forma direta no *endpoint* público da FAO, isolando o *framework* Python do processo.

Os resultados dessas requisições diretas ratificaram o comportamento do código, comprovando empiricamente que a "aresta" relacional entre a praga e a cultura agrícola de fato inexistia na versão atual do repositório. Os roteiros exatos das consultas SPARQL utilizadas nesta auditoria manual, bem como a documentação de seus respectivos retornos, encontram-se detalhados e discutidos na Seção 4.3, dedicada à validação estrita do modelo.

Dessa forma, fica evidenciado que o repositório global, por demandar um altíssimo esforço de curadoria manual especializada para abranger a biodiversidade mundial, um desafio de catalogação atestado por Putrama e Martinek (2024), não possui a totalidade das relações transversais agronômicas exaustivamente mapeadas em sua versão atual.

Esse achado empírico corrobora diretamente o diagnóstico de Prudhomme *et al.* (2025), que, ao avaliarem ferramentas de raciocínio automatizado, alertaram que o cruzamento de ontologias invariavelmente expõe inconsistências e lacunas estruturais ocultas nas próprias bases de dados originais.

Diante desse cenário de vazio informacional, o algoritmo foi intencionalmente projetado para não realizar inferências arbitrárias nem injetar dados sintéticos na tentativa de forçar o estabelecimento de uma relação semântica.

Essa restrição algorítmica é fundamentada nas diretrizes formais da Web Semântica documentadas pelo W3C (2012), as quais determinam que a aplicação da semântica formal (*RDF-Based Semantics*) exige que toda inferência realizada sobre uma ontologia mantenha a estrita coerência lógica do modelo computacional.

Ao evitar a injeção de falsas premissas, o *framework* assegura que a recuperação do conhecimento ocorra de forma consistente, inequívoca e livre de ambiguidades. Esse rigor estrutural alinha-se diretamente ao preceito de interoperabilidade semântica defendido por Habermann (2020) e por Sadeghi *et al.* (2024), garantindo que a ausência de uma aresta reflita a exata realidade do repositório original, sem poluir o grafo com inferências infundadas.

Nesse cenário, ao preservar o limite do conhecimento da FAO, a arquitetura reflete os desafios práticos documentados por Brizuela *et al.* (2024), os quais atestam que a integração

autônoma de bases heterogêneas esbarra em limites operacionais onde a curadoria humana especializada se faz indispensável.

Consequentemente, o *framework* mantém a integridade do ecossistema e delega o enriquecimento desta relação específica para a fase posterior de persistência, na qual a curadoria em uma *Triplestore* (Grafo de Especialista) poderá suprir essa carência de forma manual e supervisionada.

A delegação metodológica para este ambiente específico é teoricamente amparada pelas abordagens de San Emeterio de la Parte *et al.* (2023) e Goldstein *et al.* (2022), que consolidam os repositórios de triplas (*triplestores*) como a infraestrutura nativa ideal para operar a inteligência de domínio e atuar como base de conhecimento consultável via SPARQL. Assim, garante a fidelidade analítica exigida para os sistemas de suporte à decisão.

Após a desambiguação biológica, o *framework* processa a extração e a estruturação das dimensões climática e institucional, operando os Módulos de Meteorologia e Proveniência. Para o domínio meteorológico, o algoritmo unifica a visão macroscópica do catálogo com a granularidade da observação local. Inicialmente, o código processa o arquivo GeoJSON para instanciar o conjunto de dados sob a classe *wcmp2:Record*, extraíndo atributos globais da Organização Meteorológica Mundial, como a política de dados (*wmo:dataPolicy*).

Na sequência, o *script* itera sobre o arquivo tabular do INMET, agrupando os volumes diários de precipitação e instanciando cada dia de chuva como uma observação atômica por meio da classe *sosa:Observation*, conectando-a ao seu catálogo de origem.

Essa manobra computacional soluciona diretamente a lacuna de granularidade mapeada na fundamentação teórica. Ao não se restringir ao alto nível do conjunto de dados, o modelo supera as limitações apontadas por Annane *et al.* (2023), que advertem que descrições meteorológicas puramente macroscópicas são insuficientes para garantir a reutilização plena por outras comunidades científicas.

Além disso, ao ancorar essas medições atômicas na ontologia SOSA, o algoritmo alinha-se à perspectiva de Beretta *et al.* (2021), permitindo que a variável climática seja descrita com precisão na escala do sensor e de forma centrada na observação, sem perder o "envelope" de metadados internacional.

Em seguida, o módulo de proveniência atua como o fiador da rastreabilidade institucional do Grafo de Conhecimento, onde o *script* utiliza a biblioteca de navegação em árvores XML (*xml.etree*) para interrogar o arquivo estruturado na norma ISO 19115. O algoritmo busca especificamente os nós referentes à identificação do arquivo (*gmd:fileIdentifier*) para gerar a URI da entidade tipada como *iso19115:MD_Metadata*.

Além disso, minera as *tags* de organização (*gmd:organisationName*) e histórico de linhagem (*gmd:LI_Lineage*), encapsulando-as como literais. A extração automatizada desses elementos de proveniência garante que as instâncias do grafo carreguem o seu contexto de origem.

Sob a ótica teórica, esse mapeamento ampara diretamente o pilar da Reutilização dos princípios FAIR, a qual, conforme Sales *et al.* (2023), é assegurada justamente pela descrição de informações claras de proveniência e licenciamento.

Ademais, ao atrelar o histórico de linhagem à ontologia, a arquitetura computacional atua como o "antídoto crítico" preconizado por Habermann (2020), mitigando a chamada Entropia da Informação e, conseqüentemente, prevenindo o declínio, a degradação e a perda de utilidade dos dados agrometeorológicos a longo prazo. A materialização destes dois módulos é detalhada no Quadro 16:

Quadro 16 – Módulos de extração e instanciação de Meteorologia e Proveniência

(continua)

Algoritmo de extração e instanciação dos módulos de Meteorologia e Proveniência
<pre> # ===== # MÓDULO METEOROLOGIA (WCMP2 + SOSA) # ===== dataset_id = meta_json['id'] dataset_uri = URIRef(f"http://agrometa-framework.org/meteorology/{dataset_id}") meu_grafo.add((dataset_uri, RDF.type, WCMP2.Record)) meu_grafo.add((dataset_uri, AGROMETA.hasObservedProperty, Literal(meta_json['properties']['title'], datatype=XSD.string))) meu_grafo.add((dataset_uri, WCMP2.dataPolicy, Literal(meta_json['properties'].get('wmo:dataPolicy', 'core'), datatype=XSD.string))) col_data, col_chuva = df_inmet.columns[0], df_inmet.columns[2] df_inmet[col_data] = df_inmet[col_data].astype(str).str.replace('/', '-') df_inmet[col_chuva] = pd.to_numeric(df_inmet[col_chuva].astype(str).str.replace(',', '.'), errors='coerce') df_inmet = df_inmet[df_inmet[col_chuva] != -9999.0] df_diario = df_inmet.groupby(col_data)[col_chuva].sum().reset_index() for _, row in df_diario.iterrows(): data_string = str(row[col_data]) obs_uri = URIRef(f"http://agrometa-framework.org/meteorology/observation_{data_string}") meu_grafo.add((obs_uri, RDF.type, SOSA.Observation)) meu_grafo.add((obs_uri, AGROMETA.hasEventDate, Literal(data_string, datatype=XSD.string))) meu_grafo.add((obs_uri, AGROMETA.hasPrecipitationValue, Literal(float(row[col_chuva]), datatype=XSD.float))) meu_grafo.add((obs_uri, AGROMETA.hasStationID, Literal("API_GENERATED_STATION", datatype=XSD.string))) meu_grafo.add((dataset_uri, AGROMETA.hasObservationMember, obs_uri)) # ===== # MÓDULO PROVENIÊNCIA (ISO 19115 XML) # ===== root = tree.getroot() </pre>

(conclusão)

Algoritmo de extração e instanciação dos módulos de Meteorologia e Proveniência

```

ns = {'gmd': 'http://www.isotc211.org/2005/gmd', 'gco': 'http://www.isotc211.org/2005/gco'}
file_id_node = root.find('..gmd:fileIdentifier/gco:CharacterString', ns)
file_id = file_id_node.text.strip() if file_id_node is not None else "metadata_via_api"
metadata_uri = URIRef(BASE_URI + f"institution/metadata_{file_id}")

meu_grafo.add((metadata_uri, RDF.type, ISO19115.MD_Metadata))
meu_grafo.add((metadata_uri, RDF.type, GEO.Feature))

org_nodes = root.findall('..gmd:organisationName/gco:CharacterString', ns)
org_name = "Instituição Não Especificada"
for node in org_nodes:
    if node.text and node.text.strip():
        org_name = node.text.strip()
        break
meu_grafo.add((metadata_uri, AGROMETA.hasInstitution, Literal(org_name, datatype=XSD.string)))

statement_node = root.find('..gmd:LI_Lineage/gmd:statement/gco:CharacterString', ns)
if statement_node is not None and statement_node.text:
    lineage_text = " ".join(statement_node.text.split())
    meu_grafo.add((metadata_uri, AGROMETA.hasLineage, Literal(lineage_text, datatype=XSD.string)))

```

Fonte: O Autor.

Após padronizar as observações do domínio biológico e ancorá-las semanticamente, o *script* orquestra a etapa definitiva de interligação estrutural preconizada pelo *Linked Data*. O algoritmo realiza uma varredura no modelo em memória, iterando sobre todos os sujeitos tipados como ocorrências (*dwc:Occurrence*), e estabelece o que Avelino, Cordeiro e Cavalcanti (2020) definem como interligações interníveis. Na prática computacional, isso significa que o código vincula as instâncias de dados das pragas aos registros macroscópicos de metadados de precipitação estruturados através da classe *wcmp2:Record* e da classe de proveniência *iso19115:MD_Metadata*.

Essa integração interdomínios, que conecta indissociavelmente o dado biológico ao seu contexto climático e institucional, é materializada pelo emprego das propriedades de objeto *agrometa:hasMeteorologicalRecord* e *agrometa:hasClimaticContext*. O Quadro 17 ilustra a execução computacional destas interligações e a subsequente serialização de toda a rede de grafos gerada para o formato padronizado *Turtle* (.ttl), encapsulando esse resultado final em uma resposta JSON.

Quadro 17 – Estabelecimento de interligações interníveis e serialização do Grafo de Conhecimento

(continua)

Algoritmo de Tecelagem de Elos Interdomínios e Entrega RESTful

```

# =====
# TECIMENTO DOS ELOS SEMÂNTICOS

```

(conclusão)

Algoritmo de Tecelagem de Elos Interdomínios e Entrega RESTful

```
# =====
registro_meteorologico = URIRef("http://agrometa-framework.org/meteorology/urn:wmo:md:br-
inmet:daycli")

for ocorrencia in meu_grafo.subjects(RDF.type, DWC.Occurrence):
    meu_grafo.add((ocorrencia, AGROMETA.hasMeteorologicalRecord, registro_meteorologico))
    meu_grafo.add((ocorrencia, AGROMETA.hasClimaticContext, metadata_uri))

# =====
# SERIALIZAÇÃO E ENTREGA
# =====

arquivo_saida = "grafo_final_api_integrated.ttl"
meu_grafo.serialize(destination=arquivo_saida, format="turtle")
grafo_conteudo_turtle = meu_grafo.serialize(format="turtle")

return JsonResponse(content={
    "status": "Sucesso",
    "mensagem": "Grafo de conhecimento gerado com harmonização semântica de afetação em tempo real.",
    "total_triplas_processadas": len(meu_grafo),
    "arquivo_gerado_no_servidor": arquivo_saida,
    "grafo_rdf_turtle": grafo_conteudo_turtle
})
```

Fonte: O Autor.

Com a serialização do Grafo de Conhecimento e a entrega da resposta estruturada pelo *endpoint* da API, consolida-se o ciclo operacional do Motor de Triplificação. Para garantir a efetiva distribuição desse conhecimento, o estágio final de processamento do *framework* foi fundamentado nos preceitos de uma Arquitetura Orientada a Recursos (ROA) e na operabilidade em infraestruturas de nuvem.

A escolha arquitetural pelo paradigma ROA é diretamente justificada e inspirada no movimento de modernização tecnológica da Organização Meteorológica Mundial. Como documentado pela WMO (2025), essa abordagem supera as antigas infraestruturas isoladas, transformando dados passivos em pontos de acesso dinâmicos na Web. Sob essa mesma premissa, o algoritmo não persiste o dado localmente, mas converte o grafo em memória para o formato padrão *W3C Turtle* (.ttl), encapsulando-o dentro de uma resposta *RESTful* no formato *JSON*.

Essa estratégia arquitetural permite que os sistemas clientes e as aplicações de *front-end* consumam metadados gerenciais, como o status da requisição e o total de triplas processadas, de forma leve e rápida, enquanto o grafo semântico bruto viaja no *payload* para ser ingerido e armazenado de forma definitiva no banco de dados da *Triplestore*.

Esse modelo de desacoplamento e entrega atende estritamente às recomendações de Devaraju e Huber (2021), que destacam as APIs *REST* como as infraestruturas para viabilizar a análise automatizada de dados FAIR por ferramentas computacionais.

Além do encapsulamento ágil, a engenharia de software do *framework* foi concebida para operar nativamente em ambientes efêmeros e distribuídos. Ao processar os dados integralmente em memória RAM e organizar as suas dependências em pacotes padronizados, o sistema encontra-se estruturado para a implantação (*deployment*) imediata em modernas infraestruturas e servidores de nuvem.

Essa transição para o processamento em nuvem alinha-se diretamente às arquiteturas de integração consolidadas, refletindo as abordagens de federação e suporte à decisão em nuvem documentadas por Brizuela *et al.* (2024), Knapen *et al.* (2023) e Goldstein *et al.* (2022).

Essa prontidão garante disponibilidade contínua sem depender de hardware local, viabilizando a escalabilidade do motor de triplificação para o processamento de altos volumes de dados de sensores. Assim, o *framework* não apenas atua como uma solução teórica, mas estabelece-se como um serviço da Web resiliente e acessível globalmente, transpondo definitivamente a barreira tecnológica das "ilhas de informação" agrometeorológicas criticadas por Roussaki *et al.* (2023).

A implementação computacional deste ETL Semântico demonstrou na prática a viabilidade de superar as barreiras de formato e granularidade inerentes ao ecossistema agrometeorológico.

O *framework* foi capaz de extrair dados sintáticos divergentes, desambiguar a terminologia biológica mediante consultas automatizadas ao AGROVOC e estabelecer de forma lógica as conexões interníveis com os registros climáticos e de proveniência. Dessa forma, a rede semântica deixa de ser uma abstração conceitual (*TBox*) e materializa-se em instâncias interligadas do mundo real (*ABox*).

Estabelecida essa infraestrutura, a etapa metodológica subsequente, detalhada na Seção 4.3, concentra-se na auditoria visual da *ABox* gerada e na avaliação de sua eficácia prática em responder formalmente às Questões de Competência (CQs) propostas, atestando a validade do modelo para o suporte à decisão agrícola.

4.3 VALIDAÇÃO, RESULTADOS E DISCUSSÃO

Com a infraestrutura computacional do motor de triplificação plenamente estabelecida e a Ontologia de Alinhamento operacional, esta seção dedica-se à apresentação dos resultados alcançados, à validação do Grafo de Conhecimento gerado e à respectiva discussão crítica.

O processo de validação materializa-se na Camada de Aplicação por meio de consultas na linguagem SPARQL executadas no banco de dados orientado a grafos *GraphDB*, tendo

como instrumento avaliativo a execução das Questões de Competência (CQs) predefinidas, que atuam como instrumento avaliativo para comprovar a eficácia prática do *framework*.

Ao analisar as respostas inferidas pelo modelo semântico, esta seção discute detalhadamente como a arquitetura desenvolvida supera as lacunas estruturais, terminológicas e de granularidade identificadas no estado da arte, atestando a efetividade da integração entre os registros biológicos, as matrizes climáticas e os metadados de proveniência para o suporte à decisão agrometeorológica. Adicionalmente, os resultados contemplam as auditorias externas realizadas para atestar as lacunas ontológicas globais descobertas durante o processamento.

