

## Orientações para Catalogação de *Datasets* Ômicos no Repositório de Dados de Pesquisa da Embrapa



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Agricultura Digital  
Ministério da Agricultura, Pecuária e Abastecimento*

## DOCUMENTOS 179

# Orientações para Catalogação de *Datasets* Ômicos no Repositório de Dados de Pesquisa da Embrapa

*Márcia Izabel Fugisawa Souza  
Marcos Cezar Visoli  
Carla Cristiane Osawa  
Tércia Zavaglia Torres  
Alessandra Rodrigues da Silva*

**Embrapa Agricultura Digital**

Av. Dr. André Tosello, 209 - Cidade Universitária  
Campinas, SP, Brasil  
CEP. 13083-886  
Fone: (19) 3211-5700  
www.embrapa.br  
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações  
Presidente  
*Stanley Robson de Medeiros Oliveira*

Secretária-Executiva  
*Maria Fernanda Moura*

Membros  
*Adriana Farah Gonzalez, membro nato, Alexandre de Castro, membro indicado, Carla Cristiane Osawa, membro nato, Debora Pignatari Drucker, membro eleito, Ivan Mazoni, membro eleito, João Camargo Neto, membro indicado, Joao Francisco Goncalves Antunes, membro eleito, Magda Cruciol, membro nato.*

Supervisão editorial  
*Stanley Robson de Medeiros Oliveira*

Revisão de texto  
*Adriana Farah Gonzalez*

Normalização bibliográfica  
*Carla Cristiane Osawa*

Projeto gráfico da coleção  
*Carlos Eduardo Felice Barbeiro*

Editoração eletrônica  
*Magda Cruciol*

Imagem de capa  
<https://mon.uvic.cat/master-omics/>  
<https://dglab.com.br/blog/mapeamento-genetico-qual-o-seu-impacto-nos-dias-de-hoje>

**1ª edição**  
Publicação digital - PDF (2021)

**Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)**  
Embrapa Agricultura Digital

---

Orientações para catalogação de datasets ômicos no Repositório de Dados de Pesquisa da Embrapa / Márcia Izabel Fugisawa Souza ... [et al.]. – Campinas : Embrapa Agricultura Digital, 2021

PDF (113 p.) : il. color. - (Documentos / Embrapa Agricultura Digital, ISSN 2764-2488 ; 179).

1. Representação descritiva. 2. Representação temática. 3. Catalogação. 4. *Datasets* ômicos. 5. Redape. I. Souza, Márcia Izabel Fugisawa. II. Embrapa Agricultura Digital. III. Série.

CDD (21. ed.) 025.32

## Autores

### **Márcia Izabel Fugisawa Souza**

Bacharel em Biblioteconomia, doutora em Educação, analista da Embrapa Agricultura Digital, Campinas, SP.

### **Marcos Cezar Visoli**

Cientista da Computação, mestre em Informatique e Systèmes Spécialité Recherche Modèles, Systèmes, Imagerie, Robotique, pesquisador da Embrapa Agricultura Digital, Campinas, SP.

### **Carla Cristiane Osawa**

Bacharel em Química e Química Tecnológica e em Biblioteconomia, mestre em Química, analista da Embrapa Agricultura Digital, Campinas, SP.

### **Tércia Zavaglia Torres**

Administradora, doutora em Educação, analista da Embrapa Agricultura Digital, Campinas, SP.

### **Alessandra Rodrigues da Silva**

Bacharel em Biblioteconomia, doutora em Ciência da Informação, analista da Embrapa Sede, Brasília, DF.

## Colaboradores

### **Antonio Nhani Junior**

Biólogo, doutor em Bioquímica, pesquisador da Embrapa Agricultura Digital, Campinas, SP.

### **Poliana Fernanda Giachetto**

Zootecnista, doutora em Produção Animal, pesquisadora Embrapa Agricultura Digital, Campinas, SP.

### **Paula Regina Kuser Falcão**

Física, doutora em Cristalografia de Proteínas, pesquisadora da Embrapa Agricultura Digital, Campinas, SP.

### **Leandro Cintra Carrijo**

Cientista da Computação, doutor em Bioinformática, analista da Embrapa Agricultura Digital, Campinas, SP.

### **Luiz Manoel Silva Cunha**

Estatístico, mestre em Ciência da Computação e Matemática Computacional, analista da Embrapa Agricultura Digital, Campinas, SP.

### **Luiz Antonio Falaguasta Barbosa**

Cientista da Computação, mestre em Ciência da Computação, analista da Embrapa Agricultura Digital, Campinas, SP.

## Lista de figuras – Apêndice A

Figura 1. Tela de busca do AGROVOC *Thesaurus*.

Figura 2. Tela de resultado de busca do AGROVOC *Thesaurus* para o termo 'nucleotide sequence'.

Figura 3. Tela do termo do 'nucleotide sequence' contendo informações detalhadas.

Figura 4. Resultado da busca do termo 'nucleotide sequence', no AGROVOC *Thesaurus*, com destaque para o item URI.

Figura 5. Tela de busca do AGROVOC *Thesaurus*.

Figura 6. Tela de resultado de busca do AGROVOC *Thesaurus* para o termo 'nucleotide sequence'.

Figura 7. Tela com os detalhes do termo 'phenomena'.

Figura 8. Resultado da busca do termo 'phenomena', no AGROVOC *Thesaurus*, com destaque para o item URI.

Figura 9. Tela de busca do NAL *Thesaurus*.

Figura 10. Tela de busca do NAL *Thesaurus* com o termo 'mites'.

Figura 11. Resultado da busca do termo no NAL *Thesaurus*.

Figura 12. Tela do termo 'mites' contendo informações sobre classificações, relações, associações, formato, identificador persistente, tradução do termo para o espanhol e número do termo.

Figura 13. Tela de buscado NAL *Thesaurus*.

Figura 14. Tela de busca do NAL *Thesaurus* para o termo 'mites'.

Figura 15. Resultado da busca do termo 'mites' no NAL *Thesaurus*.

Figura 16. Tela do termo 'mites' contendo informações sobre classificações, relações, associações, formato, identificador persistente, tradução do termo para o espanhol e número do termo.

Figura 17. Tela da categoria 'L Animal Sciences and Animal Products' do NAL *Thesaurus*.

Figura 18. Tela de busca do NCBI *Taxonomy*.

Figura 19. Tela de resultados da busca do termo *Rhipicephalus (Boophilus) microplus* no NCBI *Taxonomy*.

