

Proposta de Matriz de Avaliação da Qualidade de Metadados Agrometeorológicos

Rafael Henrique Mauda Scos¹, Wesley Vinicius Fernandes²,
Maria Salete Marcon Gomes Vaz³

¹Universidade Estadual de Ponta Grossa, Ponta Grossa, Brasil (22014026@uepg.br)

²Universidade Estadual de Ponta Grossa, Ponta Grossa, Brasil (240401400010@uepg.br)

³Universidade Estadual de Ponta Grossa, Ponta Grossa, Brasil (salete@uepg.br)

Resumo: A diversidade de padrões de metadados (ISO 19115, Darwin Core e WCMP) atua como barreira à interoperabilidade agrometeorológica. Este trabalho objetivou construir uma Matriz de Critérios de Qualidade para avaliar a conformidade técnica desses padrões. A metodologia aplicou a taxonomia de Wang e Strong para formular regras binárias. O resultado consolidou um modelo prático que diagnostica falhas semânticas, promovendo a integração de dados e os princípios FAIR na agricultura digital.

Palavras-chave: Qualidade de Dados; Interoperabilidade Semântica; Metadados; Agrometeorologia; Princípios FAIR.

INTRODUÇÃO

A crescente demanda por dados agrometeorológicos confiáveis, impulsionada pelo atual cenário de mudanças climáticas, exige sistemas e plataformas que promovam a integração e o uso de informações precisas. Contudo, a diversidade de padrões de metadados existentes, a exemplo da norma ISO 19115 para informações geoespaciais, do *Darwin Core* para registros de agrobiodiversidade e do *WMO Core Metadata Profile* (WCMP) para observações meteorológicas, atua como uma barreira técnica e semântica para a interoperabilidade dos sistemas. A divergência estrutural entre esses padrões resulta na formação de ilhas de dados (*AgGateway*, 2023), dificultando análises integradas de variáveis climáticas e biológicas.

Para mitigar esse problema, torna-se importante promover a integração estrutural e garantir e mensurar a qualidade técnica desses metadados. A qualidade da informação não se limita apenas à sua exatidão, devendo ser avaliada sob múltiplas dimensões que atendam às necessidades dos consumidores dos dados (Wang e Strong, 1996). Tais dimensões são categorizadas em quatro grupos: i) qualidade intrínseca, a qualidade do dado por si só; ii) qualidade contextual, avaliada dentro do contexto da tarefa; iii) qualidade representacional, a clareza e inteligibilidade do formato; e iv) qualidade de acessibilidade, a facilidade e segurança na obtenção do dado. A ausência de modelos padronizados para diagnosticar tais atributos dificulta a interpretação rápida e prejudica a reusabilidade dos conjuntos de dados (Wilkinson et al., 2016; Michener, 2006).

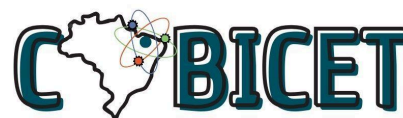
Nesse contexto, este trabalho concentrou-se na modelagem de um sistema de avaliação de qualidade para metadados agrometeorológicos. O objetivo principal da pesquisa consistiu em construir e consolidar uma Matriz de Critérios de Qualidade para os padrões *Darwin Core*, WCMP e ISO 19115, estabelecendo indicadores técnicos e objetivos baseados nas dimensões de qualidade de dados. Dessa forma, o estudo visou realizar a transição da fundamentação teórica sobre qualidade de dados para um modelo de avaliação prático e computável, fornecendo a base lógica estruturada para a identificação de lacunas semânticas e para a promoção direta da interoperabilidade na agricultura digital.

MATERIAL E MÉTODOS

Esta pesquisa caracteriza-se como um estudo de natureza aplicada, descritiva e exploratória, voltado para a modelagem de critérios de avaliação de qualidade de dados agrometeorológicos. A execução da pesquisa foi estruturada em quatro etapas consecutivas e complementares:

Etapas A: Definição Taxonomica (Teórica)

A base teórica para a avaliação foi estabelecida a partir da taxonomia de Wang e Strong (1996), que categoriza a qualidade de dados em quatro categorias fundamentais: intrínseca, contextual, representacional e de acessibilidade. A partir deste referencial, realizou-se o estudo e a seleção das dimensões específicas, tais como acurácia, completude, consistência e temporalidade, aplicáveis à avaliação de metadados agrometeorológicos.