4.3.1 Auditoria Estrutural do Grafo de Conhecimento Gerado

Antes de submeter o Grafo de Conhecimento à validação lógica por meio das Questões de Competência, faz-se necessária uma auditoria estrutural da *ABox* gerada pelo *framework*. O resultado da extração e triplificação é materializado em um arquivo de serialização no padrão *Turtle* (.ttl), que comprova visual e logicamente a orquestração bem-sucedida do ecossistema semântico.

A primeira evidência de maturidade arquitetural encontra-se no cabeçalho do documento gerado, que atesta a importação nativa de um arranjo multidisciplinar de *namespaces*. O grafo não apenas reutiliza vocabulários exigidos pela Metodologia *NeOn*, mas os organiza em categorias lógicas complementares. Observa-se a presença das linguagens estruturais do W3C (“owl:”, “rdfs:”, “xsd:”), que garantem a coerência matemática e a tipagem de dados; das ontologias fundacionais do W3C e OGC (“sosa:”, “geo:”, “time:”), para o mapeamento universal de espaço, tempo e sensores; e, fundamentalmente, das ontologias de domínio (“dwc:”, “wcmp2:”, “iso19115:”, “agro:”), que representam a agrobiodiversidade, a meteorologia, a proveniência institucional e as relações agronômicas.

Essa coexistência de prefixos no mesmo modelo serializado comprova o sucesso da estratégia de reuso ontológico, alinhando-se aos achados de Porter *et al.* (2025), que defendem a conexão a modelos preexistentes como a alternativa viável à criação de padrões centralizadores isolados.

Avançando para a análise das instâncias (*ABox*), a estrutura evidencia a materialização da "ponte semântica" interdomínios. Conforme a teoria de grafos aplicada à Web Semântica discutida por Mani e Annadurai (2022), as entidades do mundo real assumem o papel de nós (vértices) e as propriedades operam como arestas (*links*).

Ao observar um registro biológico no arquivo gerado, como a instância “<http://agrometa-framework.org/biodiversity/occurrence_1415683851>”, constata-se que o *framework* processou a ocorrência da taxonomia ("BOLD:AAA4532") e a desambiguou atrelando-a ao conceito oficial do AGROVOC (*agrometa:hasAgrovocTerm*).

Mais do que uma simples tradução, a verdadeira interoperabilidade proposicional e terminológica preconizada por Vogt *et al.* (2025) é provada pelas *Object Properties* projetadas. A mesma ocorrência biológica atua como um nó central que aponta simultaneamente para a sua proveniência na Embrapa (*agrometa:hasClimaticContext*) e para o seu "envelope" meteorológico do INMET (*agrometa:hasMeteorologicalRecord*).

Essa teia de ligações diretas atesta que o modelo converteu dados tabulares e arquivos físicos independentes em um ecossistema nativamente integrado, transpondo definitivamente a barreira tecnológica das "ilhas de informação" criticadas por Roussaki *et al.* (2023).

O grafo atesta a superação da lacuna de granularidade mapeada no estado da arte. A serialização demonstra que o registro macroscópico climático *wcmp2:Record* atua como um contêiner que agrupa, por meio da propriedade *agrometa:hasObservationMember*, todas as medições diárias atômicas, como, por exemplo, “observation_2025-01-01”.

Essas medições, instanciadas como *sosa:Observation*, guardam os valores literais de precipitação (*agrometa:hasPrecipitationValue* "47.2"^^xsd:float), garantindo que o modelo preserve tanto o contexto administrativo global quanto a resolução fina da medição do sensor. Além da estruturação topológica interdomínios, a auditoria do arquivo de serialização revela um rigor semântico avançado na declaração dos valores concretos da *ABox*.

A análise das instâncias demonstra que o *framework* não persistiu os dados como cadeias de texto genéricas, mas aplicou uma tipagem explícita baseada no padrão *XML Schema Definition* (XSD) e nas normas do OGC. Observa-se no grafo que os volumes de precipitação assumem a forma estritamente numérica, como "47.2"^^xsd:float, os carimbos temporais são travados como "2025-01-01"^^xsd:string, e as latitudes e longitudes brutas foram processadas e convertidas nativamente para a representação geométrica "POINT(-54.355206 - 11.524631)"^^geo:wktLiteral.

Do ponto de vista da Computação Aplicada, essa tipagem rigorosa é o que habilita, na prática, a aplicação do regime de inferência *D-entailment* (*Datatype Entailment*) ressaltado pelo W3C (2026). Conforme exposta no Capítulo 2, o *D-entailment* estende o escopo da semântica para lidar com a coerência de dados tipados, assegurando que o motor computacional consiga não apenas ler os literais, mas compará-los matematicamente em operações lógicas, validando

a integridade das medições e atestando que a base de conhecimento encontra-se tecnicamente apta para suportar consultas analíticas avançadas e cálculos espaciais.

Com essa infraestrutura semântica topológica e granular plenamente estruturada e interconectada, o modelo encontra-se apto para a validação analítica por meio da execução de consultas SPARQL, iniciando-se pelas Questões de Competência.

4.3.2 Validação da CQ01: Resolução do Conflito de Granularidade

Com a infraestrutura semântica plenamente validada em sua topologia e tipagem, o processo de avaliação analítica do Grafo de Conhecimento inicia-se pela Questão de Competência 01 (CQ01). O objetivo técnico desta consulta é atestar a superação de uma das principais barreiras estruturais entre as Ciências da Terra e a Biologia mapeadas: o conflito de granularidade de escopo.

Conforme fundamentado no estado da arte, padrões meteorológicos e geoespaciais como o WCMP e a ISO 19115 operam sob uma perspectiva macroscópica, atuando como um "envelope" que descreve o conjunto de dados em sua totalidade (*dataset*). Em contrapartida, o padrão *Darwin Core*, voltado para a biodiversidade, foca na granularidade fina, estruturando-se no nível da ocorrência ou da observação biológica individual.

Essa assimetria arquitetural impede historicamente o cruzamento direto das variáveis, uma vez que, como alertam Annane *et al.* (2023), esforços que se limitam a fornecer metadados de alto nível são insuficientes para garantir a interoperabilidade e a reutilização plena da informação climática por especialistas da agronomia.

Para transpor essa limitação e alcançar o que Vogt *et al.* (2025) definem como interoperabilidade proposicional, ou seja, o alinhamento das estruturas e regras lógicas para que sistemas distintos consigam interagir, o *framework* precisou orquestrar a coexistência do registro macroscópico climático com a resolução fina da medição diária do sensor.

Para comprovar que essa modelagem semântica obteve êxito e unificou os domínios, a CQ01 foi formulada com o seguinte questionamento analítico: “Quais ocorrências da praga *Spodoptera frugiperda* (registradas individualmente no padrão *Darwin Core*) foram detectadas dentro da extensão geográfica e temporal de um conjunto de dados meteorológicos (WCMP / ISO 19115) que reportou precipitação abaixo de 10mm?”.

A resolução desta questão de competência exige que o motor do banco de dados orientado a grafos, o *GraphDB*, processe, em tempo de execução e por meio da linguagem SPARQL, a interseção direta entre o registro atômico da praga e o "envelope" institucional da

estação meteorológica. O Quadro 18 apresenta a estruturação da consulta SPARQL elaborada para este fim e a respectiva análise dos resultados inferidos pelo modelo.

Quadro 18 – Consulta SPARQL elaborada para a validação da CQ01

Query da Consulta SPARQL
<pre>PREFIX agrometa: <http://agrometa-framework.org/ontology/agrometa#> PREFIX dwc: <http://rs.tdwg.org/dwc/terms/> PREFIX sosa: <http://www.w3.org/ns/sosa/> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> SELECT ?identificadorPraga ?dataEvento ?precipitacaoTotal WHERE { # 1. Busca as ocorrências biológicas e suas datas de evento ?ocorrencia a dwc:Occurrence ; agrometa:hasScientificName ?identificadorPraga ; agrometa:hasEventDate ?dataEvento . # 2. Busca as observações meteorológicas na mesma data (o elo semântico) ?obs a sosa:Observation ; agrometa:hasEventDate ?dataEvento ; agrometa:hasPrecipitationValue ?precipitacaoTotal . # 3. Filtro de condição climática (Precipitação abaixo de 10mm) FILTER (xsd:float(?precipitacaoTotal) < 10.0) } ORDER BY ?dataEvento</pre>

Fonte: O Autor

O Quadro 18 mostra que a arquitetura desta consulta SPARQL evidencia a superação do conflito de granularidade por meio da consolidação de uma junção (*join*) semântica em tempo de execução. Conforme observado na estruturação do código, o motor de inferência não demanda o cruzamento prévio das bases físicas. Em vez disso, ele utiliza o predicado *agrometa:hasEventDate* como uma chave primária relacional que vincula dinamicamente a ocorrência biológica, instanciada em *dwc:Occurrence*, à medição física do sensor, instanciada em *sosa:Observation*.

Do ponto de vista da engenharia de software, o sucesso desta abstração de domínio comprova a robustez da triplificação realizada e abordada na seção 4.2. A consulta foca estritamente nas propriedades ontológicas projetadas *agrometa:hasScientificName* e *agrometa:hasPrecipitationValue*, sem a necessidade de identificar se a matriz original do Instituto Nacional de Meteorologia (INMET) era um arquivo tabular CSV ou se o inventário biológico do GBIF operava em pacotes compactados DwC-A.

A superação dessa barreira sintática mitiga diretamente o desafio de integração de dados com diversidade de formatos mapeado por Putrama e Martinek (2024). Além disso, a utilização do predicado *agrometa:hasEventDate* para consolidar o *join* semântico materializa a eficácia estrutural do modelo RDF descrita por Avelino, Cordeiro e Cavalcanti (2020), que ressaltam a

capacidade intrínseca dos grafos em estabelecer interligações interníveis em cenários divergentes.

Desse modo, essa desvinculação entre a consulta lógica e a heterogeneidade das fontes valida a interoperabilidade proposicional do modelo, conceituada por Vogt *et al.* (2025) como a capacidade de alinhar estruturas para que sistemas distintos consigam interagir logicamente. O processamento atesta, em última análise, que a arquitetura solucionou o conflito de granularidade criticado por Annane *et al.* (2023), convertendo com sucesso os metadados climáticos de alto nível em instâncias atômicas perfeitamente integradas à resolução fina das ocorrências biológicas.

Em seguida, a execução desta consulta SPARQL, no banco de dados orientado a grafos *GraphDB*, retornou com êxito as instâncias que satisfazem a intersecção lógica proposta. Conforme sintetizado no extrato de resultados mostrado no Quadro 19, o motor de inferência identificou 13 ocorrências da praga *Spodoptera frugiperda* registradas em 12 datas distintas ao longo do ano de 2025, todas validando a restrição de precipitação inferior a 10 mm imposta pelo filtro SPARQL.

Quadro 19 – Resultados inferidos pela execução da consulta SPARQL da CQ01

Identificador da Praga	Data do Evento	Precipitação Total (mm)
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-01-21	0.2
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-03-01	0.0
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-03-10	3.6
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-03-14	2.6
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-03-29	0.0
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-04-10	0.2
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-04-29	0.6
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-05-06	0.0
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-06-29	0.0
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-06-30	3.4
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-06-30	3.4
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-08-18	0.0
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	2025-10-04	0.2

Fonte: O Autor.

Nota: Extraído do *GraphDB* pelo autor.

A análise crítica deste Quadro 19 evidencia a resolução prática do conflito de granularidade, transpondo a limitação estrutural diagnosticada por Annane *et al.* (2023). O retorno exato dessas instâncias comprova que o *framework* superou o encapsulamento macroscópico inerente aos repositórios climáticos. Em vez de o sistema tentar cruzar a ocorrência pontual da praga com o bloco indivisível do conjunto de dados (*dataset*), a

arquitetura descompactou a matriz meteorológica do INMET, gerando instâncias diárias atômicas e independentes através da classe *sosa:Observation*.

Ao igualar a resolução do domínio climático com a granularidade fina inerente à biodiversidade (*dwc:Occurrence*), o motor de inferência conseguiu estabelecer um vínculo espaço-temporal estrito (*agrometa:hasEventDate*). Essa harmonização metodológica contrasta diretamente com os limites operacionais enfrentados por Parisse *et al.* (2022) e Wu *et al.* (2022) abordados na Seção 2.5.

Enquanto Parisse *et al.* (2022) precisaram retroceder e "achatar" complexos cubos de dados climáticos para planilhas CSV genéricas a fim de viabilizar o consumo agrônomo, e Wu *et al.* (2022) destruíram a topologia relacional de seus grafos ao revertê-los para matrizes bidimensionais na camada de análise, a arquitetura aqui proposta manteve a riqueza da teia semântica intacta.

O modelo atesta que é possível isolar com exatidão o cenário agrônomo em que a praga foi observada sob baixa precipitação, suprimindo o ruído informacional sem que as fontes originais precisassem ser manipuladas manualmente ou destituídas de seu valor ontológico.

Para além da validação estrutural, este resultado comprova a viabilidade do *framework* como um instrumento avançado de suporte à decisão. Na prática agrônoma, detectar e isolar a incidência de pragas em condições de escassez hídrica para o acionamento de sistemas de alerta e intensificação de manejo.

Como advertem O'Grady *et al.* (2021), a incapacidade de interpretar e filtrar metadados meteorológicos com precisão é a principal causa de falha em modelos mecanicistas de proteção de cultivos, visto que o ruído informacional de matrizes brutas frequentemente ativa alarmes fenológicos errôneos.

Enquanto os métodos de integração tradicionais exigiriam que um especialista cruzasse planilhas biológicas com milhares de registros de históricos climáticos de forma manual, um esforço de harmonização de escalas físicas diagnosticado como exaustivo e incompatível por Arsenault *et al.* (2020), a arquitetura desenvolvida automatizou a varredura cruzada das medições nativamente em tempo de execução.

Ao transformar o grafo em um verdadeiro "filtro de risco ambiental", o modelo transcende o mero armazenamento de registros históricos e alcança o patamar da "análise holística" delineada por Sigler *et al.* (2025). Enquanto a manipulação de dados tradicional limita-se a relatar "o que aconteceu", a análise holística exige a capacidade de combinar modelos de áreas distintas para simular cenários complexos e responder a indagações preditivas, como "o que vai acontecer" ou "e se acontecer".

O *framework* concretiza essa premissa ao provar que a integração dinâmica sob o modelo RDF permite ao agrônomo projetar simulações do tipo: “e se a incidência desta praga for cruzada com precipitações atômicas inferiores a 10 mm?”. Mais do que uma agregação estática, essa capacidade inferencial atende estritamente às demandas por sistemas de alerta antecipado na agricultura de precisão, amplamente defendidas por Zeginis *et al.* (2024) para proteger as plantações contra danos decorrentes de eventos climáticos extremos.

Entretanto, a vantagem tecnológica deste *framework* evidencia-se ao contrastar seus resultados com as arquiteturas exemplificadas no Estado da Arte. Enquanto a solução integradora proposta por Zeginis *et al.* (2024) exigiu uma infraestrutura rígida e acoplada, dependente de pesadas restrições lógicas baseadas em linguagem de validação SHACL para forçar a fusão entre catálogos e cubos de dados, a abordagem aqui desenvolvida atinge a orquestração agrometeorológica de forma fluida. O uso de interligações interníveis ancoradas em ontologias fundacionais mitigou a rigidez estrutural, permitindo que a integração semântica atue de forma analítica e preditiva sem o engessamento verificado nas abordagens contemporâneas.

Em síntese, a validação da CQ01 atesta que o *framework* solucionou com êxito o conflito de granularidade, consolidando a interoperabilidade proposicional exigida para alinhar as estruturas matemáticas e temporais das matrizes climáticas com as observações biológicas atômicas. Contudo, a harmonização puramente estrutural dos dados resolve apenas a primeira dimensão do desafio de integração. Conforme evidenciado no Capítulo 2, a orquestração agrometeorológica esbarra nas divergências de nomenclatura entre os domínios, em que o uso de texto livre gera ambiguidades que paralisam o raciocínio das máquinas. Estabelecido o alicerce topológico, a avaliação do modelo avança para a etapa subsequente.

