

Filipi Miranda Soares

Universidade Federal de Minas Gerais
filipivgp2011@gmail.com

**Benildes Coura Moreira dos Santos
Maculan**

Universidade Federal de Minas Gerais
benildes@gmail.com

Debora Pignatari Drucker

*Empresa Brasileira de Pesquisa
Agropecuária*
debora.drucker@embrapa.br

Universidade Federal de Minas Gerais

Correspondência/Contato
Av. Antônio Carlos, 6627
Pampulha: 31270-901
BELO HORIZONTE - MG

Escola de Ciência da Informação da UFMG

PADRÃO DE METADADOS DARWIN CORE

Proposta de extensão para as interações ecológicas no contexto da agrobiodiversidade

RESUMO

Apresenta uma proposta de princípios para criação de uma extensão de metadados para o padrão Darwin Core para o domínio da agrobiodiversidade, com recorte temático nas interações ecológicas. O objetivo da pesquisa é apresentar princípios e recomendações de melhores práticas para a criação de metadados como uma extensão ao Darwin Core. A metodologia consistiu em analisar o modelo <onto.biodiversidade>, que é um modelo conceitual para representação de dados da biodiversidade desenvolvido pela Embrapa, para verificar a possibilidade de usar esse modelo como insumo terminológico. Além desse modelo, foi feita análise da literatura especializada para compreender as interações ecológicas. O núcleo de termos do padrão de metadados Darwin Core também foi analisado a fim de se compreender os aspectos semânticos e sintáticos do Darwin Core, que são fundamentais para a criação da extensão de metadados. Ao final do artigo, apresenta-se uma amostra dos resultados que se espera obter ao final da aplicação da metodologia que está em fase de desenvolvimento.

Palavras-Chave: Metadados, Darwin Core, Agrobiodiversidade.

ABSTRACT

It presents a proposal of principles for creating a metadata extension for the Darwin Core standard for the agrobiodiversity domain, with thematic focus on ecological interactions. The goal of the research is to present best practice principles and recommendations for metadata creation as an extension to Darwin Core. The methodology consisted of analyzing the <onto.biodiversity> model, which is a conceptual model for the representation of biodiversity data developed by Embrapa, to verify the possibility of using this model as a terminological input. In addition to this model, an analysis of the specialized literature was made to understand the eco-logical interactions. The core terms of the Darwin Core metadata standard have also been analyzed in order to understand the semantic and syntactic aspects of Darwin Core, which are key to creating metadata extension. At the end of the article, we present a sample of the results expected to be obtained at the end of the application of the methodology that is under development.

Keywords: Metadata, Darwin Core, Agrobiodiversity.

1. INTRODUÇÃO

No contexto atual de demanda crescente de produção de alimentos, conhecer a biodiversidade é essencial para que seja possível gerir os recursos biológicos de forma eficiente, o que significa o melhor aproveitamento dos recursos biológicos para o aumento da produção, com impacto ambiental reduzido. O campo científico que engloba o estudo da biodiversidade nos sistemas agrícolas é denominado agrobiodiversidade. Os trabalhos científicos desse campo têm como resultados artigos, livros, relatórios técnicos e outros, dispostos em diferentes bancos de dados. Esses dados são de suma importância para o manejo adequado da biodiversidade, entretanto, muitas vezes encontram-se dispersos em bancos de dados locais e pouco acessíveis. De forma a tornar mais eficiente a gestão desses dados, recomenda-se o uso de metadados. Despontam neste contexto, padrões de metadados voltados à biodiversidade, como Access to Biological Collections Data schema (ABCD), Darwin Core (DwC) Ecological Metadata Language (EML) e Plinian Code. Dentre esses, o padrão de metadados DwC foi escolhido como objeto de estudo desta pesquisa por ser o mais utilizado mundialmente para a descrição de registros de ocorrência de espécies e descrição de coleções de espécimes (GBIF, 2019, *online*).