Figura 20. Tela de busca do NCBI *Taxonomy*.

Figura 21. Tela de resultados da busca do termo *Rhipicephalus (Boophilus) microplus* no NCBI *Taxonomy*.

Figura 22. Hierarquia do 'Lineage (full)' do termo *Rhipicephalus microplus*.

Figura 23. Resultado da escolha do termo 'Ixodida'.

## Lista de tabelas

Tabela 1. Princípios e elementos de dados FAIR.

Tabela 2. Regras de preenchimento do formulário de catalogação no Redape.

Tabela 3. Exigências de preenchimento dos elementos no esquema Metadados de Citação (campos e subcampos).

Tabela 4. Elementos metadados e níveis de exigência de preenchimento.

Tabela 5. Elementos do esquema Metadados de Citação do formulário de catalogação do Redape.

Tabela 6. Atributos orientadores para a descrição dos elementos do esquema Metadados de Citação (campos e subcampos) no Redape.

Tabela 7. Descrição dos elementos do esquema Metadados Citação (campos e subcampos) no Redape.

Tabela 8. Elementos do esquema Metadados Ciência da Vida do formulário de catalogação do Redape.

Tabela 9. Atributos orientadores para a descrição dos elementos do esquema Metadados Ciência da Vida do Redape.

Tabela 10. Descrição dos elementos do esquema Metadados Ciência da Vida no Redape.

## Lista de tabelas – Apêndice B

Tabela 1. Exigências de preenchimento dos elementos do esquema Metadados de Citação (campos e subcampos).

Tabela 2. Descrição dos elementos do esquema Metadados de Citação (campos e subcampos) no Redape.

Tabela 3. Descrição dos elementos do esquema Metadados Ciência da Vida (campo e subcampos) no Redape.

## Lista de tabelas – Apêndice C

Tabela 1. Matriz de atividades de revisão na catalogação de *datasets* ômicos no Redape – Esquema Metadados de Citação.

Tabela 2. Matriz de atividades de revisão na catalogação de *datasets* ômicos no Redape – Esquema Metadados Ciência da Vida.

## Lista de tabelas – Apêndice D

Tabela 1. Orientações para a revisão na catalogação de *datasets* ômicos no Redape – Esquema Metadados de Citação.

Tabela 2. Orientações para a revisão na catalogação de *datasets* ômicos no Redape – Esquema Metadados Ciência da Vida.

## Lista de siglas

AACR – *Anglo-American Cataloguing Rules*

ABNT – Associação Brasileira de Normas Técnicas

Alice – Repositório Acesso Livre à Informação Científica da Embrapa

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

DOI – *Digital Object Identifier*

Embrapa – Empresa Brasileira de Pesquisa Agropecuária

ESALQ – Escola Superior de Agricultura Luiz de Queiroz

FAIR – *Findable, Accessible, Interoperable, Reusable*

FAO – *Food and Agriculture Organization of the United Nations*

Fapesp – Fundação de Amparo à Pesquisa do Estado de São Paulo

Infoteca-e – Repositório de Informação Tecnológica da Embrapa

INSDC – *International Nucleotide Sequence Database Collaboration*

ISO – *International Organization for Standardization*

LMB – Laboratório Multiusuário de Bioinformática da Embrapa

NAL – *National Agricultural Library*

NBR – Norma brasileira

NCBI – *National Center for Biotechnology Information*

ORCID – *Open Researcher and Contributor ID*

PD&I – Pesquisa, desenvolvimento e inovação

PGDIC – Política de Governança de Dados, Informação e Conhecimento da Embrapa

Redape – Repositório de Dados de Pesquisa da Embrapa

RmVAC – Vacina contra *Rhipicephalus microplus*

URI – *Uniform Resource Identifier*

URL – *Uniform Resource Locator*

WWW – *The World Wide Web*

## Sumário

|   |     |
|---|-----|
| Introdução.....   | 11  |
| Metadados e Descrição de Recursos.....  | 14  |
| Metadados e o software Dataverse.....   | 15  |
| Catalogação de dados .....  | 16  |
| Catalogação de Dados Ômicos: a Experiência do Laboratório Multiusuário de Bioinformática da Embrapa (LMB).....      | 16  |
| Regras Gerais para Descrição de <i>Datasets</i> Ômicos no Repositório de Dados de Pesquisa da Embrapa (Redape)..... | 18  |
| Elementos metadados requeridos, fortemente recomendados, recomendados e opcionais.....                              | 18  |
| Outras orientações para a catalogação de dados ômicos no Redape.....  | 21  |
| Nomes pessoais e nomes corporativos, e-mails pessoais e e-mails corporativos e afiliação .....                      | 22  |
| Descrição dos Elementos no Esquema Metadados de Citação .....   | 23  |
| Descrição dos Elementos do Esquema Metadados Ciência da Vida.....   | 43  |
| Referências .....   | 47  |
| Apendice A - Vocabulários Controlados.....  | 50  |
| Apendice B - Guia Resumido das Orientações para Catalogação de <i>Datasets</i> Ômicos no Redape.....                | 80  |
| Apendice C - Revisão na Catalogação de <i>Datasets</i> Ômicos no Redape.....  | 91  |
| Apendice D - Modelo de Relatório de Revisão na Catalogação de <i>Datasets</i> Ômicos.....                           | 104 |

## Apresentação

As atividades científicas têm passado por profundas mudanças na sua forma de produzir conhecimento. O modo de fazer ciência vem, ao longo dos tempos, se modificando. Os avanços constantes na produção de conhecimento científico têm propulsionado o desenvolvimento de novas tecnologias, produtos e serviços, voltados ao atendimento de amplos interesses da sociedade.

A passagem de um paradigma da ciência baseada em hipótese para um paradigma de ciência orientada a dados traz implicações substanciais no fazer científico em instituições de Pesquisa, Desenvolvimento e Inovação (PD&I), como a Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Atenta a essas transformações, a Embrapa tem priorizado ações voltadas à gestão de dados, informação e conhecimento, a partir do estabelecimento de mecanismos para a geração, organização, tratamento, acesso, preservação, recuperação, divulgação, compartilhamento e reúso dos seus ativos de informação.