Etapa B: Extração de Requisitos (Técnica)

Esta fase consistiu na auditoria técnica das documentações oficiais e normas internacionais dos padrões selecionados: ISO 19115-1:2014 (para dados geoespaciais), WMO *Core Metadata Profile* - WCMP (para dados meteorológicos) e *Darwin Core* - DwC (para agrobiodiversidade). Foram identificados os elementos e atributos técnicos críticos para garantir a interoperabilidade no domínio agrometeorológico.

Etapa C: Formulação dos Critérios (Métrica)

Os requisitos técnicos identificados na etapa anterior foram transformados em regras de validação lógica. Cada requisito foi convertido numa pergunta de verificação, capaz de determinar, de forma binária (sim/não), o sucesso ou a falha do metadado naquele aspecto específico. Como exemplo, para o padrão ISO 19115, formulou-se a validação do uso de vocabulários controlados (*CodeLists*), enquanto para o *Darwin Core*, verificou-se a presença de campos de identidade taxonômica, como reino e família.

Etapa D: Construção da Matriz de Critérios

Os resultados da formulação de métricas foram consolidados numa matriz estruturada. Esta matriz associou cada padrão de metadados analisado à sua dimensão de qualidade correspondente, ao campo técnico avaliado e ao critério de qualidade final sob a forma de pergunta. Este instrumento constitui a base lógica fundamental para o diagnóstico sistemático da qualidade e para a promoção da integração semântica entre os padrões analisados.

RESULTADOS E DISCUSSÃO

O principal resultado obtido foi a construção e consolidação da Matriz de Critérios de Qualidade. Este instrumento representa a transição da fundamentação teórica sobre qualidade de dados, a partir das dimensões de qualidade de Wang e Strong (1996), para um modelo de avaliação prático e computável, aplicável aos padrões de metadados agrometeorológicos. O mapeamento de dezenas de regras de validação, modeladas num formato binário (sim/não), permite verificar objetivamente a conformidade técnica e semântica dos ficheiros.

Desse modo, operacionalizou-se as dimensões de qualidade, convertendo conceitos como acurácia, completude, consistência e acessibilidade em critérios objetivos aplicáveis aos três padrões analisados. Os resultados específicos para cada padrão são detalhados a seguir:

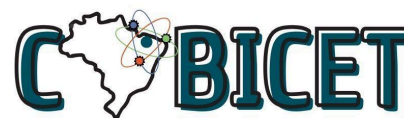
Padrão *Darwin Core* (DwC): A formulação incidiu nas dimensões de Interpretabilidade e Reputação. Evidenciou-se que a ausência de uma hierarquia taxonômica clara (família e reino) gera ambiguidades

na identificação de plantas e pragas agrícolas (Meyer et al., 2021; TDWG, 2023). Foram estabelecidos critérios para validar campos como *dwc:scientificName*, *dwc:family* e *dwc:taxonRank*, além de reforçar a rastreabilidade através da identificação do especialista (*dwc:identifiedBy*) e da data do registo (*dwc:dateIdentified*).

Padrão WMO *Core Metadata Profile* (WCMP): Os critérios focaram-se na estruturação e acessibilidade global. Estabeleceu-se a obrigatoriedade do idioma inglês e a categorização conforme a hierarquia oficial da Organização Meteorológica Mundial (*properties.wmo:topicHierarchy*) como métricas essenciais de interoperabilidade, uma vez que o idioma inglês é o padrão exigido pela WMO para viabilizar a descoberta e a troca de dados em nível global (WMO, 2023). A verificação de políticas de dados (*properties.wmo:dataPolicy*) e contatos institucionais revelou-se fundamental para garantir a reutilização das informações meteorológicas, pois essas propriedades classificam legalmente as permissões de acesso (dados essenciais ou adicionais) e asseguram o intercâmbio irrestrito conforme as resoluções internacionais (WMO, 2023).