O padrão DwC é um conjunto de metadados para descrição de ocorrências de espécies, o que significa representar “taxa, sua ocorrência na natureza conforme documentado por observações, espécimes, amostras e informações relacionadas” (DARWIN CORE TASK GROUP, 2009, *online*). Possui um núcleo básico de metadados, “com o objetivo de facilitar o compartilhamento de informações sobre a diversidade biológica, fornecendo definições de referência, exemplos e comentários” (SISTEMA DE INFORMAÇÃO SOBRE A BIODIVERSIDADE BRASILEIRA, 2018, *online*).

O DwC não possui metadados para representar dados específicos da agrobiodiversidade, como informações sobre as interações ecológicas. Considerando a importância dos dados da agrobiodiversidade para a economia e para a sociedade, aponta-se a necessidade de criação de novos metadados em complemento ao DwC para a representação deste contexto. O presente trabalho apresenta uma pesquisa de mestrado em andamento, desenvolvida no Programa de Pós-Graduação em Gestão e Organização do Conhecimento na Universidade Federal de Minas Gerais (UFMG), em parceria com a Empresa Brasileira de Pesquisa Agropecuária (Embrapa), que se propõe a criar uma

extensão de metadados ao padrão DwC para atender à temática da agrobiodiversidade, notadamente no contexto das interações entre espécies.

2. FUNDAMENTOS TEÓRICO-METODOLÓGICOS

A representação da informação proveniente de qualquer campo do conhecimento preconiza a imersão do pesquisador nesse dado campo de conhecimento, a fim de que ele entenda as relações estabelecidas entre os dados que são produzidos como resultado deste trabalho. Sendo assim, esta seção brevemente apresenta os conceitos e fundamentos teórico-metodológicos que norteiam as temáticas investigadas, para o desenvolvimento da pesquisa.

2.1. Agrobiodiversidade

A agrobiodiversidade “inclui todos os componentes da biodiversidade que têm relevância para a agricultura e alimentação”, que são “as variedades e a variabilidade de animais, plantas e de microrganismos, nos níveis genético, de espécies e de ecossistemas os quais são necessários para sustentar as funções-chaves dos agroecossistemas, suas estruturas e processos” (CONVENTION ON BIOLOGICAL DIVERSITY, 2000, p. 85).

Ela relaciona-se intrinsecamente com a Ecologia, ao estudar os relacionamentos dos organismos com os ambientes onde habitam, que, no caso da agrobiodiversidade, são os agroecossistemas (SOARES; MACULAN; DRUCKER, 2019). Nesse campo, destaca-se o estudo das interações ecológicas. Populações de diferentes espécies podem interagir entre si, o que pode resultar em variados graus de afetação a ambas as partes (CAIN; BOWMAN; HACKER, 2018; COLIN; HARPER, 2007; ODUM; BARRET, 2006).

Conhecer essas interações é de suma importância para o manejo agrícola, de maneira que possam ser utilizadas como recursos biológicos na produção de alimentos. O conhecimento de que uma espécie preda outra, por exemplo, é determinante para saber qual espécie deve ser utilizada no controle biológico de pragas em lavouras. Além da interação do tipo predatismo, existem diversas outras que estão em fase de exploração nesta pesquisa e que farão parte dos metadados que serão criados.

2.2. Metadados

A recuperação da informação depende, principalmente, do uso correto de metadados. De acordo com Zeng e Qin (2008), metadados descrevem o que a informação é (tipo de registro), sobre o que trata (assunto) e onde pode ser localizada. Em outros termos, metadados são informações criadas para descrever coisas e permitir que pessoas interajam com essas coisas para obter conhecimento sobre elas (RILEY, 2017). “Metadados são a chave para a funcionalidade dos sistemas que proveem conteúdo, possibilitando aos usuários encontrar itens de interesse, registrar informações essenciais sobre eles e compartilhar informações com outros.” (RILEY, 2017, p. 2, tradução nossa).