Nesse contexto, esta publicação traz uma contribuição efetiva para comunidade de dados ômicos da Embrapa apresentando orientações de cunho prático para que pesquisadores e analistas cataloguem seus conjuntos de dados (*datasets* ômicos) no Repositório de Dados de Pesquisa da Embrapa (Redape), visto que a catalogação é uma atividade essencial para a interoperabilidade, o compartilhamento, o uso e o reúso dos dados de pesquisa.

As orientações para catalogação de *datasets* ômicos foram elaboradas no âmbito de uma ação gerencial de melhoria de processos de gestão de dados, conduzida no Laboratório Multiusuário de Bioinformática da Embrapa (LMB). Trata-se, assim, de uma contribuição tanto para implementar a Política de Governança de Dados, Informação e Conhecimento (PGDIC), quanto para instrumentalizar o processo de catalogação de *datasets* no Redape.

*Silvia Maria Fonseca Silveira Massruhá*

Chefe-geral

## Introdução

O presente documento reúne regras básicas para a catalogação de *datasets*<sup>1</sup> com o objetivo de orientar o trabalho do catalogador e contribuir para a qualidade dos conjuntos de dados a serem disponibilizados no Repositório de Dados de Pesquisa da Embrapa (Redape). Este repositório institucional e digital visa a:

[...] preservar e facilitar a busca por dados de pesquisa produzidos pela Empresa. Ele permite a organização, o gerenciamento e a publicação de dados de acordo com os princípios que norteiam a gestão de dados científicos em todo o mundo, incluindo a acessibilidade, a interoperabilidade, a reprodutibilidade e o reúso. (Embrapa, 2021).

As regras de catalogação de *datasets* foram definidas no decorrer de uma ação gerencial conduzida em 2020, no âmbito do “Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB)”, no escopo do projeto de pesquisa “Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus microplus* e espécies relacionadas”.

Coincidentemente, nesse período, estavam em andamento na Embrapa as ações de implantação do repositório institucional para armazenar dados de pesquisa, que culminaram com adoção do software Dataverse. Essa conjunção de fatores trouxe à tona a catalogação de *datasets* como uma atividade prioritária, e, por conseguinte, dependente do estabelecimento de orientações para a descrição dos elementos metadados, as quais são objeto desta publicação.

Todavia, pode-se perguntar ‘por que catalogar *datasets*? Uma resposta possível para esta pergunta pode ser construída a partir da análise do paradigma da *e-Science* e da concepção de Ciência Aberta e Colaborativa, esta entendida como um empreendimento multi e interdisciplinar da ciência contemporânea (Chan et al., 2015). Compreende-se, pois, essa ciência como aquela que tem “capacidade de gerar e armazenar dados em escala sem precedentes muito além da capacidade humana de análise” (Cesar Junior, 2011, p. 7-8).

Atenta a esse paradigma, a Embrapa tem se posicionado ao lado das instituições líderes mundiais, no tocante a adoção de políticas de governança de dados de pesquisa. Atualmente, a Empresa tem priorizado a ação estratégica de implantação de um repositório de dados de pesquisa para abrigar o depósito, a preservação e a publicação de dados de pesquisa, com vista a promover o reúso dos dados por ela produzidos. Portanto, trata-se de uma iniciativa que se soma às experiências já consolidadas de criação de repositórios na Empresa, como é o caso do Repositório Acesso Livre à Informação Científica da Embrapa (Alice)<sup>2</sup> do Repositório de Informação Tecnológica da Embrapa (Infoteca-e)<sup>3</sup> e do GeoInfo - Infraestrutura de Dados Espaciais da Embrapa<sup>4</sup>.

---

<sup>1</sup> Optou-se pelo uso do termo ‘*dataset*’ (no idioma inglês) por considerá-lo apropriado para se referir a coleções de dados estruturados, enquanto conjuntos de dados (a forma traduzida) abarca tanto dados estruturados quanto não estruturados. “Um dataset no Dataverse é um contêiner dos arquivos de dados, da documentação, dos códigos e dos metadados descritivos desse dataset.” (Dataverse..., 2016, tradução nossa). “Como Dataverse foi desenvolvido para repositório de dados, a representação e a gestão automatizada dos conjuntos de dados é estruturada através do conceito dataset, que inclui dados, metadados de citação, metadados específicos, documentação adicional, citação, gerenciamento de versões, etc.” (Rocha et al., 2018, p. 64). Ainda no software Dataverse, “Um dataset é composto por metadados, pelos termos de uso (como licenças) e por arquivos.” (Rocha et al., 2021, p. 9).

<sup>2</sup> Disponível em: <<http://www.alice.cnptia.embrapa.br>>.

<sup>3</sup> Disponível em: <<http://infoteca.cnptia.embrapa.br>>.

<sup>4</sup> Disponível em: <<http://www.embrapa.br/geoinfo>>.

Dados de pesquisa produzidos pela Embrapa são ativos digitais de valor intangível, com potencial para beneficiar diretamente inúmeros pesquisadores e instituições por meio do reuso, seja pelo reaproveitamento, agregação, integração, metanálise ou reanálise (Curty, 2019). No entanto, a implantação de repositórios de dados traz consigo a necessidade de que os dados sejam descritos e preservados adequadamente, de acordo com requisitos de qualidade para possibilitar seu com-partilhamento e reuso eficiente, por exemplo: a recuperação de um *dataset* disponibilizados em um repositório por outro(s) pesquisador(es) e sua utilização em outros projetos<sup>5</sup>, atendendo aos princípios FAIR<sup>6</sup> e às orientações da Biblioteconomia<sup>7</sup>, Ciência da Informação<sup>8</sup> e da catalogação de dados (Curty, 2019; Semeler; Pinto, 2019).

Nesse contexto se insere a necessidade de implementação de melhorias de natureza tecnológica e processual nos repositórios, especialmente no que tange ao tratamento descritivo dos *datasets*, abarcado pela atividade de catalogação, que visa conferir qualidade tanto aos metadados quanto aos dados em si.

Ademais, o processo de catalogação impacta diretamente outros processos dele decorrentes, como a indexação, a busca, a descoberta e a recuperação de dados e informações, para os quais os vocabulários controlados<sup>9</sup> tornam-se imprescindíveis e de uso obrigatório.

O presente documento é composto de quatro capítulos e quatro apêndices. O primeiro capítulo introduz o leitor aos metadados e à descrição de recursos, abordando os princípios FAIR, o software Dataverse, a catalogação de dados e de *datasets* ômicos no LMB.