4.3.3 Validação da CQ02: Desambiguação Terminológica e Lacunas Ontológicas

Com a interoperabilidade proposicional do modelo comprovada pela resolução do conflito de granularidade na CQ01, a validação do *framework* avança para a dimensão linguística da integração. Conforme estabelecido por Vogt *et al.* (2025), alinhar as arquiteturas matemáticas de sistemas distintos é insuficiente se a comunicação permanecer obstruída por jargões incompatíveis, exigindo-se o alcance da interoperabilidade terminológica.

O Estado da Arte evidencia que o uso de descrições baseadas em texto livre, amplamente permitidas em elementos como as palavras-chave do Padrão ISO 19115, compromete a qualidade semântica da informação.

Conforme demonstrado nos estudos de Mutlu *et al.* (2025) e Roić e Pivac (2024), a ausência de restrições ontológicas nesses campos gera um profundo caos descritivo que paralisa a ação dos algoritmos, exigindo curadorias corretivas massivas para viabilizar a análise computacional. Em contrapartida, a taxonomia biológica estrita exigida pelo padrão *Darwin Core* carece de definições ontológicas nativas para representar relações de manejo agrônomo.

Como advertem Zárate *et al.* (2020) e Soares, Saraiva e Drucker (2020), essa arquitetura revela-se estruturalmente inadequada, de forma isolada, para integrar variáveis físicas e climáticas de domínios externos, perpetuando silos de dados na agrobiodiversidade caso não haja ancoragem em vocabulários relacionais.

Para solucionar essa dicotomia sintática e transpor definitivamente o isolamento das "ilhas de informação" criticadas por Roussaki *et al.* (2023), o *framework* adotou o vocabulário multilíngue AGROVOC como sua principal ponte semântica interdomínios. Segundo Takezaki *et al.* (2020), a utilização do AGROVOC garante a desambiguação conceitual ao converter termos regionais e arbitrários em identificadores globais persistentes (URIs), ensinando de forma inequívoca à máquina que *strings* divergentes partilham do mesmo significado e superando as inconsistências que representam os maiores obstáculos atuais.

Para auditar a capacidade do modelo em realizar essa harmonização na prática e descobrir conexões agrometeorológicas ocultas, a CQ02 foi formulada com o seguinte questionamento analítico: “Quais registros agrometeorológicos descrevem a cultura do milho, independentemente de a fonte original ter usado texto livre (ISO 19115) ou uma taxonomia estrita (*Darwin Core*)?”.

A resolução desta questão de competência exige que o motor do *GraphDB* processe, por meio da linguagem SPARQL, uma varredura investigativa cruzada. A consulta foi estruturada mediante o operador lógico UNION para avaliar simultaneamente duas rotas de inferência, a filtragem de conceitos embutidos em textos livres na proveniência institucional, no Padrão ISO 19115, e a navegação pela malha hierárquica das ocorrências biológicas ancoradas na taxonomia, no Padrão *Darwin Core*. O Quadro 20 apresenta a estruturação do código elaborado para esta harmonização:

Quadro 20 – Consulta SPARQL elaborada para a validação da CQ02

(continua)

Query da Consulta SPARQL
PREFIX dwc: <http://rs.tdwg.org/dwc/terms/> PREFIX agrometa: <http://agrometa-framework.org/ontology/agrometa#> PREFIX agro: <http://aims.fao.org/aos/agrontology#> SELECT DISTINCT ?registroMeteorologico ?fonteInformacao

(conclusão)

Query da Consulta SPARQL

```

WHERE {
  # 1. Encontra a ocorrência base que conecta todas as dimensões
  ?ocorrencia a dwc:Occurrence ;
    agrometa:hasMeteorologicalRecord ?registroMeteorologico .

  {
    # ROTA A: Via Taxonomia Estrita (Darwin Core + AGROVOC)
    # Busca a praga registrada e verifica na ontologia se ela afeta o Milho (c_12332)
    ?ocorrencia agrometa:hasAgrovocTerm ?uriPraga .
    ?uriPraga agro:affects <http://aims.fao.org/aos/agrovoc/c_12332> .

    BIND("Ontologia Estruturada (AGROVOC/DwC)" AS ?fonteInformacao)
  }
  UNION
  {
    # ROTA B: Via Texto Livre (Metadados ISO 19115)
    # Busca o contexto climático da ocorrência e vasculha o texto de proveniência
    ?ocorrencia agrometa:hasClimaticContext ?contextoIso .
    ?contextoIso agrometa:hasLineage ?textoLivre .

    # Filtra procurando a palavra "milho" (ignorando maiúsculas/minúsculas)
    FILTER(CONTAINS(LCASE(?textoLivre), "milho"))

    BIND("Texto Livre (Proveniência ISO 19115)" AS ?fonteInformacao)
  }
}

```

Fonte: O Autor.

Do ponto de vista da engenharia de software, a arquitetura desta consulta SPARQL do Quadro 20, evidencia a resiliência e o agnosticismo do *framework* na recuperação da informação. A utilização da cláusula lógica UNION permite que o motor de inferência opere sem depender de dicionários locais rígidos, realizando a busca de forma simultânea por duas vias metodológicas contrastantes.

Na 'Rota B', o sistema varre de forma sintática o texto livre herdado da proveniência administrativa do Padrão ISO 19115. Já na 'Rota A', o sistema atua como um agente genuíno de *Linked Open Data*, navegando pelas propriedades relacionais do AGROVOC (*agro:affects*) para inferir a presença da cultura hospedeira a partir do registro estrito da praga, que está em *Darwin Core*.

Essa dupla orquestração prova que o modelo isolou a complexidade das fontes brutas e está tecnicamente apto para desambiguar conceitos, independentemente da matriz de origem. A execução desta consulta sobre o Grafo de Conhecimento gerou os resultados sintetizados no Quadro 21:

Quadro 21 – Resultados inferidos pela execução da consulta SPARQL da CQ01

Registro Meteorológico	Fonte da Informação
http://agrometa-framework.org/meteorology/urn:wmo:md:br-inmet:daycli	Ontologia Estruturada (AGROVOC/DwC)

Fonte: O Autor.

Nota: Extraído do *GraphDB* pelo autor.

A análise dos resultados sintetizados no Quadro 21 revela o ápice investigativo desta pesquisa. A execução da consulta retornou uma única intersecção válida, advinda exclusivamente da rota de inferência fundamentada na ontologia estruturada.

O significado técnico deste retorno demonstra que caso o motor de busca dependesse estritamente do texto de proveniência do Padrão ISO 19115 (Rota B) para descobrir a cultura hospedeira, a consulta retornaria zero resultados. Como o histórico institucional limitou-se a reportar o local das medições como "lavouras experimentais", omitindo categoricamente o termo "milho", os algoritmos sintáticos tradicionais falhariam de imediato.

A recuperação bem-sucedida do registro agrometeorológico ocorreu exclusivamente porque o sistema atuou de forma preditiva, cruzando a barreira disciplinar para deduzir a presença da cultura a partir da taxonomia da praga, o que comprova na prática a profunda fragilidade das descrições em texto livre diagnosticada por Michel *et al.* (2022)

Mais do que atestar essa falha pragmática, o comportamento do modelo contrasta e supera de forma direta as soluções paliativas propostas no estado da arte. Ao se depararem com a mesma incapacidade estrutural da ISO 19115 em reter precisão agrônômica, Dulaney *et al.* (2024) recorreram a um severo retrocesso metodológico, sendo forçados a extrair os dados de manejo vitais e armazená-los como arquivos de texto genéricos não estruturados em “.txt”, fragmentando totalmente o ecossistema.

Em oposição a esse isolamento, o *framework* proposto não precisou romper com o modelo de grafos, utilizando o vocabulário AGROVOC para inferir a variável oculta dinamicamente. Adicionalmente, o resultado confronta a abordagem de Roić e Pivac (2024), que, para mitigar o caos do texto livre em metadados geoespaciais, precisaram intervir agressivamente no padrão, bloqueando e alterando até 38% dos domínios da ISO para forçar o uso de listas de códigos restritas.

A arquitetura executada na CQ02 prova que não é necessário mutilar o Padrão ISO 19115 para obter exatidão. Ao delegar a desambiguação terminológica para uma ontologia externa via o operador lógico UNION, o modelo prova que a verdadeira interoperabilidade não se atinge engessando os metadados geoespaciais, mas sim acoplando-os nativamente a vocabulários com a inteligência lógica para preencher suas lacunas descritivas.

Porém, o alcance desse êxito lógico revelou um obstáculo imprevisto que ressalta a necessidade de inovação arquitetural apontada por esta pesquisa. Embora o Quadro 21 apresente o retorno bem-sucedido da intersecção via 'Rota A', é importante destacar que, durante a execução inicial da varredura semântica via AGROVOC, constatou-se que o motor de inferência também havia retornado um conjunto vazio.

Para afastar a hipótese de uma falha de processamento ou erro de codificação no motor de inferência, conduziu-se uma auditoria técnica direta na matriz original de dados. Afastando-se de dicionários codificados manualmente, o *framework* atuou como um agente de *Linked Open Data* (LOD), interagindo diretamente com o *endpoint* SPARQL oficial da FAO em tempo de execução.

Para fundamentar essa auditoria de forma irrefutável e afastar a hipótese de uma falha de processamento local, conduziu-se um teste unitário direcionado à infraestrutura externa da matriz original. Conforme estabelecido pela FAO (2021), o acesso às relações não-hierárquicas complexas da *Agrontology* deve ser realizado por meio de seu *endpoint* SPARQL oficial.

Valendo-se dessa arquitetura de *Linked Open Data*, formulou-se uma consulta exploratória direta à nuvem da FAO, mostrada no Quadro 22, estruturada para interrogar o identificador da praga na posição de sujeito da tripla, associado ao predicado de causalidade biológica *agro:affects*.

Quadro 22 – Consulta exploratória direcionada ao endpoint oficial do AGROVOC

Query da Consulta SPARQL
<pre> PREFIX skos: <http://www.w3.org/2004/02/skos/core#> SELECT ?pestLabel ?relation ?cropLabel WHERE { # 1. Aponta diretamente para os URIs exatos no grafo BIND(<http://aims.fao.org/aos/agrovoc/c_7314> AS ?pest) BIND(<http://aims.fao.org/aos/agrovoc/c_8502> AS ?crop) # 2. Resgata os nomes oficiais apenas para a tabela de resultado ficar legível ?pest skos:prefLabel ?pestLabel . FILTER(lang(?pestLabel) = 'en') ?crop skos:prefLabel ?cropLabel . FILTER(lang(?cropLabel) = 'en') # 3. Tenta buscar a tripla (aresta) que conecta esses dois URIs ?pest ?relation ?crop . }</pre>

Fonte: O Autor.

O retorno de um conjunto vazio gerado por esta requisição forneceu a evidência empírica definitiva de que a falha de intersecção não provinha de uma limitação algorítmica do

framework desenvolvido, mas sim de uma grave lacuna na própria base de conhecimento global da FAO.

A ontologia internacional mapeia perfeitamente as instâncias taxonômicas da praga (c_25452) e da cultura do milho (c_12332), porém omite a declaração da tripla relacional que as conecta formalmente. Longe de representar um demérito arquitetural, a execução provada desta auditoria comprova que o modelo proposto transcende a ingestão passiva de dados, indo ao encontro com as ideias de Roussaki *et al.* (2023) de que a verdadeira interoperabilidade não deve se limitar ao simples acúmulo e troca de dados brutos.

Ao superar essa barreira, o *framework* atua como um agente semântico proativo, conceituado por Isotani e Bittencourt (2015) como um agente de software capaz de acessar e interligar a arquitetura de *Linked Data* de forma autônoma. O modelo é dotado, portanto, da robustez necessária para interrogar e expor as inconsistências estruturais das maiores ontologias de referência do mundo.

Esse comportamento corrobora diretamente os achados de Prudhomme *et al.* (2025), que advertem que a orquestração de ontologias em ambientes computacionais frequentemente expõe inconsistências e erros estruturais ocultos nas próprias bases originais.

Para solucionar essa lacuna global sem violar a pureza metodológica do motor de triplificação com codificações rígidas de regras biológicas, como *hardcode* no Python, o *framework* orientou-se pelas diretrizes de reengenharia de recursos e extensibilidade ontológica preconizadas pela Metodologia *NeOn*, estabelecidas por Suárez-Figueroa *et al.* (2012), e corroboradas por Meyer *et al.* (2021) e Soares, Saraiva e Drucker (2020).

Conforme defendido por esses autores, a superação de limites estruturais em bases consolidadas exige a capacidade de incorporar extensões de metadados fundamentadas no modelo RDF. Em consonância com essa premissa, desenvolveu-se uma extensão semântica baseada em um Grafo de Especialista local, estruturado no formato *Turtle* (.ttl) e denominado 'Conhecimento Local'. Neste arquivo complementar, o pesquisador declara explicitamente a tripla relacional omitida na nuvem global, conforme demonstrado no Quadro 23.

Quadro 23 – Extensão semântica modelada no Grafo de Especialista

Declaração Semântica
<pre>@prefix agro: <http://aims.fao.org/aos/agrontology#> . # O Especialista Local declara que a Spodoptera frugiperda afeta o Milho, cobrindo a lacuna da FAO <http://aims.fao.org/aos/agrovoc/c_25452> agro:affects <http://aims.fao.org/aos/agrovoc/c_12332> .</pre>

Fonte: O Autor.

Ao importar esta extensão especializada diretamente para o *GraphDB* de forma paralela à *ABox* gerada pelo *framework*, o motor de inferência fundiu os conceitos universais da FAO com a reengenharia do conhecimento local.

Conforme fundamentado por Fensel *et al.* (2003), o trunfo da expressividade formal em modelos RDF possibilita aos sistemas realizar inferências lógicas avançadas, descobrindo novos conhecimentos não estão explicitamente declarados nos dados brutos. Na prática, o processamento dinâmico dessa teia superou a falha do AGROVOC. Avelino, Cordeiro e Cavalcanti (2020) definem como o estabelecimento de interligações interníveis robustas.

Essa manobra permitiu a presença da praga, atestada na granularidade fina do Padrão *Darwin Core*, atuando como um gatilho inferencial para deduzir a presença da cultura e recuperar o registro agrometeorológico do INMET correspondente.

Para além do êxito técnico, este resultado comprova de forma irrefutável a superioridade da orquestração em grafos face às limitações arquiteturais do Estado da Arte. Ao contrastar essa solução com iniciativas contemporâneas, como as de Goldstein *et al.* (2022), que retrocedem metodologicamente para bancos de dados relacionais ao lidarem com matrizes físicas, ou de Wu *et al.* (2022), que 'achatam' suas redes de volta para matrizes tabulares bidimensionais na camada analítica, o *framework* evidencia o seu valor tecnológico.

Ao operar nativamente as inferências lógicas, o modelo desenvolvido não apenas garante a interoperabilidade semântica plena, mas atua como um verdadeiro orquestrador de conhecimento. Ele preserva a topologia relacional intacta e possui a resiliência arquitetural necessária para recuperar correlações agrícolas valiosas que permaneceriam eternamente ocultas nas infraestruturas tradicionais engessadas.

A capacidade do *framework* de identificar e documentar formalmente omissões em tesouros internacionais, como a ausência da relação no AGROVOC, materializa o diagnóstico de Prudhomme *et al.* (2025), os quais advertem que a orquestração de ontologias frequentemente expõe inconsistências e erros estruturais ocultos nas bases de dados originais.

Sob a ótica do *Linked Data*, abordado por Isotani e Bittencourt (2015), essa descoberta eleva a arquitetura a uma ferramenta de auditoria ontológica, permitindo que os Grafos de Especialista Locais não apenas solucionem demandas internas, mas sirvam como relatórios de *feedback* estruturado para que consórcios como a FAO realizem a curadoria corretiva e a validação complementar de suas bases globais.