De acordo com Pomerantz (2015, p. 26, tradução nossa) “metadados são declarações sobre um objeto potencialmente informativo”. Para que essas declarações sejam feitas, é necessário que haja duas condições básicas: “alguma coisa para se fazer uma declaração sobre” e “algo para se dizer sobre essa coisa” (POMERANTZ, 2015, p. 27, tradução nossa). Este “objeto potencialmente informativo” é algo sobre o qual [está sendo feita] uma declaração. Este objeto é mais comumente chamado de recurso. A descrição, então, é o que está sendo dito sobre o recurso” (POMERANTZ, 2015, p. 27, tradução nossa). Ainda segundo o mesmo autor, uma declaração possui três partes básicas, conforme a Figura 1



Figura 1. As três partes de uma declaração (POMERANTZ, 2015, p. 27, tradução nossa)

A primeira parte da declaração é o sujeito, ou seja, a entidade que está sendo descrita, representado na Figura 1 pela obra de arte Mona Lisa. A segunda parte é o relacionamento (predicado) que há entre o recurso informacional e o objeto, apresentado na Figura 1 como a relação de criador, ou seja, o criador da obra Mona Lisa. A terceira parte é o objeto que está vinculado ao sujeito por meio do predicado. “O objeto é uma entidade utilizada para descrever o sujeito”. (POMERANTZ, 2015, p. 28, tradução nossa). Na Figura 1, o objeto é Da Vinci, o criador da obra de arte Mona Lisa. Por ser composta por três partes, esta estrutura é denominada na literatura sobre metadados como *tripla* (sujeito-predicado-objeto).

Há diversos padrões de metadados, aplicados a finalidades específicas. O Dublin Core (DC) foi o primeiro padrão de metadados desenvolvido para descrição de recursos digitais na Web. Surgiu a partir de um congresso sediado na cidade de Dublin, Ohio, onde bibliotecários apresentaram propostas baseadas em registros elaborados usando Código de Catalogação Anglo-Americano (AACR2) como norma e o Machine Readable Cataloging (MARC21) como formato de padronização de dados (ZENG; QIN, 2008).

O Dublin Core Abstract Model (DCAM) é a base do padrão de metadados DC e foi desenvolvido pela DCMI. Esse modelo é composto por três submodelos: a) DCMI Resource Model; b) DCMI Description Set Model; c) The DCMI Vocabulary Model. De acordo com Powell et al. (2007, *online*) esse modelo permite “obter uma melhor compreensão dos tipos de descrições [...] [que estão sendo codificadas] e facilita o desenvolvimento de melhores mapeamentos e traduções de sintaxe cruzada.”

Padrões de metadados são criados para a representação de dados de áreas específicas. Tratando-se da biodiversidade, destaca-se entre os padrões de metadados o DwC, utilizado mundialmente para o compartilhamento de dados sobre a biodiversidade.

2.3. Padrão de metadados Darwin Core

O padrão de metadados DwC foi concebido para facilitar a “descoberta, recuperação e integração de informação sobre espécimes biológicos modernos, suas ocorrências espaço-temporais, e suas evidências de comprovação hospedadas em coleções (físicas ou digitais)”. (DARWIN CORE TASK GROUP, 2009c, *online*). Ele foi criado em 2009 pelo Darwin Core Task Group, subdivisão do Biodiversity Information Standards (TDWG¹), tendo por base os princípios do DC, portanto, segue os princípios do DCAM. Tem sido aceito e aplicado por mais de 340 instituições, em 43 países, incluindo o Brasil (GBIF, 2019). Em 2012, estimava-se que a GBIF indexasse mais de 300 milhões de registros no formato DwC (WIECZOREK et al., 2012). Hoje, o número de registros de ocorrência é de aproximadamente 1.302.735.625², sendo desse montante 13.834.599³ registros provenientes do Brasil (GBIF, 2019).

¹ O nome do grupo foi alterado de Taxonomic Databases Working Group para Biodiversity Information Standards, mas decidiram manter a sigla TDWG.

² Dado obtido a partir da busca simples no repositório da GBIF. Disponível em: <https://www.gbif.org/occurrence/charts>. Acesso em: 28 abr. 2019.