O segundo capítulo traz as regras gerais que devem ser observadas durante a etapa de descrição de *datasets* no Redape, no tocante ao preenchimento do formulário de catalogação, aos elementos metadados requeridos, fortemente recomendados, recomendados e opcionais. Contém ainda orientações para o uso de pontuação, maiúsculas, tecla *Enter*, botões de acesso, ajuda rápida, nomes e e-mails pessoais e corporativos, afiliação, palavras-chave, termos-livre e data.

O terceiro capítulo é constituído das orientações para a descrição dos elementos do esquema Metadados de Citação, predefinidos pelo software Dataverse, os quais compõem um conjunto genérico de 33 elementos, entre campos e subcampos. Neste capítulo são apresentados: os elementos metadados e o tipo de exigência de preenchimento para cada um deles, bem como os atributos que orientam a descrição desses elementos. É neste capítulo que o catalogador encontra as orientações para a descrição de cada elemento e respectivos campos e subcampos, compostas de: a)

---

<sup>5</sup> Os dados de pesquisa podem ser reutilizados integralmente, parcialmente ou de forma combinada, uma ou várias vezes, para propósitos iguais, semelhantes ou completamente diferentes daqueles em que foram coletados. O reuso envolve todo e qualquer novo uso de dados pré-existent, podendo ser feito pelos pesquisadores que coletaram/produziram os dados primários ou por outras partes não necessariamente envolvidas na pesquisa original. (Curty, 2019).

<sup>6</sup> FAIR é o acrônimo de *Findable, Accessible, Interoperable, Reusable*, princípios que preconizam que os dados necessitam ser encontráveis, acessíveis, interoperáveis e reutilizáveis (Go FAIR, 2018).

<sup>7</sup> Área do conhecimento que se ocupa com a organização e a administração das bibliotecas e outras unidades de informação, além da seleção, aquisição, organização e disseminação de publicações sob diferentes suportes físicos (Targino, 2006 citado por Russo, 2010, p. 47).

<sup>8</sup> A Ciência da Informação engloba o conjunto de disciplinas voltadas à produção, comunicação e consumo da informação (Ortega, 2004). Ela recebe a colaboração de várias disciplinas, que leva a interações e a uma certa reciprocidade para haver enriquecimento mútuo, dentre elas, destacam-se: Psicologia, Sociologia, Informática, Matemática, Lógica, Estatística, Economia, Comunicação, Direito, Política e Telecomunicações (Le Coadic, 1994).

<sup>9</sup> Um vocabulário controlado é essencialmente uma lista de termos autorizados, que geralmente constam em uma lista adotada por uma instituição. Inclui, em geral, uma forma de estrutura semântica destinada, especialmente, a: a) controlar sinônimos, optando por uma única forma padronizada, com remissiva de todas as outras; b) diferenciar homógrafos; e c) reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si. (Lancaster, 2004, p. 19).

título do elemento, campo e subcampo; b ) identificador; c) ocorrência; d) exigência de preenchimento; e) definição; f) comentário; g) exemplo e h) documento orientador.

No quarto capítulo são apresentadas as regras para a descrição dos elementos do esquema Metadados Ciência da Vida, também oferecidos pelo Dataverse; neste caso, constituídos por elementos específicos voltados para a descrição de dados biológicos. Neste capítulo são apresentados os oito elementos do esquema, todos de preenchimento opcional, além dos atributos que norteiam a descrição. As orientações para a descrição de cada elemento são as mesmas estabelecidas para o esquema Metadados de Citação, mencionadas acima.

Ao final encontram-se quatro apêndices preparados para apoiar e complementar a atividade de catalogação, no intuito de ilustrar e facilitar o trabalho do catalogador. O Apêndice A orienta o uso de vocabulários controlados na atribuição de palavras-chave e de categorias de assunto para os *datasets*. Neste apêndice encontram-se as orientações para o uso dos vocabulários controlados AGROVOC *Thesaurus*, NAL *Thesaurus* e NCBI *Taxonomy*.

No Apêndice B o catalogador encontra uma versão resumida do documento completo das orientações para a catalogação de *datasets* ômicos, objeto principal desta publicação. Este apêndice foi preparado para atender a eventuais necessidades do catalogador, interessado em encontrar orientações que se apresentem de forma objetiva e sucinta, com menos detalhamento.

O Apêndice C apresenta as orientações para a execução da atividade de revisão na catalogação de *datasets* no Redape, salientando a necessidade de controle e garantia de qualidade, por meio da descrição padronizada de metadados. Neste apêndice estão estabelecidos os elementos e as etapas de controle de qualidade para a revisão de *datasets* ômicos, bem como apresentadas as matrizes de atividade para orientar a revisão.

O Apêndice D traz um modelo de relatório de revisão na catalogação, que contempla os seguintes itens: a) elemento (campo/subcampo); b) descrição do erro e/ou inconsistência; c) orientação para a correção; d) elemento(s) de impacto na qualidade; e e) documento orientador para a correção.

## Metadados e Descrição de Recursos

Este capítulo reúne as orientações para a catalogação de *datasets* no Redape, as quais se resumem nas regras gerais de descrição de *datasets* aplicáveis às especificidades do software Dataverse. Em especial, o documento traz orientações para os usuários (autores de *datasets*, produtores de *datasets*, especialistas de domínio, especialistas no tratamento da informação, curadores etc.) na realização da atividade de representação descritiva e de representação temática de *datasets*.

As orientações foram elaboradas com base em recomendações técnicas, nas regras básicas do Código de Catalogação Anglo-Americano – AACR2 (Ribeiro, 1995; 2003; Código..., 2004), e no uso de vocabulários controlados para tratamento de assunto (palavras-chave e classificação taxonômica), visando à padronização de nomes pessoais e corporativos, e-mails e afiliação etc.

Os metadados e a catalogação de dados no escopo deste documento foram considerados na perspectiva da gestão de dados de pesquisa e do ciclo de vida dos dados. Assim, metadados e catalogação de *datasets* correspondem à etapa Descrição, do ciclo de vida dos dados e, como tais devem estar alinhadas aos princípios FAIR (acrônimo formado pelas palavras em inglês: *Findable, Accessible, Interoperable, Reusable* (Go FAIR, 2018)). Esses princípios orientam a publicação de dados de pesquisa de modo a torná-los encontráveis, acessíveis, interoperáveis e reutilizáveis. Os princípios FAIR definem as características que recursos de dados contemporâneos, ferramentas, vocabulários e infraestruturas devem exibir para auxiliar a descoberta e o reuso por terceiros. (Wilkinson et al., 2016). Os quatro princípios FAIR são compostos de 15 elementos a eles associados e que contribuem para complementar e enriquecer as informações essenciais, tanto para metadados como para dados. Estão relacionados entre si, mas são independentes e separáveis (Wilkinson et al., 2016), conforme apresentado na Tabela 1, a seguir.