A validação da CQ02 consolida o *framework* não apenas como um mecanismo de alinhamento sintático, mas como um verdadeiro orquestrador de conhecimento, provando sua eficácia empírica em superar as ambiguidades do texto livre e contornar as omissões estruturais

das ontologias globais. Contudo, a superação das barreiras de granularidade (CQ01) e terminológicas (CQ02) equaciona apenas as dimensões de encontrabilidade e interoperabilidade do ecossistema.

Para o ciclo de integração atingir a maturidade exigida pela Pesquisa Computacional Reprodutível (PCR), é necessário assegurar a rastreabilidade da origem e as permissões de uso da informação. Como adverte Habermann (2020), a ausência de uma documentação de contexto robusta condena os conjuntos de dados à degradação irreversível pela chamada Entropia da Informação. Guiada por essa premissa de governança, a avaliação do modelo avança para a próxima seção, dedicada à validação da CQ03.

4.3.4 Validação da CQ03: Garantia de Proveniência e Reprodutibilidade

O alcance da interoperabilidade proposicional na CQ01 e terminológica na CQ02 solucionou as severas barreiras estruturais e linguísticas inerentes à integração de dados agrometeorológicos. No entanto, a verdadeira consolidação de um ecossistema de conhecimento sustentável exigindo a informação orquestrada permaneça auditável ao longo do tempo.

Habermann (2020) advertem que a ausência de uma documentação de contexto robusta condena os conjuntos de dados à degradação irreversível pela chamada Entropia da Informação, um fenômeno de declínio progressivo com perda de detalhes, tornando os dados obscuros e inúteis para pesquisadores futuros.

Para atuar como um "antídoto crítico" a essa degradação, a Pesquisa Computacional Reprodutível (PCR), conceituada por Leipzig *et al.* (2021), estabelece os metadados na camada de *inputs* da pilha analítica atuando como a "cola" inseparável da reprodutibilidade.

Sales *et al.* (2023) determinam que o pilar da Reutilização (*Reusable*) dos princípios FAIR só é efetivamente atingido quando os recursos são descritos com informações claras de licenciamento e proveniência metodológica.

Para garantir essa governança e evitar o fenômeno advertido por Knapen *et al.* (2023), que evidenciaram que a tentativa de orquestrar matrizes heterogêneas em processos genéricos frequentemente resulta na perda irreversível de dados de contexto, a arquitetura foi concebida para não corromper as propriedades originais durante a triplificação.

Em vez de suprimir o histórico do dado, o *framework* foi desenhado para extrair e federar automaticamente os atributos administrativos das fontes originais. Esse rigor materializa-se na extração do histórico de processamento institucional, como a linhagem

ancorada na propriedade *hasLineage*, derivada da classe *LI_Lineage* do Padrão ISO 19115, e das permissões legais de acesso, como a política de dados ancorada em *wcmp2:dataPolicy*, extraída do Padrão WCMP.

A capacidade do modelo em reter e orquestrar essa teia de governança de forma simultânea às observações biológicas instanciou a formulação da terceira Questão de Competência (CQ03): “Qual é a linhagem (*lineage* da ISO 19115) e a política de dados (*dataPolicy* do WCMP) do recurso climático que serviu de entrada para a monitorização em tempo real da *Spodoptera frugiperda*?”.

A resolução desta questão demanda que o motor do *GraphDB* execute uma varredura de rastreabilidade transversal. A consulta SPARQL estruturada para esta validação navega a partir da instância atômica do organismo biológico monitorado, no padrão *Darwin Core*, e percorre os elos semânticos do grafo para resgatar os metadados macroscópicos de contexto associados a ele.

O Quadro 24 apresenta a estruturação desse código exploratório, cujos resultados atestarão empiricamente a capacidade do *framework* em garantir a transparência metodológica e a rastreabilidade integral da pesquisa agrícola:

Quadro 24 – Consulta SPARQL elaborada para a validação da CQ03

Query da Consulta SPARQL
PREFIX dwc: <http://rs.tdwg.org/dwc/terms/> PREFIX agrometa: <http://agrometa-framework.org/ontology/agrometa#> PREFIX wcmp2: <http://wmo-im.github.io/wcmp2/> SELECT DISTINCT ?organismo ?linhagemISO ?politicaDadosWCMP WHERE { # 1. Identifica a ocorrência biológica (Organismo monitorado) ?ocorrencia a dwc:Occurrence ; agrometa:hasScientificName ?organismo . # 2. Rastreia o metadado de Proveniência (ISO 19115) e extrai a Linhagem ?ocorrencia agrometa:hasClimaticContext ?contextoIso . ?contextoIso agrometa:hasLineage ?linhagemISO . # 3. Rastreia o catálogo climático (WCMP2) e extrai a Política de Dados ?ocorrencia agrometa:hasMeteorologicalRecord ?registroMeteorologico . ?registroMeteorologico wcmp2:dataPolicy ?politicaDadosWCMP . }

Fonte: O Autor.

A arquitetura da consulta SPARQL apresentada no Quadro 24 comprova a aderência do *framework* aos Princípios FAIR, mais especificamente ao pilar da Reutilização, conforme ressaltado por Sales *et al.* (2023).

Ao analisar a anatomia da consulta, observa-se que o motor de inferência utiliza a mesma instância central, referente à ocorrência atômica da praga, para irradiar a busca simultaneamente para dois domínios administrativos distintos.

Através da propriedade transversal *agrometa:hasClimaticContext*, o sistema navega até o Padrão ISO 19115 para resgatar a linhagem institucional *agrometa:hasLineage*, enquanto a propriedade *agrometa:hasMeteorologicalRecord* acessa o Padrão WCMP para recuperar as credenciais de acesso legal da OMM (*wcmp2:dataPolicy*).

Ao unificar a linhagem técnica e o respaldo legal de forma atrelada à observação biológica, o modelo atende nativamente à exigência da Pesquisa Computacional Reprodutível (PCR) descrita por Leipzig *et al.* (2021), garantindo que os metadados atuem como a "cola" da reprodutibilidade. Essa manobra confere rastreabilidade total à pesquisa agrometeorológica e atua ativamente contra a Entropia da Informação advertida por Habermann (2020), atestando que os dados não se perderão na obscuridade por falta de contexto.

Contudo, ao analisar a anatomia da consulta, observa-se que o motor de inferência utiliza a mesma instância central referente à ocorrência atômica da praga para irradiar a busca simultaneamente para dois domínios administrativos distintos.

Através da propriedade transversal *agrometa:hasClimaticContext*, o sistema navega até o Padrão ISO 19115 para resgatar a linhagem institucional *agrometa:hasLineage*, enquanto a propriedade *agrometa:hasMeteorologicalRecord* acessa o Padrão WCMP para recuperar as credenciais de acesso legal da OMM (*wcmp2:dataPolicy*). Ao unificar a linhagem técnica e o respaldo legal de forma atrelada à observação biológica, o modelo atende nativamente à exigência da Pesquisa Computacional Reprodutível (PCR) descrita por Leipzig *et al.* (2021), garantindo que os metadados atuem como a "cola" da reprodutibilidade.

Essa alternativa confere rastreabilidade total à pesquisa agrometeorológica e atua ativamente contra a Entropia da Informação advertida por Habermann (2020), atestando que os dados não se perderão na obscuridade por falta de contexto.

A execução da consulta SPARQL, descrita no Quadro 24, no Grafo de Conhecimento gerou o retorno consolidado apresentado no Quadro 25. O resultado demonstra a capacidade do sistema em varrer a teia semântica e resgatar os atributos administrativos atrelados às diferentes nomenclaturas do organismo monitorado.

Quadro 25 – Resultados inferidos pela execução da consulta SPARQL da CQ03

Organismo	Linhagem (ISO 19115)	Política de Dados (WCMP)
BOLD:AAA4532	Dado coletado via armadilhas de feromônio em lavouras experimentais e sistematizado no repositório institucional da Embrapa.	core
<i>Spodoptera frugiperda</i> J.E.Smith, 1797	Dado coletado via armadilhas de feromônio em lavouras experimentais e sistematizado no repositório institucional da Embrapa.	core

Fonte: O Autor.

Nota: Extraído pelo *GraphDB*.

Os resultados expostos no Quadro 25 atestam de forma categórica a eficácia da arquitetura em materializar a governança de dados no ecossistema agrometeorológico. Ao processar a requisição, o motor de inferência conectou com sucesso a taxonomia da praga, identificada tanto por seu código de barras genético BOLD quanto por sua nomenclatura científica tradicional, à exata linhagem metodológica de sua captura.

Essa proveniência, resgatada nativamente do nó *LI_Lineage* do Padrão ISO 19115, provê o contexto agrônomo indispensável para atestar a qualidade da amostra. Concomitantemente, o sistema extraiu a política de acesso da matriz climática subjacente, chancelada pelo nó *wmo:dataPolicy* do Padrão WCMP sob a categoria *core*.

Conforme fundamentado pela WMO (2025), a classificação nesta categoria atende à Política Unificada de Dados da OMM, garantindo o acesso totalmente livre, aberto e irrestrito ao recurso meteorológico. Ao prover e unificar essas condições de acesso claras exigidas por Sales *et al.* (2023), o suporte nativo a essa diretriz atesta que o compartilhamento e o consumo automatizado da informação pelo *framework* respeitam integralmente o pilar de Acessibilidade dos princípios FAIR.

Aprofundando a análise estrutural deste resultado à luz da classificação da *National Information Standards Organization* (NISO), abordada por Ulrich *et al.* (2022), constata-se que a CQ03 orquestra com precisão a extração dos chamados metadados administrativos.

Ao federar de forma autônoma o histórico de gestão do recurso do Padrão ISO 19115 e as informações de direitos autorais e acesso legal do Padrão WCMP, o *framework* transcende a mera recuperação de variáveis físicas brutas. Essa arquitetura impulsiona a informação ao longo do que Habermann (2020) define como o *Continuum* da Compreensão. Em vez de apenas evitar a perda de detalhes, a disponibilidade contínua desse contexto metodológico anexo ao dado gera a negentropia, ou seja, a entropia negativa, advertida pelo autor como um processo em que o valor informacional do registro agrometeorológico não se degrada, mas se expande ao permitir o reuso seguro por novas comunidades científicas.

A garantia dessa rastreabilidade automatizada materializa a premissa de Řezn k *et al.* (2022), estabelecendo a liga  o inquebr vel entre o acesso a dados abertos confi veis e a confiabilidade do resultado da decis o final em sistemas de alerta e suporte agr cola. A extra  o automatizada e simult nea dessas dimens es atesta a capacidade do modelo em materializar a governan a de dados e mitigar a forma  o das chamadas "ilhas de informa  o", severamente criticadas por Roussaki *et al.* (2023).

O ineditismo arquitetural deste resultado destaca-se frontalmente ao ser contrastado com as limita  es das infraestruturas contempor neas mapeadas no Estado da Arte. Diferentemente do ecossistema agr cola relatado por Dulaney *et al.* (2024), o qual foi for ado a isolar suas vari veis metodol gicas e de manejo em arquivos de texto n o estruturados no formato ".txt", ou da plataforma de Nath *et al.* (2024), que evidenciou o aprisionamento das pol ticas e proveni ncias em documentos est ticos como PDFs, o *framework* proposto n o fragmenta a informa  o em "silos de dados", fen meno advertido por Meyer *et al.* (2021).

Pelo contr rio, ao superar o gargalo tecnol gico da Reutiliza  o diagnosticado por Petrosyan *et al.* (2023), o modelo assegura que nenhum algoritmo ou pesquisador consuma as vari veis biol gicas e clim ticas na ponta do ecossistema sem herdar, de forma simult nea, obrigat ria e unificada, a linhagem de sua autoria institucional e a sua licen a legal de uso.

A valida  o da CQ03 consolida a blindagem administrativa do *framework*, garantindo a total transpar ncia sobre a origem e as permiss es da informa  o. Entretanto, a materializa  o dessa governan a documental equaciona apenas a dimens o passiva do ecossistema agrometeorol gico.

Para que a arquitetura atinja a maturidade exigida para o suporte direto   decis o,   imperativo que o sistema processe o dinamismo do ambiente f sico em alta resolu  o. Como advertem Zeginis *et al.* (2024), a integra  o cont nua desses dados   relevante para analisar a intera  o entre as vari veis clim ticas e a produ  o agr cola, viabilizando o monitoramento em tempo real e a cria  o de sistemas de alerta antecipado para proteger as planta  es contra danos causados por eventos clim ticos extremos. Guiada por essa premissa de an lise preditiva, a avalia  o do modelo avan a para a valida  o da CQ04.

4.3.5 Valida  o da CQ04: Correla  o Agrometeorol gica e Suporte   Decis o

A consolida  o da rastreabilidade institucional e do acesso legal alcan ada na CQ03 estabeleceu a base de governan a e de proveni ncia exigida para a ci ncia aberta. No entanto, para que o *framework* supere a fun  o de um reposit rio passivo e atue efetivamente como um

Sistema de Suporte à Decisão (DSS) no campo, é necessário que ele processe o dinamismo do ambiente físico em alta resolução.

Conforme apontado por Sigler *et al.* (2025), a fronteira da integração de dados reside na chamada "análise holística", que capacita os ecossistemas a combinarem modelos de diferentes áreas para simular cenários complexos, permitindo responder a perguntas preditivas analíticas como "o que vai acontecer" ou "e se acontecer".

Essa urgência é corroborada por Zeginis *et al.* (2024), os quais destacam que a verdadeira utilidade da interoperabilidade agrometeorológica consiste em analisar a interação entre as variáveis climáticas e a produção, viabilizando o monitoramento em tempo real e a criação de sistemas de alerta antecipado contra danos causados por eventos extremos.

Para materializar essa análise holística, o modelo precisa superar o isolamento entre a observação biológica e o sensoriamento físico. Como diagnosticado por Beretta *et al.* (2021), os padrões dependentes de domínio adotam estritamente a visão de seu produtor, fazendo com que a semântica permaneça encapsulada e inviabilizando que uma mesma matriz de dados seja consumida por campos científicos distintos.

A arquitetura proposta contorna esse obstáculo ao ancorar nativamente as medições climáticas do Instituto Nacional de Meteorologia (INMET) na ontologia universal SOSA (*Sensor, Observation, Sample, and Actuator*), propiciando uma granularidade fina na escala atômica do sensor. Paralelamente, o modelo interliga essas observações de precipitação aos registros biológicos da praga *Spodoptera frugiperda* mapeados no Padrão *Darwin Core*. O elo causal definitivo dessa topologia é orquestrado pela ontologia de suporte *Agrontology*, do vocabulário AGROVOC.

Conforme estabelecido pela FAO (2021) e detalhado em suas diretrizes editoriais (FAO, 2022), essa ontologia foi desenvolvida para acomodar associações semânticas transversais, provendo as propriedades relacionais não-hierárquicas necessárias para o domínio. É a aplicação dessas diretrizes, a exemplo do uso da propriedade de causalidade *agro:affects*, que permite ao sistema inferir formalmente qual cultura agrícola, como o Milho, sofrerá o impacto imediato dessa cadeia de eventos.

A capacidade do *framework* em processar esse entrelaçamento multidisciplinar e identificar situações de risco ambiental, como o estresse hídrico severo, instanciou a formulação da quarta e última Questão de Competência (CQ04): “Quais medições específicas de sensores (*MeasurementOrFact* do *Darwin Core*) indicam fatores climáticos extremos que afetam (propriedade *affects* da *Agrontology*) o desenvolvimento da cultura do Milho?”.

A resolução desta questão de complexidade analítica exige que o *GraphDB* execute uma varredura transversal simultânea nas frentes meteorológica (SOSA), biológica (*Darwin Core*) e agronômica (AGROVOC). O Quadro 26 apresenta a estruturação desse código SPARQL, desenhado para extrair os limiares críticos de precipitação subjacentes às ocorrências da praga e atestar o potencial prático do *framework* como um instrumento de alerta e correlação causal para a agricultura.