Padrão ISO 19115: A matriz destacou as dimensões de Completude, Consistência e Acurácia. A validação baseada em vocabulários controlados (*CodeLists*), em detrimento do uso de texto livre no campo *MD_TopicCategoryCode*, mostrou-se crucial para assegurar a consistência semântica e a integridade estrutural exigida por sistemas de informação geográfica, prevenindo divergências terminológicas (ISO, 2014; Han e Han, 2022). Adicionalmente, critérios relacionados com a resolução espacial e a linhagem dos dados (*lineage*) reforçaram a acurácia necessária para aplicações de agricultura de precisão, ao documentarem com rigor a proveniência e a densidade dos dados geoespaciais (ISO, 2014).

De forma integrada, os resultados evidenciaram que os desafios da interoperabilidade agrometeorológica envolvem profundas diferenças conceituais entre os padrões. A matriz desenvolvida permitiu explicitar essas divergências e estabelecer um ponto de convergência baseado em critérios técnicos de qualidade, constituindo um avanço para a integração de dados heterogêneos no contexto da agricultura digital.



CONCLUSÃO

O desenvolvimento deste estudo permitiu a sistematização e a consolidação de um modelo técnico para a avaliação de metadados agrometeorológicos. A construção da Matriz de Critérios de Qualidade representou o alcance do objetivo principal, viabilizando a operacionalização das dimensões teóricas de acurácia, completude, consistência e acessibilidade em métricas objetivas e aplicáveis.

A estruturação das regras de validação em formato binário (sim/não) constituiu um avanço metodológico significativo, pois eliminou a subjetividade na análise da conformidade técnica e estabeleceu a base lógica necessária para o diagnóstico escalável de metadados. Evidenciou-se que a integração de dados heterogêneos no domínio da agricultura digital depende, primordialmente, da superação de barreiras semânticas e estruturais, as quais foram devidamente mapeadas e mitigadas através dos pontos de convergência estabelecidos entre os padrões *Darwin Core*, *WCMP* e *ISO 19115*.

Concluiu-se que o levantamento de requisitos e a modelação das métricas foram finalizados com êxito, resultando num instrumento robusto que favorece a comparação entre diferentes padrões e contribui diretamente para a transparência e reusabilidade de dados científicos. Este trabalho consolidou, portanto, uma base metodológica consistente para apoiar a governança de dados e a integração semântica no contexto agrometeorológico.

AGRADECIMENTOS

Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de iniciação científica concedida. Agradecimentos também à orientadora Prof^a. Dr^a. Maria Salete Marcon Gomes Vaz e ao coorientador Wesley Vinicius Fernandes pelo direcionamento e apoio, e à Universidade Estadual de Ponta Grossa (UEPG) pela infraestrutura e fomento à pesquisa.

REFERÊNCIAS

- AGGATEWAY. Interoperability in Agriculture: 2023 Status Report. [S. l.], 2023.
- BIODIVERSITY INFORMATION STANDARDS (TDWG). Darwin Core Quick Reference Guide. [S. l.], 2023.
- HAN, S.; HAN, J. Case study on an integrated interoperable metadata model for geoscience information resources. *Geoscience Data Journal*, [S. l.], v. 9, p. 355-370, 2022.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 19115-1:2014:

geographic information: metadata: part 1: fundamentals. Geneva: ISO, 2014.

- MEYER, R. et al. Aligning Standards Communities: Sustainable Darwin Core MIxS Interoperability. *Biodiversity Information Science and Standards*, [S. l.], v. 5, e73775, 2021.
- MICHENER, William K. Meta-information concepts for ecological data management. *Ecological Informatics*, [S. l.], v. 1, n. 1, p. 3–7, 2006.
- WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems*, [S. l.], v. 12, n. 4, p. 5-33, 1996.
- WILKINSON, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, [S. l.], v. 3, 2016.
- WORLD METEOROLOGICAL ORGANIZATION. Manual on the WMO Information System: Volume I - Annex VII to the WMO Technical Regulations. Geneva: WMO, 2023.