**Tabela 1.** Princípios e elementos de dados FAIR.

| <b>Princípio F - <i>Findable</i> (Encontrabilidade) e seus elementos</b>        |   |
|---|---|
| F1  | Os metadados e dados devem ter identificadores globais, persistentes e identificáveis.  |
| F2  | Os dados devem ser descritos com metadados enriquecidos (impacta diretamente R1).   |
| F3  | Os metadados devem incluir claramente e explicitamente os identificadores dos dados que descrevem.  |
| F4  | Os metadados e dados devem ser recuperáveis ou indexados em recursos que ofereçam capacidade de busca.  |
| <b>Princípio A - <i>Accessible</i> (Acessibilidade) e seus elementos</b>        |   |
| A1  | Metadados e dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado.  |
| A1.1  | O protocolo deve ser aberto, gratuito e universalmente implementável.   |
| A1.2  | O protocolo deve permitir procedimentos de autenticação e autorização, quando necessário.   |
| A2  | Metadados devem ser acessíveis, mesmo quando os dados não estão mais disponíveis.   |
| <b>Princípio I - <i>Interoperable</i> (Interoperabilidade) e seus elementos</b> |   |
| I1  | Metadados e dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento. |
| I2  | Metadados e dados devem usar vocabulários que assegurem os princípios FAIR.   |
| I3  | Metadados e dados devem incluir referências qualificadas para outros metadados e dados.   |
| <b>Princípio R - <i>Reusable</i> (Reusabilidade) e seus elementos</b>           |   |
| R1  | Metadados e dados são descritos com uma pluralidade de atributos precisos e relevantes.   |
| R1.1  | Metadados e dados devem ser disponibilizados com licenças de uso claras e acessíveis.   |
| R1.1  | Metadados e dados devem estar associados à sua proveniência.  |
| R1.3  | Metadados e dados devem estar alinhados com padrões relevantes ao seu domínio.  |

O detalhamento desses princípios não é objeto deste documento orientador de catalogação. Contudo, é oportuno enfatizar que os metadados e a catalogação são imprescindíveis para que dados de pesquisa possam ser “[...] identificáveis, citáveis, visíveis, recuperáveis, interpretáveis, contextualizáveis, interoperáveis e reutilizáveis onde quesitos de consistência e procedência são considerados.” (Semeler; Pinto, 2019, p. 116).

Metadados possibilitam a exploração de outras dimensões e facetas do dado, que ao serem reveladas pela catalogação passam a contribuir para a melhoria da gestão e qualidade dos dados de pesquisa, favorecendo a descoberta das coleções de dados para a comunidade científica. Metadados são indispensáveis para que, no futuro, os conteúdos digitais possam ser acessados e interpretados. Sem metadados, de acordo com Gray (2009) citado por Sayão e Sales (2016, slide 83), os usuários

[...] não saberão os detalhes de como os dados foram obtidos e preparados: 1) como os instrumentos foram projetados e construídos; 2) quando, onde e como os dados foram coletados; e, 3) não terão uma descrição dos processos que levaram aos dados derivados, que são tipicamente usados para análises científicas.

Metadados também são indispensáveis à interoperabilidade técnica e semântica, ou seja, sem eles os repositórios e plataformas de dados não poderão intercambiar dados e informações, tornando impossível a descoberta e o reuso dos dados. Metadados são constituídos por elementos descritivos bem definidos, como por exemplo: a) autor; b) título; c) descrição; d) assunto; e) palavra-chave; f) identificador; g) produtor; h) tipo de dados; i) restrições de acesso; j) termo de uso das coleções etc., formando a partir da catalogação de dados um corpo de informações capaz de contextualizar os dados, sobretudo, quanto à proveniência, história, natureza e propósito.

A geração de metadados enriquecidos<sup>10</sup> traz benefícios diretos para a gestão de dados, impactando positivamente no arquivamento e preservação, bem como na interoperabilidade e na recuperação de *datasets* de pesquisa. Dados somente serão úteis para análise se tiverem sido descritos com metadados de qualidade, e para que isso aconteça, a melhor recomendação é a adoção dos princípios FAIR para orientar a catalogação.

O conteúdo deste documento orientador deverá ser submetido à constante avaliação e adequação, procurando atender às necessidades dos usuários do Laboratório Multiusuário de Bioinformática da Embrapa e ao aprimoramento das recomendações e orientações, ajustando-as a possíveis e necessárias customizações do software Dataverse com vistas a melhorar a experiência de uso do Redape.

## Metadados e o software Dataverse

O software Dataverse é uma aplicação web, de código-aberto, dedicada ao compartilhamento, à preservação, à citação, à exploração e à análise de dados de pesquisa. Este software é utilizado na criação de repositórios institucionais de dados de pesquisa, os quais se estruturam por meio de metadados descritivos, documentação, códigos e arquivos de dados. (Dataverse, 2020). Após estudos e análises sobre ferramentas para gerir repositórios de dados, o software Dataverse foi escolhido pela Empresa para a implantação do Repositório de Dados de Pesquisa da Embrapa (Redape). Esses estudos levaram em consideração as experiências internacionais, as características tecnoló-

<sup>10</sup> Metadados enriquecidos semanticamente que descrevem os conjuntos de dados, de forma abrangente, incluindo aspectos descritivos, estruturais e de qualidade, para que sejam facilmente compreensíveis por humanos e máquinas. (Brasileiro, 2018).

gicas, os recursos e as iniciativas existentes no desenvolvimento e implantação de tecnologias voltadas para a gestão de dados.

## Catálogo de dados

A catalogação ou representação bibliográfica é uma disciplina da Biblioteconomia e da Ciência da Informação que consiste em um conjunto de informações que simbolizam um registro do conhecimento, podendo ser definida como:

O estudo, preparação e organização de mensagens, com base em registros do conhecimento, reais ou ciberespaciais, existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir a interseção entre as mensagens contidas nestes registros do conhecimento e as mensagens internas dos usuários. (Mey; Silveira, 2009, p. 7).