Quadro 26 – Consulta SPARQL elaborada para a validação da CQ04

Query da Consulta SPARQL
<pre> PREFIX dwc: <http://rs.tdwg.org/dwc/terms/> PREFIX agrometa: <http://agrometa-framework.org/ontology/agrometa#> PREFIX agro: <http://aims.fao.org/aos/agrontology#> PREFIX sosa: <http://www.w3.org/ns/sosa/> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> SELECT DISTINCT ?medicao ?valorExtremo ?culturaAfetada ?padraoSensor WHERE { # 1. O Ponto Central: A ocorrência biológica (Praga ou Evento Fisiológico) ?ocorrencia a dwc:Occurrence ; agrometa:hasAgrovocTerm ?uriPragaOuFator . # 2. A Ponte Semântica: Verifica na FAO qual cultura sofre o impacto ?uriPragaOuFator agro:affects ?culturaAfetada . { # ROTA A: Formato de Extensão do Darwin Core (MeasurementOrFact) # Ideal para amostragens manuais em campo ?ocorrencia dwc:measurementOrFact ?medicao . ?medicao a dwc:MeasurementOrFact ; dwc:measurementValue ?valorExtremo . BIND("Darwin Core (MeasurementOrFact)" AS ?padraoSensor) } UNION { # ROTA B: Formato de IoT e Sensores (SOSA Observation) # O padrão que o framework implementou nativamente para o INMET ?ocorrencia agrometa:hasEventDate ?dataEvento . ?medicao a sosa:Observation ; agrometa:hasEventDate ?dataEvento ; agrometa:hasPrecipitationValue ?valorExtremo . BIND("SOSA (Sensor/Observation)" AS ?padraoSensor) } # 3. Filtro de Fator Climático Extremo (Estresse Hídrico Severo - Chuva < 1mm) FILTER (xsd:float(?valorExtremo) < 1.0) } ORDER BY ASC(xsd:float(?valorExtremo)) </pre>

Fonte: O Autor.

A consulta SPARQL apresentada no Quadro 26 atesta a capacidade do *framework* em superar a heterogeneidade estrutural para correlacionar matrizes granulares de diferentes

ecossistemas, solucionando o desafio técnico de integração em ambientes de *Big Data* alertado por Putrama e Martinek (2024).

Ao dissecar o código sob a lógica de correspondência de subgrafos elucidada por El Idrissi *et al.* (2022), observa-se que o motor de inferência utiliza a ocorrência biológica como âncora inicial, projetando imediatamente a busca para a ontologia *Agrontology* por meio da propriedade transversal *agro:affects*, a qual forma a ponte causal que identifica a cultura vulnerável à praga. Porém, a originalidade materializa-se no uso da cláusula UNION para orquestrar uma verdadeira federação de dados em tempo de execução.

Na "Rota A", o algoritmo busca por observações registradas sob a extensão *MeasurementOrFact* do Padrão *Darwin Core*, classe oficializada pelo *Darwin Core Maintenance Group* (2021) como a ponte de interoperabilidade central para suportar a associação e o armazenamento das medições ou fatos atrelados à ocorrência biológica manual.

Já na "Rota B", o sistema navega pela ontologia SOSA, resgatando de forma direta as medições físicas geradas em alta frequência pelas redes de sensores IoT do Instituto Nacional de Meteorologia (INMET).

A orquestração simultânea desta "Rota B" materializa a arquitetura defendida por San Emeterio de la Parte *et al.* (2023) para o processamento de grandes volumes de telemetria no campo, e consagra a diretriz de Beretta *et al.* (2021), rompendo o encapsulamento estrito dos produtores ao usar a ontologia SOSA para separar a observação atômica do sensor físico da característica temática de interesse.

A harmonização simultânea desses dois padrões através da federação semântica resolve de forma definitiva o engessamento arquitetural advertido por Beretta *et al.* (2021), os quais criticam os sistemas atuais por alienarem a interpretação da observação física nos jargões estritos de seus produtores originais.

Sob essa malha integrada, a aplicação do filtro matemático para limiares climáticos extremos, como o estresse hídrico severo ($\text{FILTER} < 1.0$), transcende a extração de um número frio e consolida a "análise holística" citada por Sigler *et al.* (2025).

A consulta estabelece uma teia lógica inquebrável e dinâmica de causalidade, em que o clima orquestra o risco de alerta, a biologia registra a praga, e a semântica infere a cultura ameaçada. O sucesso desta operação consolida o *framework* não apenas como um integrador de metadados, mas como um poderoso Sistema de Suporte à Decisão (DSS), assegurando que variáveis heterogêneas interajam de forma autônoma para transformar a abundância de dados brutos no conhecimento produtivo e preditivo exigido por Roussaki *et al.* (2023).

A execução da consulta SPARQL descrita no Quadro 26 materializa a correlação holística exigida pelo modelo. Conforme solucionado e atestado na CQ02 através do Grafo de Especialista Local e visto no Quadro 23, o motor de inferência pôde navegar pela propriedade *agro:affects* previamente injetada na base, contornando a omissão estrutural da FAO.

Essa manobra valida na prática a premissa destacada por Alharbi *et al.* (2024) de que as ontologias globais contemporâneas frequentemente exigem instâncias adicionais para alimentar inferências lógicas em domínios locais. Com o Grafo de Conhecimento federado e enriquecido por essa ponte de domínio, a consulta superou a limitação do AGROVOC e foi executada com sucesso, gerando o retorno consolidado exposto no Quadro 27.

Quadro 27 – Resultados da correlação entre clima extremo e impacto cultural (CQ04)

Medição (SOSA/DwC)	Valor Extremo (mm)	Cultura Afetada (AGROVOC)	Padrão do Sensor
.../observation_2025-05-06	0.0	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-08-18	0.0	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-03-01	0.0	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-03-29	0.0	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-06-29	0.0	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-01-21	0.2	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-04-10	0.2	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-10-04	0.2	c_12332 (Milho)	SOSA (Sensor/Observation)
.../observation_2025-04-29	0.6	c_12332 (Milho)	SOSA (Sensor/Observation)

Fonte: O Autor.
Nota: Extraído pelo *GraphDB*.

Os 9 registros críticos recuperados no Quadro 27 atestam empiricamente a eficácia do *framework* como um genuíno Sistema de Suporte à Decisão (DSS) para o domínio agrônômico. Do ponto de vista técnico, a extração comprova que o motor de inferência navegou perfeitamente pelas instâncias atômicas da ontologia de sensores (SOSA), isolando na escala temporal exata as datas em que a proliferação da praga no campo coincidiu com um regime climático severo de estresse hídrico, evidenciado pelos escassos volumes de precipitação, entre 0.0 e 0.6 mm.

Ao processar simultaneamente as rotas do Darwin Core e do Padrão WMCP, o modelo consolida-se como um filtro analítico de precisão, eliminando o "ruído" estatístico de matrizes

massivas. Esse isolamento atômico contrapõe-se diretamente às limitações de *frameworks* contemporâneos, como o BDAKG de Nath *et al.* (2024), cuja modelagem em cubos de dados sacrificou a resolução fina para o monitoramento contínuo em favor de agregações estatísticas.

Além disso, ao orquestrar a telemetria nativamente, a arquitetura proposta elimina a dependência crônica de módulos tradutores intermediários exigidos por ecossistemas recentes de IoT, a exemplo do STSDaMaS proposto por San Emeterio de la Parte *et al.* (2023).

Mais do que recuperar valores numéricos de precipitação, esta validação consegue propiciar a agregação causal alcançada pelo sistema diretamente no motor lógico. Diferentemente de infraestruturas ambientais como a *Link-Climate* proposta por Wu *et al.* (2022), que falharam metodologicamente ao ter que "achatar" os grafos de volta para matrizes tabulares bidimensionais para viabilizar o processamento preditivo, destruindo, assim, a topologia relacional original, o *framework* preserva a integridade lógica e semântica da informação.

Conforme advertem Roussaki *et al.* (2023), a verdadeira comunicação entre plataformas deve superar o simples acúmulo de dados brutos e transformar a abundância de informações em conhecimento produtivo e acionável. Ao cruzar com êxito a observação meteorológica e a ocorrência do inseto para inferir o impacto prático na cultura afetada (c_12332), a resolução da CQ04 consagra a "análise holística" almejada por Sigler *et al.* (2025).

O ecossistema prova que as variáveis biológicas e climáticas deixaram de operar como silos isolados, convertendo-se em uma malha de inteligência preditiva capaz de suportar sistemas de alerta precoce em cenários de risco extremo.

Aprofundando a análise deste resultado, constata-se que nem as matrizes meteorológicas do INMET, tampouco os registros de ocorrência biológica do GBIF, possuíam a capacidade nativa de vincular a ameaça climática e entomológica ao impacto direto na cultura agrícola.

Essa ausência de contexto agrônômico nas fontes primárias reflete com exatidão o diagnóstico de Soares, Saraiva e Drucker (2020), os quais alertam que o padrão *Darwin Core*, quando operado de forma isolada, carece das classes semânticas para representar a complexidade das interações e dos impactos na agrobiodiversidade.

Para superar essa severa limitação estrutural, a arquitetura do *framework* orquestra de forma exata o que Vogt *et al.* (2025) definem como as camadas terminológica e proposicional da interoperabilidade. Em vez de recorrer a deduções arbitrárias, o motor de inferência utiliza a regra causal explícita (*agro:affects*) previamente declarada no Grafo de Especialista Local para conectar os domínios.

Em tempo de execução, o algoritmo alinha a taxonomia da praga aos limiares de estresse hídrico extraídos da ontologia SOSA, aplicando o conhecimento do especialista sobre as matrizes brutas. A verdadeira inovação deste processamento reside na capacidade computacional de orquestrar essas regras formais sobre bases isoladas e massivas sem a necessidade de reestruturá-las fisicamente.

O resultado consolida o ápice da integração semântica, em que a máquina torna-se capaz de alertar o impacto exato na cultura do milho, com precisão de milímetros e exatidão temporal, transformando dados que nasceram fragmentados em uma malha de inteligência preditiva ambiental e agrônômica.

Em síntese, a validação da CQ04 coroa o desenvolvimento do framework, comprovando sua maturidade em transcender a função de um mero repositório de metadados para atuar como um orquestrador semântico na agrometeorologia.

Ao longo deste capítulo, demonstrou-se de forma empírica que a arquitetura resolveu o conflito de granularidade (CQ01), desambiguou terminologias baseadas em texto livre (CQ02), garantiu a rastreabilidade institucional sob a ótica dos princípios FAIR (CQ03) e interligou o monitoramento físico de sensores IoT a relações lógicas de causalidade biológica (CQ04).

Ao harmonizar as estruturas espaciais rígidas da ISO 19115, a granularidade biológica do Darwin Core e as matrizes climáticas do WCMP por meio do vocabulário AGROVOC, o modelo soluciona as barreiras arquiteturais e a fragmentação terminológica severamente mapeadas no estado da arte.

Assentadas as bases tecnológicas e validadas todas as questões de competência exigidas para a interoperabilidade interdomínios, a pesquisa avança para o seu encerramento no Capítulo 5, dedicado às considerações finais, à síntese das contribuições científicas e às diretrizes para trabalhos futuros.

5 CONSIDERAÇÕES FINAIS

A transição da agricultura tradicional para os ecossistemas de agricultura de precisão e tomada de decisão baseada em dados esbarra, inevitavelmente, na fragmentação da informação. Conforme evidenciado ao longo deste trabalho, a proliferação de padrões de metadados independentes culminou na criação de "ilhas de informação" tecnológicas, citadas por Roussaki *et al.* (2023), nas quais matrizes climáticas, registros biológicos e históricos de proveniência permanecem semanticamente isolados.

Diante desta barreira tecnológica e da necessidade de viabilizar a verdadeira "análise holística" preditiva demandada por Sigler *et al.* (2025), o presente trabalho propôs-se a solucionar a dicotomia entre a multidisciplinaridade exigida pelo campo e o isolamento dos formatos originais.

Nesse sentido, atesta-se que o Objetivo Geral desta pesquisa foi plenamente alcançado com o desenvolvimento e a implementação bem-sucedida de um *framework* semântico AgroMeta, capaz de integrar metadados agrometeorológicos e promover a interoperabilidade entre os padrões.

A concepção arquitetural da solução transcendeu o simples alinhamento sintático, materializando-se em um *middleware* computacional que automatiza o cruzamento de esquemas e a orquestração do conhecimento sem corromper a proveniência dos dados originais.

Ao transformar arquivos heterogêneos dispersos em um Grafo de Conhecimento unificado, o modelo provou ser capaz de sustentar sistemas de alerta sem depender de intervenções manuais exaustivas.

O percurso metodológico que culminou nesta arquitetura exigiu, inicialmente, o alcance do primeiro objetivo específico, que buscar realizar uma análise profunda das estruturas dos padrões WCMP, ISO 19115 e *Darwin Core*.

Por meio da revisão sistemática da literatura, este trabalho mapeou com ineditismo as convergências, como a necessidade primária de localização no tempo e no espaço, e, sobretudo, as divergências críticas que inviabilizavam a integração interdomínios.

Diagnosticou-se que o cerne da incompatibilidade não residia apenas nas diferenças de linguagem de codificação, mas em uma severa lacuna de granularidade. Constatou-se que padrões consolidados nas Ciências da Terra, como o WCMP e a ISO 19115, operam sob uma perspectiva macroscópica, atuando como "envelopes" para o conjunto de dados completo (*dataset*).

Essa característica reflete exatamente a limitação apontada por Annane *et al.* (2023), que advertem que descrições meteorológicas limitadas ao alto nível são insuficientes para garantir a reutilização plena por outras comunidades científicas.

Em contrapartida, a documentação da agrobiodiversidade, regida pelo Padrão *Darwin Core*, concentra-se na resolução micro, focada na ocorrência atômica e na medição individual, conforme as diretrizes estruturais de Wieczorek *et al.* (2012).

A constatação teórica desta assimetria de escopos foi o alicerce que direcionou o design do *framework*, comprovando que o cruzamento direto dessas variáveis seria impossível sem uma intervenção ontológica capaz de orquestrar diferentes níveis de abstração simultaneamente.

Para transpor a assimetria estrutural identificada no estado da arte, o segundo objetivo específico focou em definir o mapeamento entre os padrões originais WCMP, ISO 19115 e *Darwin Core*, e as ontologias de referência selecionadas.

Esse objetivo foi atingido por meio do desenvolvimento de uma Ontologia de Alinhamento baseada nas diretrizes da Metodologia *NeOn*, a qual preconiza o reuso de *core ontologies* para evitar a redundância e o "inchaço do modelo".

Em vez de propor uma arquitetura centralizadora e rígida, uma abordagem considerada impraticável em larga escala por Sadeghi *et al.* (2024), o *framework* delegou a representação do tempo à ontologia *OWL-Time*, do espaço à *GeoSPARQL*, e a granularidade da medição climática à ontologia *SOSA*.

O sucesso deste mapeamento ontológico materializou a interoperabilidade proposicional preconizada por Vogt *et al.* (2025) e revelou um avanço contundente frente às pesquisas atuais.

Enquanto trabalhos de ponta, como os de Parisse *et al.* (2022) e Wu *et al.* (2022), esbarraram na complexidade estrutural e precisaram retroceder metodologicamente, achatando matrizes multidimensionais e grafos em planilhas tabulares CSV para viabilizar a análise, a arquitetura aqui proposta preservou a riqueza topológica e relacional da informação de ponta a ponta.

A separação inteligente entre o catálogo macroscópico e a observação atômica consolidou um modelo em que a integração ocorre na camada semântica, sem a necessidade de destruição do formato físico original.

Porém, a harmonização proposicional não é suficiente se a comunicação permanecer obstruída por jargões incompatíveis e ambiguidades textuais. Desse modo, o terceiro objetivo

específico foi plenamente alcançado ao se estabelecerem regras de desambiguação terminológica utilizando o vocabulário multilíngue AGROVOC.

A validação do *framework* provou que a adoção deste vocabulário como ponte semântica garantiu a tão almejada interoperabilidade terminológica evidenciada por Vogt *et al.* (2025), transformando nomes informais de pragas e culturas em Identificadores Globais Persistentes (URIs) vinculados à *Linked Open Data*.