3. METODOLOGIA

Quanto à abordagem do problema, esta investigação caracteriza-se como qualitativa, devido à maneira intuitiva de analisar os dados, com o suporte de especialista do domínio (GIL, 1994). Do ponto de vista de seus objetivos, a pesquisa é exploratória e descritiva, uma vez que envolveu o estudo exaustivo das temáticas tratadas (GIL, 1994), e, quanto aos procedimentos, é uma pesquisa aplicada (GERHARDT; SILVEIRA, 2009).

O objeto estudado é o conjunto de conceitos da agrobiodiversidade e o núcleo de metadados do padrão DwC. Para refinar a representação dos conceitos de interações entre espécies, será utilizado como insumo o *Final Report of the Task Group on GBIF Data Fitness for Use in Agrobiodiversity* (ARNAUD et al., 2016), que apresenta os principais dados produzidos no contexto da agrobiodiversidade, apontando como representar alguns desses dados por meio de metadados já desenvolvidos pela GBIF. Segundo Arnaud et al. (2016), os principais e mais relevantes dados que podem ser reutilizados em sistemas de cultivo são sobre interações entre espécies (interações biológicas), aspectos agronômicos e práticas agronômicas.

Decidiu-se adotar como amostra desta pesquisa os conceitos da agrobiodiversidade que compõem o escopo das interações entre espécies, conceito não representado nos metadados do DwC e de suas extensões. Além da inexistência de metadados sobre interações entre espécies, a escolha da amostra se justifica pela necessidade de gestão desses dados específicos no contexto da Embrapa, instituição parceira no desenvolvimento desta pesquisa.

A metodologia foi dividida em três etapas:

- a) a primeira etapa consistiu em identificar as classes mais gerais do núcleo básico de termos do DwC, levantar dados sobre os termos existentes nas extensões do DwC disponíveis, assim como estudar a estrutura e os termos do modelo <onto.biodiversidade> desenvolvido na Embrapa. A escolha do modelo <onto.biodiversidade> se justifica pelo fato de ele ter sido desenvolvido como suporte a um projeto da Embrapa, instituição referência na pesquisa agropecuária brasileira, e por não haver outros modelos semelhantes na literatura. Entretanto, constatou-se após análise que esse modelo não possui conceitos relacionados às interações ecológicas que pudessem ser reaproveitados. Desta forma, o insumo conceitual para o campo das interações ecológicas está sendo pesquisado na literatura especializada;

³ Dado obtido a partir da busca simples no repositório da GBIF. Disponível em: <https://www.gbif.org/occurrence/charts?country=BR>. Acesso em: 28 abr. 2019.

- b) a segunda etapa desta pesquisa, que ainda não foi realizada, consistirá em extrair da literatura especializada os termos que representam as interações ecológicas e testar se esses atendem ao modelo de triplas (Figura 1), ou seja, se podem ligar um sujeito a um objeto; posteriormente, os termos selecionados serão definidos de acordo com as recomendações da ISO/IEC 11179-4 (2004).
- c) na terceira etapa, a extensão a ser criada seguirá o modelo dos metadados DwC (vide **Erro! Fonte de referência não encontrada.** e **Erro! Fonte de referência não encontrada.**) e os princípios para construção de metadados do DCMI Abstract Model simplificado (COYLE, 2008). Para isso, serão aplicadas as três formas de se fazer a extensão do padrão de metadados DwC: a) criando novos metadados, utilizando-se termos que não estão presentes no core do padrão de metadados; b) utilizando-se qualificadores para metadados existentes para que esses possam representar valores específicos; c) esquemas de codificação, como os Códigos Internacionais de Nomenclatura.

4. RESULTADOS PRELIMINARES

As relações ecológicas podem ocorrer entre indivíduos da mesma espécie (intraespecífica) ou entre indivíduos de espécies diferentes (interespecífica) (CASSINI, 2005). As relações ecológicas propriamente ditas são: colônia, sociedade, canibalismo, competição, comensalismo, inquilinismo, mutualismo, simbiose, foresia, amensalismo, competição, esclavagismo, parasitismo e predatismo.