A catalogação requer o emprego de formato e padrões internacionais (de metadados<sup>11</sup>, inclusive) e de normas técnicas para executar a representação descritiva, bem como a adoção de taxonomias, tesouros e vocabulários controlados para formular a representação temática. Isso se aplica a quaisquer artefatos, objetos e unidades de informação, sejam impressos ou digitais. Portanto, a catalogação vai além da simples inserção de objetos de dados e de informação em base de dados, repositórios ou plataforma de dados.

No contexto da Ciência Aberta e da Biblioteconomia de Dados, a catalogação de *datasets* deve ser implementada em conformidade com os princípios FAIR, para que metadados e dados sejam descritos com padrão de qualidade, visando o reúso desses dados para a pesquisa. Ademais, a catalogação de dados beneficia a transparência da pesquisa, o que favorece a descoberta de dados e a produção de novos conhecimentos. Isso cria condições para a reprodutibilidade e verificação da pesquisa publicada. (Gordana; Dragan, 2017). Na mesma direção, Chan et al. (2015) ao discorrem sobre o paradigma da ciência contemporânea (*e-Science*), baseada em grandes volumes de dados, também discorrem sobre a reprodutibilidade e a verificação da pesquisa.

## Catálogo de Dados Ômicos<sup>12</sup>: a Experiência do Laboratório Multiusuário de Bioinformática da Embrapa (LMB)

O LMB é uma estrutura computacional de vanguarda criada em 2001 e sediada nas dependências da Embrapa Informática Agropecuária, atual Embrapa Agricultura Digital, localizada em Campinas, SP. Sua atuação faz frente aos avanços do conhecimento nas áreas de recursos genéticos, da biotecnologia e do melhoramento genético. Com a missão de viabilizar soluções de Bioinformática para projetos de pesquisa, desenvolvimento e inovação na Embrapa em um ambiente colaborativo, o LMB compartilha equipamentos de alto custo, otimiza recursos econômicos e humanos, incorpora e disponibiliza competências na área de computação de alto desempenho para a Embrapa e para qualquer outra instituição de ensino, pesquisa, desenvolvimento e inovação

<sup>11</sup> Podem ser considerados como dados sobre outros dados. É o termo que os bibliotecários colocaram em catálogos e que se refere comumente à informação descritiva sobre recursos da web. Um registro de metadados consiste em um conjunto de atributos, ou elementos, necessários para descrever o recurso em questão. (Mey; Silveira, 2009, p. 133).

<sup>12</sup> Dados gerados pelas tecnologias de sequenciamento de alto rendimento, que permitem extrair informações genômicas, transcriptômicas e proteômicas em grande escala de maneira rápida, confiável e barata, sendo também conhecidos como dados multi-ômicos. Alguns tipos de dados ômicos que podem ser extraídos da: 1) Genômica: a) polimorfismos de nucleotídeo único; b) variantes raras; c) variação no número de cópias; 2) Epigenômica: a) metilação de DNA; 3) Transcriptômica: a) expressão de RNA mensageiro (mRNA); b) expressão de micro RNA (miRNA); e c) expressão de RNA longo não codificante (lncRNA); 4) Proteômica: a) expressão de proteína. (Antonelli et al., 2019, p. 267-268).

do Brasil e do mundo relacionada à agropecuária. (Embrapa Agricultura Digital, 2021).

No LMB, a implementação de processos de gestão de dados é de fundamental importância porque otimiza o processamento de alto desempenho, facilitando a execução de análises que visam à descoberta, bem como favorece o compartilhamento e o reuso dos dados por outros pesquisadores em projetos de pesquisa ou em outros sistemas, por exemplo. Por conseguinte, a gestão de dados ômicos e, por sua vez, o processo de catalogação de *datasets*, além de serem processos contíguos e inerentes à missão do LMB, também são determinantes para assegurar o compartilhamento, favorecer o reuso e garantir a sustentabilidade dos dados.

Em relação à experiência do LMB no uso da ferramenta Dataverse, por uma decisão de seus pesquisadores e analistas, escolheu-se o projeto de pesquisa “Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus microplus* e espécies relacionadas”, doravante denominado Projeto RmVAC, para ser o projeto-piloto de avaliação do uso deste software. Neste projeto-piloto foram realizadas experiências pioneiras de descrição de metadados e de catalogação de *datasets* no Redape, por meio da ferramenta Dataverse.

Portanto, o processo de catalogação de *datasets* se desenvolveu no ambiente do software Dataverse, utilizando-se um formulário para descrição dos metadados e upload de arquivos de dados. O software Dataverse oferece ao usuário a possibilidade de escolher qual esquema de metadados deseja usar para melhor atender aos interesses e ao contexto dos dados a serem depositados. O *Citation Metadata*, doravante denominado esquema Metadados de Citação, é predefinido pelo software Dataverse e exportável para o padrão *Dublin Core*, inclusive. Em razão disso, este é o esquema genérico de metadados adotado no LMB para a catalogação de *datasets*. No entanto, para atender diretamente ao domínio dos dados biológicos, o Dataverse oferece também o esquema específico de metadados - o *Life Science Metadata* -, aqui denominado de esquema Metadados Ciência da Vida, considerado mais apropriado para descrever *datasets* ômicos.

Desse modo, os esquemas Metadados de Citação e Metadados Ciência da Vida podem ser utilizados concomitantemente. Essa recomendação foi baseada nos estudos realizados no LMB para a escolha de esquemas para catalogar dados ômicos. A utilização simultânea dos dois esquemas está condicionada à estruturação do *dataverse*<sup>13</sup> ao qual o *dataset* estiver vinculado, ou seja, o *dataverse* quando criado deve prever o uso de ambos os esquemas.

O processo de catalogação de *datasets* inicia-se com a atividade de **autodepósito**, a qual deve ser executada pelo autor do *dataset*, na função de catalogador, ou por outra pessoa sob sua responsabilidade. O momento do autodepósito de *datasets* no software Dataverse corresponde à fase de **pré-catalogação**, quando são preenchidos os campos correspondentes aos elementos metadados requeridos, a saber: a) título; b) autor; c) contato; d) descrição; e) assunto; e e) depositante.

Na fase seguinte ocorre a **catalogação** propriamente dita, começando pela revisão do preenchimento prévio dos conteúdos relativos aos elementos metadados requeridos. A descrição catalográfica de *datasets* se ocupa essencialmente do preenchimento dos elementos metadados recomendados e opcionais (apresentados no capítulo seguinte), a ser executado pelo autor do *dataset*, na função de catalogador. Ao bibliotecário e/ou cientista da informação cabe a responsabilidade técnica por revisar esta atividade que requer conhecimentos de normas e padrões oriundos da Biblioteconomia e da Ciência da Informação.