A eficácia dessa estratégia de desambiguação foi evidenciada ao se confrontar os resultados do *framework* com as falhas operacionais descritas no estado da arte. Mutlu *et al.* (2025) demonstraram que a permissão de texto livre em metadados gera um colapso descritivo que exige rotinas massivas de curadoria corretiva.

Para tentar mitigar esse caos, Roić e Pivac (2024) precisaram mutilar até 38% dos domínios da ISO 19115 para forçar o uso de listas restritas, enquanto Dulaney *et al.* (2024) optaram por isolar variáveis agronômicas em arquivos de texto ".txt" não estruturados.

A solução desenvolvida nesta pesquisa contrapõe-se diretamente a essas práticas ao utilizar a ontologia *Agrontology*, e propriedades causais como *agro:affects*, o modelo provou que não é necessário engessar o padrão de geoinformação.

O sistema demonstrou capacidade preditiva de inferir a cultura agrícola oculta em metadados administrativos vagos, conectando a praga, o clima e a lavoura sem gerar novos "silos de dados", conforme o alerta frequentemente reiterado por Roussaki *et al.* (2023) e Meyer *et al.* (2021).

Consolidadas as bases ontológicas e as regras de desambiguação, o quarto objetivo específico foi atingido mediante a implementação computacional do modelo por meio de um motor de triplicação semântico, seguindo o fluxo entrada, transformação e saída.

Diferentemente de abordagens que dependem de conversões manuais e curadoria exaustiva de dados, a arquitetura foi operacionalizada na forma de uma *API RESTful* em linguagem Python, garantindo o processamento eficiente em memória e o alinhamento a infraestruturas de nuvem.

Essa infraestrutura permitiu ingerir ativamente arquivos sintáticos divergentes, como matrizes tabulares em CSV, catálogos geoespaciais em GeoJSON e árvores de proveniência em XML, extraindo e convertendo essas fontes nativamente para o modelo estrutural de grafos em RDF.

Ao serializar essas informações heterogêneas em instâncias logicamente interligadas (*ABox*), o *framework* não apenas validou o esquema teórico, mas entregou um artefato de

software dinâmico capaz de superar o gargalo de processamento sintático inerente ao volume massivo de dados agrometeorológicos.

O percurso metodológico culminou no alcance do quinto objetivo específico, focado em validar a arquitetura proposta por meio de Questões de Competência (CQs) para verificar a efetividade da integração semântica.

Esse objetivo foi plenamente consolidado ao se submeter o modelo a um cenário de teste real e de criticidade agronômica, envolvendo a interação dinâmica entre a praga *Spodoptera frugiperda*, a cultura do milho e limiares climáticos extremos de precipitação.

A execução bem-sucedida de consultas na linguagem SPARQL no banco de dados orientado a grafos *GraphDB* provou, de forma empírica, que o *framework* construído é capaz de navegar entre o domínio institucional da ISO 19115, as matrizes climáticas do WCMP e as observações biológicas granulares do *Darwin Core* sem corromper a topologia da informação.

A materialização dessa validação atingiu seu ápice na CQ04, a qual atestou a capacidade do sistema em atuar como um genuíno Sistema de Suporte à Decisão (DSS) agrometeorológico. Ao cruzar nativamente os dados atômicos de chuva, via ontologia SOSA, com as propriedades causais de impacto agrícola *agro:affects*, a arquitetura transcendeu a mera agregação passiva de informações e alcançou o patamar da "análise holística" ressaltada por Sigler *et al.* (2025).

Enquanto arquiteturas recentes voltadas para a agricultura de precisão, como a de Zeginis *et al.* (2024), esbarraram em engessamentos técnicos causados por pesadas restrições sintáticas, e modelos como o de Alharbi *et al.* (2024) demonstraram limitações de escalabilidade devido à alimentação estritamente manual, o *framework* aqui desenvolvido orquestrou o conhecimento de forma fluida.

A capacidade de inferir qual cultura seria atingida sob um estresse hídrico severo corrobora a eficácia da ferramenta em suportar sistemas de alerta antecipado, uma demanda apontada por Zeginis *et al.* (2024) como vital para proteger plantações contra eventos extremos.

Além disso, a etapa de validação prática revelou um subproduto científico de elevado impacto metodológico frente à manutenção de tesouros internacionais. Ao constatar, durante a execução de consultas SPARQL, a ausência da relação causal explícita entre a praga *Spodoptera frugiperda* e a cultura do milho diretamente no *endpoint* oficial da FAO, a pesquisa evidenciou a fragilidade inerente aos repositórios globais.

Conforme diagnosticado por Putrama e Martinek (2024), a construção e a manutenção de ontologias em larga escala demandam um altíssimo esforço de curadoria manual especializada, o que torna essas bases suscetíveis a omissões críticas. O fato de o AGROVOC não possuir a tripla relacional mapeada, o que Oliveira *et al.* (2023) afirmam ser a principal

praga-chave do milho na América do Sul, prova que a dependência passiva de bases centralizadas pode paralisar a "análise holística" almejada por Sigler *et al.* (2025).

Diante desse achado, a pesquisa materializou na prática a advertência estabelecida por Prudhomme *et al.* (2025), os quais constataram que a orquestração de ontologias em ambientes computacionais frequentemente expõe inconsistências e erros estruturais ocultos nas bases de dados originais.

Sob a ótica do *Linked Open Data* (LOD), conceituada por Isotani e Bittencourt (2015), essa descoberta eleva a arquitetura proposta de uma mera consumidora de dados para uma ferramenta autônoma e proativa de auditoria ontológica.

A solução adotada, baseada na injeção formal da propriedade *agro:affects* por meio de um Grafo de Especialista Local para suprir a omissão da FAO, provou que o modelo não apenas resolve as demandas internas de integração, mas estabelece um mecanismo metodológico de *feedback* estruturado, possuindo o potencial de gerar relatórios automatizados para que consórcios internacionais realizem a curadoria corretiva de suas bases globais.

Desse modo, o *framework* assegura que o conhecimento agrometeorológico evolua de forma colaborativa, rompendo definitivamente com os "silos de ontologias", citados por Prudhomme *et al.* (2025), e fortalecendo a rede global de dados conectados.

A transposição tecnológica empreendida nesta pesquisa atesta que a gestão de metadados agrometeorológicos não deve se restringir ao encapsulamento sintático e ao armazenamento físico em silos.

Como defende Habermann (2020), metadados ricos e semanticamente alinhados atuam como o "antídoto crítico" contra a Entropia da Informação. O *framework* desenvolvido permitiu que os dados avançassem no *Continuum* da Compreensão, transformando o acúmulo de arquivos heterogêneos na inteligência preditiva e acionável tão requisitada por Roussaki *et al.* (2023) para a agricultura moderna.

Apesar do *framework* AgroMeta ter atingido seus objetivos centrais de promover a interoperabilidade estrutural e semântica, a sua validação prática e implementação computacional evidenciaram limitações inerentes à arquitetura de dados e à dependência de serviços externos, as quais devem ser criticamente ponderadas.

A primeira limitação refere-se ao que pode ser considerado uma ilusão de automação plena frente à incompletude de bases globais. O modelo foi projetado para atuar como um orquestrador automatizado, mas a sua eficácia depende intimamente da completude da ontologia AGROVOC.

Durante a validação (CQ02 e CQ04), constatou-se uma lacuna estrutural no repositório da FAO relacionada a ausência da relação direta de afetação entre a praga *Spodoptera frugiperda* e a cultura do milho. Para garantir a inferência esperada, exigiu-se a injeção manual dessa tripla relacional por meio de um "Grafo de Especialista Local". Este cenário evidencia que, no primeiro contato com a incompletude do mundo real, a automação do *framework* cedeu espaço à necessidade do mesmo esforço de curadoria manual especializada que o presente trabalho criticou em abordagens de terceiros.

A segunda limitação está atrelada à vulnerabilidade da arquitetura de rede e à fragilidade do mecanismo de tolerância a falhas (*fallback*). O motor de triplificação depende criticamente de consultas dinâmicas e em tempo real ao *endpoint* SPARQL da FAO para realizar a desambiguação dos termos, o que o sujeita a riscos constantes de indisponibilidade, como bloqueios de firewall e erros "*400 Bad Request*".

Para evitar a interrupção abrupta do *pipeline*, o sistema implementou uma estratégia de *fallback* que converte a *string* de texto livre em uma URI local (ex: http://agrometa-framework.org/taxonomy/Spodoptera_frugiperda). Contudo, assume-se que esta estratégia sacrifica a interoperabilidade global, visto que, ao camuflar um texto livre sob a forma de uma URI arbitrária e local, o sistema deixa de desambiguar o conceito e acaba por gerar novas "ilhas de informação".

Por fim, a terceira limitação reside na validação da robustez do pipeline de ETL Semântico frente à degradação de dados reais. Durante a preparação do cenário de testes para a extração de proveniência institucional (CQ03) no padrão ISO 19115, identificou-se que o arquivo XML da Embrapa continha omissões estruturais e *tags* críticas vazias (marcadas como *gco:nilReason="missing"*). Ao invés de o sistema demonstrar como lidaria analiticamente com metadados ausentes ou corrompidos, a validação ocorreu sobre uma cópia do arquivo populada e "consertada" manualmente em ambiente controlado.

Reconhece-se que essa validação sob dados simulados mascara a real (in)capacidade do modelo em tratar a "sujeira" e as omissões severas características do *Big Data* agrícola, enfraquecendo a prova de sua total resiliência operacional.

Como trabalhos futuros, sugere-se a ampliação do mapeamento ontológico desenvolvido, inserindo novas classes de domínio para responder a futuras Questões de Competência. Recomenda-se a expansão metodológica do Grafo de Especialista Local para abranger novas culturas, pragas e fenômenos agrometeorológicos transversais, além da integração com novos padrões de sensores e maquinário agrícola não contemplados no escopo atual.

Recomenda-se, também, a integração da arquitetura proposta com algoritmos de Processamento de Linguagem Natural (PLN) e *Machine Learning* para automatizar a extração de conceitos agronômicos em literaturas não estruturadas, mitigando o esforço de curadoria manual.

Finalmente, a aplicação de modelos de predição climática baseados em Inteligência Artificial diretamente sobre a rede de grafos gerada poderá elevar o *framework* a um ecossistema autônomo de recomendação de manejo agrícola, otimizando ainda mais a resiliência do setor primário.

REFERÊNCIAS

- ABHILASH, C. B.; MAHESH, K. Ontology-based data interestingness: A state-of-the-art review. **Natural Language Processing Journal**, [S.l.], v. 4, p. 100021, 2023.
- ALDANA-MARTÍN, J. F. *et al.* Semantic modelling of Earth Observation remote sensing. **Expert Systems With Applications**, [S.l.], v. 187, p. 115838, 2022.
- ALHARBI, A. F. *et al.* An ontology-based agriculture decision-support system with an evidence-based explanation model. **Smart Agricultural Technology**, [S.l.], v. 9, p. 100659, 2024.
- ALMEIDA, A. P. *et al.* **Carta de Recomendação para o Uso da Inteligência Artificial na Educação: Desafios e Potencialidades**. São Paulo: Editora Nelpa, 2025. Disponível em: https://www.educamaisai.com.br/_files/ugd/019da9_5d7c0b401bbd41dcbb6a7220058af875.pdf. Acesso em: 26 mai. 2026.
- ANNANE, A. *et al.* Improving FAIRness of the SYNOP meteorological data set with semantic metadata. **International Journal of Metadata, Semantics and Ontologies**, [S.l.], v. 16, n. 2, p. 118-137, 2023.
- ARSENAULT, R. *et al.* A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. **Scientific Data**, [S.l.], 2020.
- AVELINO, J. de O.; CORDEIRO, K. de F.; CAVALCANTI, M. C. An RDF based approach for integrating data at different levels of abstraction. *In: Brazilian Symposium On Multimedia And The Web (WebMedia '20)*, 2020, Nova York. **Anais [...]**. Nova York: ACM, 2020.
- BATISTA, D. *et al.* Machine actionable metadata models. **Scientific Data**, [S.l.], 2022.
- BERETTA, V. *et al.* A user-centric metadata model to foster sharing and reuse of multidisciplinary datasets in environmental and life sciences. **Computers & Geosciences**, [S.l.], v. 154, p. 104807, 2021.
- BRIZUELA, J. *et al.* A roadmap for a middleware as a federation service for integrative data retrieval of agricultural data. **Journal of Integrative Bioinformatics**, [S.l.], v. 21, n. 3, p. 20240027, 2024.
- DARWIN CORE MAINTENANCE GROUP. **List of Darwin Core terms**. [S.l.]: Biodiversity Information Standards (TDWG), 2021. Disponível em: <http://rs.tdwg.org/dwc/doc/list/>. Acesso em: 25 mai. 2026.
- DEVARAJU, A.; HUBER, R. An automated solution for measuring the progress toward FAIR research data. **Patterns**, [S.l.], v. 2, n. 11, p. 100370, 2021.
- DRURY, B. *et al.* A survey of semantic web technology for agriculture. **Information Processing in Agriculture**, [S.l.], v. 6, n. 4, p. 487-501, 2019.

DULANEY, W. P. et al. Development of a gridded yield data archive for farm management and research at the USDA Beltsville Agricultural Research Center. **Agrosystems, Geosciences & Environment**, [S.l.], v. 7, p. e20474, 2024. DOI: 10.1002/agg2.20474.

EL IDRISSE, B. *et al.* RDF/OWL storage and management in relational database management systems: A comparative study. **Journal of King Saud University – Computer and Information Sciences**, [S.l.], v. 34, p. 7604-7620, 2022.

FENSEL, D. *et al.* (ed.). **Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential**. Cambridge: The MIT Press, 2003.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). **Agricultural Metadata Element Set (AgMES): Namespace Specification Version 1.1**. Rome: FAO, 2010. Disponível em: <https://aims.fao.org/standards/agmes/namespace-specification>. Acesso em: 25 ago. 2025.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). **AGROVOC: Semantic data interoperability on food and agriculture**. Rome: FAO, 2021. Disponível em: <https://doi.org/10.4060/cb2838en>. Acesso em: 10 set. 2025.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). **The AGROVOC Editorial Guidelines**. 2. ed. Rome: FAO, 2022. Disponível em: <https://doi.org/10.4060/cb8640en>. Acesso em: 10 set. 2025.

GILLILAND, A. J. Setting the stage. In: BACA, M. (ed.). **Introduction to Metadata**. 2. ed. Los Angeles: Getty Research Institute, 2008. p. 1-19.

GOLDSTEIN, A.; FINK, L.; RAVID, G. A Cloud-Based Framework for Agricultural Data Integration: A Top-Down-Bottom-Up Approach. **IEEE Access**, v. 10, p. 88527-88537, 2022.

HABERMANN, T. Metadata and Reuse: Antidotes to Information Entropy. **Patterns**, [S.l.], v. 1, p. 1-5, abr. 2020. Disponível em: <https://doi.org/10.1016/j.patter.2020.100004>. Acesso em: 07 ago. 2025.

HAN, S.; HAN, J. Case study on an integrated interoperable metadata model for geoscience information resources. **Geoscience Data Journal**, v. 9, p. 355-370, 2022.

HAYNES, D. **Metadata for information management and retrieval**. Cambridge: Facet Publishing, 2004.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO 19115:2014 Geographic information — Metadata**. Genebra: ISO, 2014.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec Editora, 2015.

KHAN, A. A.; ABONYI, J. Information sharing in supply chains – Interoperability in an era of circular economy. **Cleaner Logistics and Supply Chain**, v. 5, p. 100074, 2022.

KLEPPMANN, M.; RICCOMINI, C. **Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems**. 2. ed. Sebastopol, CA: O'Reilly Media, 2026.

KNAPEN, R.; LOKERS, R.; JANSSEN, S. Evaluating the D4Science virtual research environment platform for agro-climatic research. **Agricultural Systems**, [S.l.], v. 210, p. 103706, 2023.