Os percevejos das espécies *Euschistus heros*, *Piezodorus guildinii* e *Nezara viridula* são pragas que atacam a soja. De acordo Pacheco e Corrêa-Ferreira (2000), espécies parasitoides de ovos são utilizadas em diversos países para controlar as populações de percevejos que atacam a soja. No estudo conduzido pelos autores, que são pesquisadores da Embrapa, foi utilizada a espécie de microvespa *Telenomus podisi* Ashmead como parasita para controlar a população dos percevejos.

A partir da situação apresentada, infere-se que há uma relação ecológica do tipo parasitismo entre a microvespa e o percevejo da soja. Essa relação poderia ser representada em um modelo de triplas conforme a Figura 2.



Figura 2. Relação de parasitismo no modelo de triplas (elaborado pelos autores, 2019)

O predicado *parasiteOf* descreve a relação entre a espécie de vespa e o percevejo da Figura 2 **Erro! Fonte de referência não encontrada.** pode ser formalmente descrito conforme a ficha terminológica do Quadro 1.

Quadro 1. Elemento da classe Relação ecológica

parasiteOf		Property
Definition	Organismo que vive às custas de outro organismo, prejudicando-o.	
Comments	A melhor prática recomendada é inserir o nome científico completo da espécie hospedeira, com autoria e data se possível, de acordo com os Códigos Internacionais de Nomenclatura.	
Examples	<i>Euschistus heros</i> (Fabricius, 1798)	

Fonte: elaborado pelos autores, 2019.

Levando-se em consideração que o objeto do exemplo do Quadro 1, ou seja, a espécie *Euschistus heros* (Fabricius, 1798), também pode ser sujeito caso a tripla seja invertida, pode-se apresentar outro predicado, com o percevejo como sujeito, conforme o Quadro 2.

Quadro 2. Relação inversa

hostTo		Property
Definition	Organismo que serve de abrigo para outro organismo.	
Comments	A melhor prática recomendada é inserir o nome científico completo do parasito, com autoria e data se possível, de acordo com os Códigos Internacionais de Nomenclatura.	
Examples	<i>Telenomus podisi</i> (Ashmead, 1893)	

Fonte: elaborado pelos autores, 2019.

As triplas apresentadas nos exemplos dos Quadro 1 e Quadro 2 serão desenvolvidas para todos os termos da amostra, que resultarão na composição da extensão de metadados que criar-se-á como resultado desta pesquisa.

5. CONSIDERAÇÕES FINAIS

Espera-se ao final desta pesquisa criar uma extensão de metadados capaz de atender à representação de dados das interações ecológicas, que são fator determinante para o desenvolvimento da agricultura sustentável.

A metodologia para criação de metadados aplicada nesta pesquisa poderá servir como modelo para construção de metadados em outras áreas, uma vez que se baseia nos princípios fundamentais para construção de metadados, que são representados pelo DCAM.

REFERÊNCIAS

BEGON, M.; TOWNSEND, C. R.; HARPER, J. L. **Ecologia**. Tradução de Adriano Sanches Melo et al. 4. ed. Porto Alegre: Artmed, 2007.

CAIN, M. L.; BOWMAN, W. D.; HACKER, S. D. **Ecologia**. Revisão técnica de Fernando Joner e Paulo Luiz de Oliveira. 3. ed. Porto Alegre: Artmed, 2018. Traduzido por: Armando Molina Divan Junior et al.

CONVENTION ON BIOLOGICAL DIVERSITY. **COP 5 Decisions**: Agricultural biological diversity: Review of phase I of the programme of work and adoption of a multi-year work programme: Annex: Programme of work on agricultural biodiversity. Nairobi, 2000. Available from: <https://www.cbd.int/decisions/cop/?m=cop-05>. Accessed on: 9 April 2019.

COYLE, K. **DCAM Explained**. [S. l.], 2008. Disponível em: http://kcoyle.net/dcam_simple.html. Acesso em: 28 jul. 2019.