<sup>13</sup> *Dataverse* é aqui entendido como um projeto que reúne conjuntos de dados (*datasets*).

## Regras Gerais para Descrição de *Datasets* Ômicos no Repositório de Dados de Pesquisa da Embrapa (Redape)

Nesta seção são apresentadas as regras que devem ser observadas na descrição de *datasets* no Redape. São orientações para o preenchimento do formulário de catalogação, baseadas nas regras mínimas da AACR2 (Ribeiro, 1995; 2003; Código..., 2004), que visam à clareza e à padronização das informações. Todos os elementos metadados, campos e subcampos exigem atenção no seu preenchimento a fim de evitar erros de qualquer natureza. O preenchimento correto dos campos e subcampos tem impacto direto sobre a qualidade dos *datasets* descritos, bem como de seus meta-dados, que são indexados e recuperados por mecanismos de busca tanto do software Dataverse como por quaisquer outros buscadores na Internet, sobretudo os metabuscadores de repositórios de dados de pesquisa.

As regras básicas estabelecidas que orientam o preenchimento do formulário de catalogação estão indicadas na Tabela 2.

**Tabela 2.** Regras para preenchimento do formulário de catalogação do Redape.

| Orientações/Regras     |   |
|------------------------|---|
| Usar                   | Elementos metadados requeridos, fortemente recomendados, recomendados e opcionais |
|                        | Pontuação   |
|                        | Maiúscula   |
|                        | Ajuda rápida  |
|                        | Palavras-chave e termos livres  |
|                        | Padrão para descrição de datas  |
| Não usar               | Tecla <i>Enter</i>  |
| Controlar e padronizar | Nome de autor pessoal e nome autor corporativo                                    |
|                        | E-mail pessoal  |
|                        | E-mail corporativo  |
|                        | Afiliação   |

Fonte: Adaptado de Souza et al. (2020).

As regras mostradas na Tabela 2 estão detalhadas no item “Outras orientações para a catalogação de dados ômicos no Redape”, apresentado mais adiante.

### Elementos metadados requeridos, fortemente recomendados, recomendados e opcionais

No tocante ao preenchimento do formulário on-line de catalogação de *datasets* no Redape, os elementos metadados são classificados em: a) requeridos; b) fortemente recomendados; c) recomendados; e d) opcionais. **Observação:** Em relação à exigência de preenchimento dos campos, a equipe do LMB decidiu estabelecer mais dois níveis de exigência (fortemente recomendado e recomendado) à estrutura do software Dataverse. O acréscimo desses dois níveis objetivou a ampliação da capacidade de recuperação, reúso e integração dos *datasets* com outros conjuntos e repositórios, de acordo com os princípios FAIR.

*Elementos requeridos* são aqueles considerados ‘essenciais’ à descrição do *dataset*; são de preenchimento obrigatório e podem ser identificados visualmente por um asterisco (\*) após o nome de cada elemento (campo ou subcampo). Os elementos requeridos foram assim definidos

pelo software Dataverse, e caso não sejam preenchidos o sistema impedirá a continuidade da catalogação do *dataset*. São os seguintes os elementos requeridos: a) Título; b) Autor (Nome); c) Contato (E-mail); d) Descrição (Texto); e) Assunto; e f) Depositante.

*Elementos fortemente* recomendados para preenchimento não são de preenchimento obrigatório, visto que o software Dataverse não os considera como elementos requeridos. No entanto, quando preenchidos, os elementos fortemente recomendados fornecem informações adicionais indispensáveis à qualificação do *dataset*, contribuindo para a geração de metadados enriquecidos e aderentes aos princípios FAIR. Daí decorre a importância de seu preenchimento defendida neste documento. Os elementos e/ou campos são: a) Contato (Nome); b) Palavras-chave (Termo-chave); c) Palavras-chave (Vocabulário controlado do termo-chave); d) Palavras-chave (URL do vocabulário controlado); e) Classificação (Termo de classificação); f) Classificação (Vocabulário controlado da classificação); g) Classificação (URL do vocabulário controlado); h) Data da produção; i) Colaborador (Nome); e j) Distribuidor (Nome).

*Elementos recomendados* também não são de preenchimento obrigatório; no entanto, quando preenchidos, os elementos recomendados fornecem informações adicionais relevantes à descrição do *dataset*, contribuindo igualmente para a geração de metadados com qualidade FAIR. Os elementos e/ou campos recomendados são: a) URL alternativa; b) Autor (Afiliação); c) Autor (Sistema identificador); d) Autor (Identificador); e) Contato (Afiliação); f) Publicação relacionada (Citação); g) Publicação relacionada (Tipo de identificador); h) Publicação relacionada (Número de identificador); i) Produtor (Nome do produtor); j) Produtor (Afiliação); k) Local de produção; l) Financiamento (Nome do órgão financiador); m) Financiamento (Número do contrato); n) Distribuidor (Afiliação); o) Data da distribuição; p) Data do depósito; q) Período de cobertura dos dados (Data de início); r) Período de cobertura dos dados (Data de término); s) Período de coleta dos dados (Data de início); t) Período de coleta dos dados (Data de término); u) Tipos de dados; v) Software (Nome); e w) Software (Versão).

*Elementos opcionais* quando preenchidos fornecem informações complementares julgadas pertinentes à descrição do *dataset* pelo catalogador ou revisor da catalogação, visando a descoberta e reuso dos dados. Os elementos e/ou campos opcionais são: a) Título (Subtítulo); b) Título (Título alternativo); c) Outro identificador (Agência); d) Outro identificador (Identificador); e) Descrição (Data); f) Publicação relacionada (URL da publicação); g) Notas; h) Idioma; i) Produtor (Abreviatura do nome do produtor); j) Produtor (URL do produtor); k) Produtor (URL do logotipo do produtor); l) Colaborador (Tipo); m) Distribuidor (Abreviatura); n) Distribuidor (URL do distribuidor); o) Distribuidor (URL do logotipo); p) Série (Nome); q) Série (Informação); r) Material relacionado; s) *Datasets* relacionados; t) Outras referências; u) Fontes de dados; v) Origem das fontes dos dados; w) Características das fontes observadas; e x) Documentação e acesso a fontes de dados.