KRISNAWIJAYA, N. N. K. *et al.* Implementing FAIR principles in data management systems: A multi-case study in precision farming. **Computers and Electronics in Agriculture**, [S.l.], v. 230, p. 109855, 2025.

LEIPZIG, J. *et al.* The role of metadata in reproducible computational research. **Patterns**, [S.l.], v. 2, p. 1-28, set. 2021. Disponível em: <https://doi.org/10.1016/j.patter.2021.100322>. Acesso em: 07 ago. 2025.

MA, X.; MA, C.; WANG, C. A new structure for representing and tracking version information in a deep time knowledge graph. **Computers & Geosciences**, [S.l.], v. 145, p. 104620, 2020.

MANI, S.; ANNADURAI, S. An Improved Structural-Based Ontology Matching Approach Using Similarity Spreading. **International Journal on Semantic Web and Information Systems**, [S.l.], v. 18, n. 1, 2022. Disponível em: <https://doi.org/10.4018/IJSWIS.300825>. Acesso em: 07 ago. 2025.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2017.

MARTÍNEZ, D. *et al.* A conceptual data modeling framework with four levels of abstraction for environmental information. **Environmental Modelling and Software**, [S.l.], v. 183, p. 106248, 2025. Disponível em: <https://doi.org/10.1016/j.envsoft.2024.106248>. Acesso em: 07 ago. 2026.

MCGUINNESS, D. L. Ontologies Come of Age. In: FENSEL, D. *et al.* (ed.). **Spinning the Semantic Web: bringing the World Wide Web to Its Full Potential**. Cambridge: The MIT Press, 2003. p. 1-19.

MECKLER, S. *et al.* The Web and Linked Data as a Solid Foundation for Dataspaces. In: ACM WEB CONFERENCE 2023 (WWW '23 Companion), 2023, Nova York. **Companion Proceedings** [...]. Nova York: ACM, 2023.

MEYER, R. *et al.* Aligning Standards Communities: Sustainable Darwin Core MIxS Interoperability. **Biodiversity Information Science and Standards**, [S.l.], v. 5, e73775, 2021.

MICHEL, F. *et al.* Semantic Indexing of Open Scientific Literature to Help Users Discover and Navigate through Publications Networks. **Biodiversity Information Science and Standards**, [S.l.], v. 6, e93640, 2022.

MOTTA, M. *et al.* A framework for FAIR robotic datasets. **Scientific Data**, v. 10, n. 320, 2023.

MUTLU, I. *et al.* Automated curation of spatial metadata in environmental monitoring data. **Ecological Informatics**, v. 86, p. 103038, 2025.

NATH, R. P. D. *et al.* Knowledge Graph Generation and Enabling Multidimensional Analytics on Bangladesh Agricultural Data. **IEEE Access**, [S.l.], v. 12, p. 85028-85043, 2024.

NEVEU, P.; DAVID, R.; JONQUET, C. Improving data identification and tagging for more effective decision making in agriculture. In: ARMSTRONG, L. (ed.). **Improving Data Management and Decision Support Systems in Agriculture**. [S.l.]: Burleigh Dodds Science Publishing, 2020. p. 85.

O'GRADY, P. *et al.* The new WMO Core Metadata Profile (WCMP): A profile of OGC API - Records for the WMO Information System (WIS). In: AGU FALL MEETING 2021, 2021, New Orleans, LA. **Anais [...]**. New Orleans, LA, 2021.

OECD; FAO. **OECD-FAO Agricultural Outlook 2023-2032**. Paris: OECD Publishing, 2023. Disponível em: <https://doi.org/10.1787/08801ab7-en>. Acesso em: 20 abr. 2026.

OLIVEIRA, I. R. de *et al.* **Manejo de lagarta-do-cartucho em sistemas de produção integrados com braquiária**. Sete Lagoas, MG: Embrapa Milho e Sorgo, 2023. 27 p. (Comunicado Técnico, 260). Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1159228/1/Manejo-de-lagarta-do-cartucho-em-sistemas-de-producao-integrados.pdf>. Acesso em: 15 jun. 2026.

PAGE, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, [S.l.], v. 372, n. 71, p. 1-9, 2021.

PARISSE, B. *et al.* MADIA - Meteorological variables for agriculture: A dataset for the Italian area. **Data in Brief**, [S.l.], v. 46, p. 108852, 2022.

PETROSYAN, L. *et al.* FAIR degree assessment in agriculture datasets using the F-UJI tool. **Ecological Informatics**, [S.l.], v. 76, p. 102126, 2023.

PORTER, J. H. *et al.* Using a units ontology to annotate pre-existing metadata. **Scientific Data**, [S.l.], 2025.

PRUDHOMME, T. *et al.* **A semantic approach to mapping the Provenance Ontology to Basic Formal Ontology**. [S.l.: s.n.], 2025.

PUTRAMA, I M.; MARTINEK, P. Heterogeneous data integration: Challenges and opportunities. **Data in Brief**, [S.l.], p. 110853, 2024.

RANATUNGA, S. *et al.* Use of Semantic Web Technologies to Enhance the Integration and Interoperability of Environmental Geospatial Data: A Framework Based on Ontology-Based Data Access. **ISPRS International Journal of Geo-Information**, [S.l.], v. 14, n. 2, p. 52, 2025.

ŘEZNÍK, T. *et al.* Improving the documentation and findability of data services and repositories: A review of (meta) data management approaches. **Computers & Geosciences**, [S.l.], v. 169, p. 105194, 2022. Disponível em: <https://doi.org/10.1016/j.cageo.2022.105194>. Acesso em: 07 ago. 2025.

ROIĆ, M.; PIVAC, D. Standardization of Metadata of Analog Cadastral Documents Resulting from Systematic Cadaster Establishment. **Land**, Basel, v. 13, p. 1343, 2024. Disponível em: <https://doi.org/10.3390/land13091343>. Acesso em: 07 ago. 2025.

ROMANI, L. A. S. *et al.* AgroAPI platform: An initiative to support digital solutions for agribusiness ecosystems. **Smart Agricultural Technology**, [S.l.], v. 5, p. 100247, 2023.

ROUSSAKI, I. *et al.* Building an interoperable space for smart agriculture. **Digital Communications and Networks**, [S.l.], v. 9, p. 183-193, 2023. Disponível em: <https://doi.org/10.1016/j.dcan.2022.02.004>. Acesso em: 07 ago. 2025.

SADEGHI, M. *et al.* Interoperability of heterogeneous Systems of Systems: from requirements to a reference architecture. **The Journal of Supercomputing**, [S.l.], v. 80, p. 8954–8987, 2024.

SALES, T. P. *et al.* A FAIR catalog of ontology-driven conceptual models. **Data & Knowledge Engineering**, [S.l.], v. 147, p. 102210, 2023.

SAMOURKASIDIS, A.; ATHANASIADIS, I. N. A semantic approach for timeseries data fusion. **Computers and Electronics in Agriculture**, [S.l.], v. 169, p. 105171, 2020.

SAN EMETERIO DE LA PARTE, M. *et al.* Big Data and precision agriculture: a novel spatio-temporal semantic IoT data management framework for improved interoperability. **Journal of Big Data**, [S.l.], v. 10, n. 52, 2023.

SIGLER, L. *et al.* A Review of Technologies and Challenges for Integrated Modeling Analysis. **Archives of Computational Methods in Engineering**, [S.l.], v. 32, p. 1733-1762, 2025. Disponível em: <https://doi.org/10.1007/s11831-024-10187-3>. Acesso em: 07 ago. 2025.

SOARES, F. M.; SARAIVA, A. M.; DRUCKER, D. P. Linking Agrobiodiversity Data through Metadata Standards. **Biodiversity Information Science and Standards**, [S.l.], v. 4, e58928, 2020.

SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M. The NeOn Methodology for Ontology Engineering. In: SUÁREZ-FIGUEROA, M. C. *et al.* (ed.). **Ontology Engineering in a Networked World**. Berlin, Heidelberg: Springer, 2012. p. 9-34.

TAKEZAKI, A. *et al.* Common Agriculture Vocabulary for Enhancing Semantic-level Interoperability in Japan. **JARQ - Japan Agricultural Research Quarterly**, [S.l.], v. 54, n. 3, p. 219-225, 2020.

TAYUR, V. M.; SUCHITHRA, R. Multi-ontology mapping generative adversarial network in internet of things for ontology alignment. **Internet of Things**, [S.l.], v. 20, p. 100616, 2022.

THALHATH, N.; NAGAMORI, M.; SAKAGUCHI, T. Metadata application profile as a mechanism for semantic interoperability in FAIR and open data publishing. **Data and Information Management**, [S.l.], v. 9, p. 100068, 2025.

ULRICH, H. *et al.* Understanding the Nature of Metadata: Systematic Review. **Journal of Medical Internet Research**, [S.l.], v. 24, n. 1, e25440, 2022.

URDU, D. *et al.* Aligning interoperability architectures for digital agri-food platforms. **Computers and Electronics in Agriculture**, [S.l.], v. 224, p. 109194, 2024.

VARADHARAJAN, C. *et al.* BASIN-3D: A brokering framework to integrate diverse environmental data. **Computers & Geosciences**, [S.l.], v. 159, p. 105024, 2022.

VOGT, L. *et al.* Suggestions for extending the FAIR Principles based on a linguistic perspective on semantic interoperability. **Scientific Data**, [S.l.], 2025.

WORLD METEOROLOGICAL ORGANIZATION (WMO). **Manual on the WMO Information System: Volume II. WMO-No. 1060**. Geneva: WMO, 2023. Disponível em: <https://library.wmo.int/viewer/68731/?offset=#page=1&viewer=picture&o=bookmark&n=0&q=>. Acesso em: 08 jun. 2026.

WORLD METEOROLOGICAL ORGANIZATION (WMO). **WMO Core Metadata Profile (WCMP) Version 2**. Geneva: WMO, 2025.

WORLD WIDE WEB CONSORTIUM (W3C). **OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition)**. W3C Recommendation. Editado por Michael Schneider. [S.l.]: World Wide Web Consortium, 2012. Disponível em: <http://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/>. Acesso em: 15 out. 2025.

WORLD WIDE WEB CONSORTIUM (W3C). **RDF 1.2 Concepts and Abstract Data Model**. W3C Working Draft / Candidate Recommendation. Editado por Andy Seaborne *et al.* [S.l.]: World Wide Web Consortium, 2026a. Disponível em: <https://www.w3.org/TR/rdf12-concepts/>. Acesso em: 15 out. 2025.

WORLD WIDE WEB CONSORTIUM (W3C). **SPARQL 1.2 Protocol**. W3C Working Draft. Editado por Andy Seaborne *et al.* [S.l.]: World Wide Web Consortium, 2026b. Disponível em: <https://www.w3.org/TR/sparql12-protocol/>. Acesso em: 15 out. 2025.

WORLD WIDE WEB CONSORTIUM (W3C). **SPARQL 1.2 Query Language**. W3C Working Draft. Editado por Olaf Hartig *et al.* [S.l.]: World Wide Web Consortium, 2026c. Disponível em: <https://www.w3.org/TR/sparql12-query/>. Acesso em: 16 out. 2025.

WIECZOREK, J. *et al.* Darwin Core: An evolving community-developed biodiversity data standard. **PLoS ONE**, [S.l.], v. 7, n. 1, e29715, 2012. Disponível em: <https://doi.org/10.1371/journal.pone.0029715>. Acesso em: 18 out. 2025.

WU, J. *et al.* Boosting Climate Analysis With Semantically Uplifted Knowledge Graphs. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, [S.l.], v. 15, p. 4708-4718, 2022.

WU, J. *et al.* LinkedGeoClimate: An Interoperable Platform for Climate Data Access Within Geographical Context. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, [S.l.], 2024. Disponível em: <https://doi.org/10.1109/JSTARS.2024.3391922>. Acesso em: 07 ago. 2025.

XU, F. *et al.* Features of a FAIR vocabulary. **Journal of Biomedical Semantics**, [S.l.], v. 14, n. 6, 2023.

ZÁRATE, M. *et al.* Linked Open Biodiversity Data (LOBD): A semantic application for integrating biodiversity information. **Biodiversity Information Science and Standards**, [S.l.], v. 4, e58975, 2020.

ZEGINIS, D. *et al.* **A semantic meta-model for data integration and exploitation in precision agriculture and livestock farming**. [S.l.: s.n.], 2024.

ZHANG, F. *et al.* A multi-strategy ontology mapping method based on cost-sensitive SVM. **Journal of Cloud Computing**, [S.l.], v. 13, p. 144, 2024.

ZHANG, L.; LOBOV, A. Semantic Web Rule Language-based approach for implementing [design automation/knowledge models]. **Advanced Engineering Informatics**, [S.l.], v. 62, p. 102587, 2024.

APÊNDICE A – CÓDIGO-FONTE DO MOTOR DE TRIPLIFICAÇÃO EM PYTHON

```

import os
import json
import pandas as pd
import urllib.request
import urllib.parse
import xml.etree.ElementTree as ET
from fastapi import FastAPI, UploadFile, File, HTTPException
from fastapi.responses import JSONResponse
from rdflib import Graph, Namespace, URIRef, Literal, RDF, XSD

# Inicializa a API
app = FastAPI(
    title="AgroMeta Framework API",
    description="API REST para triplificação, resolução e integração semântica agrometeorológica.",
    version="1.0.0"
)

# Configuração de Namespaces
BASE_URI = "http://agrometa-framework.org/"
DWC = Namespace("http://rs.tdwg.org/dwc/terms/")
WCMP2 = Namespace("http://wmo-im.github.io/wcmp2/")
ISO19115 = Namespace("http://www.isotc211.org/2005/gmd/")
SOSA = Namespace("http://www.w3.org/ns/sosa/")
GEO = Namespace("http://www.opengis.net/ont/geosparql#")
TIME = Namespace("http://www.w3.org/2006/time#")
AGRO = Namespace("http://aims.fao.org/aos/agrontology#")
AGROMETA = Namespace("http://agrometa-framework.org/ontology/agrometa#")

CACHE_RESOLVER_AGROVOC = {}
CACHE_RELACOES_AGROVOC = {}

def buscar_uri_agrovoc_automatico(nome_cientifico: str) -> str:
    """Busca a URI do AGROVOC corrigindo codificação e idioma"""
    partes = nome_cientifico.replace('J.E.', '').split(',')
    palavras = partes[0].strip().split(' ')
    nome_limpo = " ".join(palavras[:2]) if len(palavras) >= 2 else palavras[0]

    if nome_limpo in CACHE_RESOLVER_AGROVOC:
        return CACHE_RESOLVER_AGROVOC[nome_limpo]

    headers = {'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)', 'Accept': 'application/json'}

    # 1. Tenta API REST sem restrição de idioma (pois nomes científicos são Latim)
    try:
        params = urllib.parse.urlencode({'query': nome_limpo})

```

```

url = f"https://agrovoc.fao.org/agrovoc/rest/v1/search?{params}"
req = urllib.request.Request(url, headers=headers)
with urllib.request.urlopen(req, timeout=10) as response:
    data = json.loads(response.read().decode('utf-8'))
    if data and data.get('results'):
        uri_fao = data['results'][0].get('uri')
        if uri_fao:
            CACHE_RESOLVER_AGROVOC[nome_limpo] = uri_fao
        return uri_fao
except Exception:
    pass

# 2. Tenta SPARQL Endpoint como fallback
try:
    query = f"""
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT ?uri WHERE {{
    ?uri skos:prefLabel|skos:altLabel ?label .
    FILTER (strstarts(lcase(str(?label)), "{nome_limpo.lower()}"))
}} LIMIT 1
"""

    params = urllib.parse.urlencode({'query': query, 'format': 'json'})
    url = f"https://agrovoc.fao.org/sparql?{params}"
    req = urllib.request.Request(url,
                                headers={'User-Agent': 'Mozilla/5.0', 'Accept': 'application/sparql-
results+json'})
    with urllib.request.urlopen(req, timeout=10) as response:
        data = json.loads(response.read().decode('utf-8'))
        bindings = data.get('results', {}).get('bindings', [])
        if bindings:
            uri_fao = bindings[0]['uri']['value']
            CACHE_RESOLVER_AGROVOC[nome_limpo] = uri_fao
            return uri_fao
except Exception:
    pass

# Fallback local seguro
uri_local = f"http://agrometa-framework.org/taxonomy/{nome_limpo.lower().replace(' ',
_)}"
CACHE_RESOLVER_AGROVOC[nome_limpo] = uri_local
return uri_local

def obter_relacoes_afetacao(uri_praga: str):
    """Consulta dinamicamente as afetações da praga de forma segura e limpa"""

    # PROTEÇÃO CRÍTICA: Só envia para a FAO se for uma URI válida da FAO.
    # Isso evita os erros 400 Bad Request que apareceram no seu print.
    if not uri_praga.startswith("http://aims.fao.org/aos/agrovoc"):
        return []