DARWIN CORE TASK GROUP. **Darwin Core**. [S. l.]: Biodiversity Information Standards, [2009]. Available from: <http://rs.tdwg.org/dwc/>. Accessed on: 8 Sept. 2018.

GLOBAL BIODIVERSITY INFORMATION FACILITY. **Padrões de dados**. Copenhagen: GBIF Secretariat Universitetsparken, [2019]. Disponível em: <https://www.gbif.org/pt/standards>. Acesso em: 22 fev. 2019.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 11179-4**: information technology: metadata registries (MDR): part 4: formulation of data definitions. 2nd ed. Geneva, 2004. 9 p.

MOYES, C. D.; SCHULTE, P. M. **Princípios de fisiologia animal**. Tradução de Adriane Belló Klein et al. 2. ed. Porto Alegre: Artmed, 2010.

ODUM, E. P.; BARRET, G. W. **Fundamentos de ecologia**. Tradução da 5. ed. norte americana por Pégasus Sistemas e Soluções. São Paulo: Cengage Learning, 2006. 632 p. Título original: Fundamentals of ecology.

PACHECO, D. J. P.; CORRÊA-FERREIRA, B. S. Parasitismo de *Telenomus podisi* Ashmead (Hymenoptera: Scelionidae) em Populações de Percevejos Pragas da Soja. **Anais da Sociedade Entomológica do Brasil**, [s. l.], v. 29, n. 2, p. 295-302, jun. 2000. Disponível em: <http://www.scielo.br/pdf/aseb/v29n2/v29n2a11.pdf>. Acesso em: 29 abr. 2019.

POMERANTZ, J. **Metadata**. Cambridge, MA: The MIT Press, 2015.

RILEY, J. **Understanding Metadata**: What is Metadata, and What is it For?: A Primer. Baltimore: National Information Standards Organization, 2017. 45 p. Disponível em: https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 16 fev. 2019.

SISTEMA DE INFORMAÇÃO SOBRE A BIODIVERSIDADE BRASILEIRA. **Termos Darwin Core**. Brasília, DF: Ministério da Ciência, Tecnologia, Inovações e Comunicações, [2018?].

Disponível em: <http://www.sibbr.gov.br/areas/index.php?area=publicar&subarea=termos-dwc>. Acesso em: 3 junho 2018.

SOARES, F. M.; MACULAN, B. C. M. dos S.; DRUCKER, D. P. Darwin Core for Agricultural Biodiversity: a metadata extension proposal. **Biodiversity Information Science and Standards**, v. 3, p. e37053, 2019. DOI: <https://doi.org/10.3897/biss.3.37053>

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-Scguman Publishers, 2008.

Filipi Miranda Soares

Mestrando em Gestão e Organização do Conhecimento PPGOC-UFMG, bacharel em Biblioteconomia pela ECI-UFMG, intercâmbio acadêmico no curso Ciências e Tecnologias da Documentação e Informação pelo Instituto Superior de Contabilidade e Administração do Porto, Portugal, ISCAP (2016-2017).

Benildes Coura Moreira dos Santos Maculan

Doutora e Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação, na Escola de Ciência da Informação (ECI) da Universidade Federal de Minas Gerais (UFMG). Formada em Biblioteconomia pela ECI/UFMG. Atualmente, Professora Adjunto no Departamento de Organização e Tratamento da Informação, na ECI/UFMG.

Debora Pignatari Drucker

Possui graduação em Engenharia Florestal pela ESALQ/Universidade de São Paulo (2001), mestrado em Biologia (Ecologia) pelo Instituto Nacional de Pesquisas da Amazônia/INPA (2005) e doutorado em Ambiente e Sociedade pelo NEPAM - Universidade Estadual de Campinas/UNICAMP (2012). Tem experiência nas áreas de Engenharia Florestal, Ecologia, Uso sustentável de Recursos Naturais, Gestão de Dados Científicos e Inovação Tecnológica. É analista da Embrapa Informática Agropecuária e atua na área de transferência de tecnologia.