```

```

if uri_praga in CACHE_RELACOES_AGROVOC:
    return CACHE_RELACOES_AGROVOC[uri_praga]

try:
    query = f"""
    SELECT DISTINCT ?cultura WHERE {{
        {{ <{uri_praga}> <http://aims.fao.org/aos/agrontology#affects> ?cultura . }}
        UNION
        {{ <{uri_praga}> <http://aims.fao.org/aos/agrontology#isPestOf> ?cultura . }}
        UNION
        {{ <{uri_praga}> <http://aims.fao.org/aos/agrontology#afflicts> ?cultura . }}
        UNION
        {{ ?cultura <http://aims.fao.org/aos/agrontology#hasPest> <{uri_praga}> . }}
        UNION
        {{ <{uri_praga}> <http://aims.fao.org/aos/agrontology#hasHost> ?cultura . }}
    }}
    """
    # Correção do Erro 400: urlencode padronizado
    params = urllib.parse.urlencode({'query': query, 'format': 'json'})
    url = f"https://agrovoc.fao.org/sparql?{params}"

    req = urllib.request.Request(url, headers={
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)',
        'Accept': 'application/sparql-results+json'
    })
    with urllib.request.urlopen(req, timeout=10) as response:
        data = json.loads(response.read().decode('utf-8'))
        culturas = [res['cultura']['value'] for res in data['results']['bindings']]
        CACHE_RELACOES_AGROVOC[uri_praga] = culturas
        return culturas
except Exception as e:
    print(f"Aviso interno: Falha ao buscar cultura na FAO para {uri_praga} - {str(e)}")
    return []

```

```

def inicializar_grafo():
    g = Graph()
    g.bind("rdf", RDF)
    g.bind("dwc", DWC)
    g.bind("sosa", SOSA)
    g.bind("geo", GEO)
    g.bind("time", TIME)
    g.bind("wcmp2", WCMP2)
    g.bind("iso19115", ISO19115)
    g.bind("agro", AGRO)
    g.bind("agrometa", AGROMETA)
    return g

```

```

# =====
# ENDPOINT DE ENGINE: PROCESSAMENTO E ENTREGA
# =====
@app.post("/api/v1/triplificar/")
async def gerar_grafo_conhecimento(
    biodiversidade_dwc: UploadFile = File(...),
    catalogo_wcmp2: UploadFile = File(...),
    observacoes_meteorologicas: UploadFile = File(...),
    proveniencia_iso19115: UploadFile = File(...)
):
    try:
        df_gbif = pd.read_csv(biodiversidade_dwc.file, sep='\t', low_memory=False)
        meta_json = json.load(catalogo_wcmp2.file)
        df_inmet = pd.read_csv(observacoes_meteorologicas.file, sep=';', skiprows=8,
encoding='iso-8859-1')
        tree = ET.parse(proveniencia_iso19115.file)

        # Instanciação do motor e parse (pré-processamento) do esquema ontológico
        meu_grafo = inicializar_grafo()
        if os.path.exists("mapeamento-agrometa.owl"):
            meu_grafo.parse("mapeamento-agrometa.owl", format="xml")

# =====
# MÓDULO BIODIVERSIDADE (Darwin Core)
# =====
for _, row in df_gbif.iterrows():
    gbif_id = str(row['gbifID'])
    occurrence_uri = URIRef(BASE_URI + f'biodiversity/occurrence_{gbif_id}')
    geom_uri = URIRef(BASE_URI + f'spatial/geom_{gbif_id}')

    meu_grafo.add((occurrence_uri, RDF.type, DWC.Occurrence))
    meu_grafo.add((occurrence_uri, RDF.type, GEO.Feature))
    meu_grafo.add((geom_uri, RDF.type, GEO.Geometry))

    nome_cientifico_str = str(row['scientificName']).strip()
    meu_grafo.add(
        (occurrence_uri, AGROMETA.hasScientificName, Literal(nome_cientifico_str,
datatype=XSD.string)))

    lat, lon = float(row['decimalLatitude']), float(row['decimalLongitude'])
    meu_grafo.add((occurrence_uri, AGROMETA.hasLatitude, Literal(lat,
datatype=XSD.float)))
    meu_grafo.add((occurrence_uri, AGROMETA.hasLongitude, Literal(lon,
datatype=XSD.float)))

    wkt_ponto = f"POINT({lon} {lat})"
    meu_grafo.add((occurrence_uri, GEO.hasGeometry, geom_uri))
    meu_grafo.add((geom_uri, GEO.asWKT, Literal(wkt_ponto,
datatype=GEO.wktLiteral)))

```

```

if pd.notna(row['eventDate']):
    meu_grafo.add(
        (occurrence_uri, AGROMETA.hasEventDate, Literal(str(row['eventDate'][:10],
datatype=XSD.string)))

# 1. Resolve a URI no AGROVOC
uri_final_vocabulo = buscar_uri_agrovoc_automatico(nome_cientifico_str)
uri_agrovoc = URIRef(uri_final_vocabulo)
meu_grafo.add((occurrence_uri, AGROMETA.hasAgrovocTerm, uri_agrovoc))

# 2. Materializa a relação com a cultura (Ponte Semântica)
culturas_afetadas = obter_relacoes_afetacao(str(uri_agrovoc))
for cultura in culturas_afetadas:
    meu_grafo.add((uri_agrovoc, AGRO.affects, URIRef(cultura)))

# =====
# MÓDULO METEOROLOGIA (WCMP2 + SOSA)
# =====
dataset_id = meta_json['id']
dataset_uri = URIRef(f'http://agrometa-framework.org/meteorology/{dataset_id}')

meu_grafo.add((dataset_uri, RDF.type, WCMP2.Record))
meu_grafo.add(
    (dataset_uri, AGROMETA.hasObservedProperty,
    Literal(meta_json['properties']['title'], datatype=XSD.string)))
meu_grafo.add((dataset_uri, WCMP2.dataPolicy,
    Literal(meta_json['properties'].get('wmo:dataPolicy', 'core'),
datatype=XSD.string)))

col_data, col_chuva = df_inmet.columns[0], df_inmet.columns[2]
df_inmet[col_data] = df_inmet[col_data].astype(str).str.replace('/', '-')
df_inmet[col_chuva] = pd.to_numeric(df_inmet[col_chuva].astype(str).str.replace(',', '.'),
errors='coerce')
df_inmet = df_inmet[df_inmet[col_chuva] != -9999.0]
df_diario = df_inmet.groupby(col_data)[col_chuva].sum().reset_index()

for _, row in df_diario.iterrows():
    data_string = str(row[col_data])
    obs_uri = URIRef(f'http://agrometa-
framework.org/meteorology/observation_{data_string}')
    meu_grafo.add((obs_uri, RDF.type, SOSA.Observation))
    meu_grafo.add((obs_uri, AGROMETA.hasEventDate, Literal(data_string,
datatype=XSD.string)))
    meu_grafo.add((obs_uri, AGROMETA.hasPrecipitationValue,
    Literal(float(row[col_chuva]), datatype=XSD.float)))
    meu_grafo.add((obs_uri, AGROMETA.hasStationID,
    Literal("API_GENERATED_STATION", datatype=XSD.string)))
    meu_grafo.add((dataset_uri, AGROMETA.hasObservationMember, obs_uri))

# =====

```

```

# MÓDULO PROVENIÊNCIA (ISO 19115 XML)
# =====
root = tree.getroot()
ns = {'gmd': 'http://www.isotc211.org/2005/gmd', 'gco':
'http://www.isotc211.org/2005/gco'}

file_id_node = root.find('..gmd:fileIdentifier/gco:CharacterString', ns)
file_id = file_id_node.text.strip() if file_id_node is not None else "metadata_via_api"
metadata_uri = URIRef(BASE_URI + f'institution/metadata_{file_id}')

meu_grafo.add((metadata_uri, RDF.type, ISO19115.MD_Metadata))
meu_grafo.add((metadata_uri, RDF.type, GEO.Feature))

org_nodes = root.findall('..gmd:organisationName/gco:CharacterString', ns)
org_name = "Instituição Não Especificada"
for node in org_nodes:
    if node.text and node.text.strip():
        org_name = node.text.strip()
        break
meu_grafo.add((metadata_uri, AGROMETA.hasInstitution, Literal(org_name,
datatype=XSD.string)))

statement_node = root.find('..gmd:LI_Lineage/gmd:statement/gco:CharacterString', ns)
if statement_node is not None and statement_node.text:
    lineage_text = " ".join(statement_node.text.split())
    meu_grafo.add((metadata_uri, AGROMETA.hasLineage, Literal(lineage_text,
datatype=XSD.string)))

# =====
# TECIMENTO DOS ELOS SEMÂNTICOS
# =====
registro_meteorologico = URIRef("http://agrometa-
framework.org/meteorology/urn:wmo:md:br-inmet:daycli")

for ocorrencia in meu_grafo.subjects(RDF.type, DWC Occurrence):
    meu_grafo.add((ocorrencia, AGROMETA.hasMeteorologicalRecord,
registro_meteorologico))
    meu_grafo.add((ocorrencia, AGROMETA.hasClimaticContext, metadata_uri))

# =====
# SERIALIZAÇÃO E ENTREGA
# =====
arquivo_saida = "grafo_final_api_integrated.ttl"
meu_grafo.serialize(destination=arquivo_saida, format="turtle")
grafo_contenido_turtle = meu_grafo.serialize(format="turtle")

return JsonResponse(content={
    "status": "Sucesso",
    "mensagem": "Grafo de conhecimento gerado com harmonização semântica de
afetação em tempo real.",

```

```
        "total_triplas_processadas": len(meu_grafo),
        "arquivo_gerado_no_servidor": arquivo_saida,
        "grafo_rdf_turtle": grafo_conteudo_turtle
    })

except Exception as e:
    raise HTTPException(status_code=500, detail=f"Erro no pipeline de triplificação:
{str(e)}")
```

APÊNDICE B – ONTOLOGIA DE ALINHAMENTO COMPLETA

```

<?xml version="1.0"?>
<rdf:RDF xmlns="http://agrometa-framework.org/ontology/agrometa#"
  xml:base="http://agrometa-framework.org/ontology/agrometa"
  xmlns:dwc="http://rs.tdwg.org/dwc/terms/"
  xmlns:geo="http://www.opengis.net/ont/geosparql#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:agro="http://aims.fao.org/aos/agrontology#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:sosa="http://www.w3.org/ns/sosa/"
  xmlns:time="http://www.w3.org/2006/time#"
  xmlns:wcmp2="http://wmo-im.github.io/wcmp2/"
  xmlns:agrometa="http://agrometa-framework.org/ontology/agrometa#"
  xmlns:iso19115="http://www.isotc211.org/2005/gmd/">
  <owl:Ontology rdf:about="http://agrometa-framework.org/ontology/agrometa"/>

  <!--
  //////////////////////////////////////
  //
  // Object Properties
  //
  //////////////////////////////////////
  -->

  <!-- http://agrometa-framework.org/ontology/agrometa#hasClimaticContext -->

  <owl:ObjectProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasClimaticContext">
    <rdfs:domain rdf:resource="http://rs.tdwg.org/dwc/terms/Occurrence"/>
    <rdfs:range rdf:resource="http://www.isotc211.org/2005/gmd/MD_Metadata"/>
  </owl:ObjectProperty>

  <!-- http://agrometa-framework.org/ontology/agrometa#hasMeteorologicalRecord -->

  <owl:ObjectProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasMeteorologicalRecord">
    <rdfs:domain rdf:resource="http://rs.tdwg.org/dwc/terms/Occurrence"/>
    <rdfs:range rdf:resource="http://wmo-im.github.io/wcmp2/Record"/>
  </owl:ObjectProperty>

  <!-- http://aims.fao.org/aos/agrontology#affects -->

  <owl:ObjectProperty rdf:about="http://aims.fao.org/aos/agrontology#affects"/>

  <!--
  //////////////////////////////////////

```

```

//
// Data properties
//
////////////////////////////////////
-->

<!-- http://agrometa-framework.org/ontology/agrometa#hasEventDate -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasEventDate">
  <rdfs:domain rdf:resource="http://rs.tdwg.org/dwc/terms/Occurrence"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasInstitution -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasInstitution">
  <rdfs:domain rdf:resource="http://www.isotc211.org/2005/gmd/MD_Metadata"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasLatitude -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasLatitude">
  <rdfs:domain rdf:resource="http://rs.tdwg.org/dwc/terms/Occurrence"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasLineage -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasLineage">
  <rdfs:domain rdf:resource="http://www.isotc211.org/2005/gmd/MD_Metadata"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasLongitude -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasLongitude">
  <rdfs:domain rdf:resource="http://rs.tdwg.org/dwc/terms/Occurrence"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasObservedProperty -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasObservedProperty">

```

```

    <rdfs:domain rdf:resource="http://wmo-im.github.io/wcmp2/Record"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasScientificName -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasScientificName">
    <rdfs:domain rdf:resource="http://rs.tdwg.org/dwc/terms/Occurrence"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<!-- http://agrometa-framework.org/ontology/agrometa#hasStationID -->

<owl:DatatypeProperty rdf:about="http://agrometa-
framework.org/ontology/agrometa#hasStationID">
    <rdfs:domain rdf:resource="http://wmo-im.github.io/wcmp2/Record"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>

<!--
////////////////////////////////////
//
// Classes
//
////////////////////////////////////
-->

<!-- http://rs.tdwg.org/dwc/terms/Occurrence -->

<owl:Class rdf:about="http://rs.tdwg.org/dwc/terms/Occurrence">
    <rdfs:subClassOf rdf:resource="http://www.opengis.net/ont/geosparql#Feature"/>
</owl:Class>

<!-- http://wmo-im.github.io/wcmp2/Record -->

<owl:Class rdf:about="http://wmo-im.github.io/wcmp2/Record">
    <rdfs:subClassOf rdf:resource="http://www.opengis.net/ont/geosparql#Feature"/>
</owl:Class>

<!-- http://www.isotc211.org/2005/gmd/MD_Metadata -->

<owl:Class rdf:about="http://www.isotc211.org/2005/gmd/MD_Metadata">
    <rdfs:subClassOf rdf:resource="http://www.opengis.net/ont/geosparql#Feature"/>
</owl:Class>

<!-- http://www.opengis.net/ont/geosparql#Feature -->

<owl:Class rdf:about="http://www.opengis.net/ont/geosparql#Feature"/>

```

```

<!-- http://www.w3.org/2006/time#Instant -->

<owl:Class rdf:about="http://www.w3.org/2006/time#Instant">
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2006/time#TemporalEntity"/>
</owl:Class>

<!-- http://www.w3.org/2006/time#Interval -->

<owl:Class rdf:about="http://www.w3.org/2006/time#Interval">
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2006/time#TemporalEntity"/>
</owl:Class>

<!-- http://www.w3.org/2006/time#TemporalEntity -->

<owl:Class rdf:about="http://www.w3.org/2006/time#TemporalEntity"/>

<!-- http://www.w3.org/ns/sosa/Observation -->

<owl:Class rdf:about="http://www.w3.org/ns/sosa/Observation"/>
</rdf:RDF>

<!-- Generated by the OWL API (version 4.5.29.2024-05-13T12:11:03Z)
https://github.com/owlcs/owlapi -->

```