A Tabela 3 descreve as recomendações para cada nível de exigência de preenchimento dos metadados e a Tabela 4 apresenta os elementos metadados e seus níveis de exigência de preenchimento.

**Tabela 3.** Exigências de preenchimento dos elementos no esquema Metadados de Citação (campos e subcampos).

| Exigência de preenchimento | Descrição  |
|----------------------------|--|
| Requerido                  | Elemento e/ou campo e/ou subcampo requerido - deve ser preenchido. Software Dataverse exige o preenchimento.   |
| Fortemente recomendado     | Elemento e/ou campo e/ou subcampo - preenchimento fortemente recomendado. Embora não seja exigência do software, seu preenchimento fornecerá informações para análise de indicadores de qualidade dos metadados. |
| Recomendado                | Elemento e/ou campo e/ou subcampo – preenchimento recomendado. Embora não seja exigência do software, o seu preenchimento fornecerá informações para aumentar o nível de qualidade dos metadados                 |
| Opcional                   | Elemento e/ou campo e/ou subcampo – preenchimento opcional, ficando a critério do catalogador o fornecimento de informações adicionais. Preencher caso pertinente.   |

Fonte: Adaptado de Souza et al. (2020).

**Tabela 4.** Elementos metadados e níveis de exigência de preenchimento.

| Metadados   | Níveis de exigência               |
|---|-----------------------------------|
| Título  | Elementos requeridos              |
| Autor (Nome)  |                                   |
| Contato (E-mail)  |                                   |
| Descrição (Texto)                                       |                                   |
| Assunto   |                                   |
| Depositante   |                                   |
| Contato (Nome)  | Elementos fortemente recomendados |
| Palavras-chave (Termo-chave)                            |                                   |
| Palavras-chave (Vocabulário controlado do termo-chave)  |                                   |
| Palavras-chave (URL do vocabulário controlado)          |                                   |
| Classificação (Termo de classificação)                  |                                   |
| Classificação (Vocabulário controlado da classificação) |                                   |
| Classificação (URL do vocabulário controlado)           |                                   |
| Data da produção  | Elementos recomendados            |
| Colaborador (Nome)                                      |                                   |
| Distribuidor (Nome)                                     |                                   |
| URL alternativa   |                                   |
| Autor (Afiliação)                                       |                                   |
| Autor (Sistema identificador)                           |                                   |
| Contato (Afiliação)                                     |                                   |
| Publicação relacionada (Citação)                        |                                   |
| Publicação relacionada (Tipo de identificador)          |                                   |
| Publicação relacionada (Número de identificador)        |                                   |
| Produtor (Nome)   |                                   |
| Produtor (Afiliação)                                    |                                   |
| Local de produção                                       |                                   |
| Financiamento (Nome do órgão financiador)               |                                   |
| Financiamento (Número do contrato)                      |                                   |
| Distribuidor (Afiliação)                                |                                   |
| Data da distribuição                                    |                                   |
| Data do depósito  |                                   |
| Período de cobertura dos dados (Data de início)         |                                   |
| Período de cobertura dos dados (Data de término)        |                                   |
| Período de coleta dos dados (Data de início)            |                                   |
| Período de coleta dos dados (Data de término)           |                                   |
| Tipo de dados   |                                   |
| Software (Nome)   |                                   |
| Software (Versão)                                       |                                   |

Continua...

|  |                     |
|--|---------------------|
| Título (Subtítulo)                         | Elementos opcionais |
| Título (Título alternativo)                |                     |
| Outro identificador (Agência)              |                     |
| Outro identificador (Identificador)        |                     |
| Descrição (Data)                           |                     |
| Publicação relacionada (URL da publicação) |                     |
| Notas                                      |                     |
| Idioma                                     |                     |
| Produtor (Abreviatura do nome do produtor) |                     |
| Produtor (URL do produtor)                 |                     |
| Produtor (URL do logotipo)                 |                     |
| Colaborador (Tipo)                         |                     |
| Distribuidor (Abreviatura)                 |                     |
| Distribuidor (URL do distribuidor)         |                     |
| Distribuidor (URL do logotipo)             |                     |
| Série (Nome)                               |                     |
| Série (Informação)                         |                     |
| Material relacionado                       |                     |
| <i>Datasets</i> relacionados               |                     |
| Outras referências                         |                     |
| Fontes de dados                            |                     |
| Origem das fontes de dados                 |                     |
| Características das fontes observadas      |                     |
| Documentação acesso a fontes de dados      |                     |

Fonte: Adaptado de Souza et al. (2020).

## Outras orientações para a catalogação de dados ômicos no Redape

### *Pontuação*

Não usar ponto final no preenchimento dos campos, exceto naqueles em que o fechamento do período é necessário, como é o caso dos elementos Descrição, Citação e Notas.

### *Maiúsculas*

Usar letra maiúscula inicial para o preenchimento de campos dos elementos Título e Palavras-chaves, e para nomes próprios. No elemento Descrição, o uso de maiúscula deve obedecer à norma culta da língua. Nenhuma palavra deve ser escrita em letras maiúsculas, exceto siglas.

### *Tecla Enter*

A tecla *Enter* não deve ser usada, pois, caso isso aconteça, o registro será fechado imediatamente. Entretanto, se isso ocorrer, o usuário poderá utilizar a opção de editar os metadados.

### *Botões de acesso*

O botão  que aparece em alguns campos e/ou subcampos do formulário de catalogação indica que é possível adicionar mais de uma ocorrência durante a descrição do *dataset*. O botão  indica que é possível excluir uma ocorrência de um campo e/ou subcampo; ou seja, com este botão é possível eliminar do formulário um campo ou subcampo preenchido anteriormente.

### *Ajuda rápida*

Para cada elemento metadado ou campo a ser preenchido dentro do formulário de catalogação está disponível uma “ajuda rápida”. Para acessar essa ajuda basta clicar sobre o ícone  posicionado à frente do elemento metadado, campo ou subcampo.

## **Nomes pessoais e nomes corporativos, e-mails pessoais e e-mails corporativos e afiliação**

Para facilitar, agilizar e padronizar o preenchimento desses campos, é desejável que o Redape esteja integrado com o banco de dados oficial de empregados da Embrapa. Caso este recurso não esteja disponível, o catalogador deverá consultar a página oficial de empregados da Embrapa<sup>14</sup>. Essas ações ajudarão na consistência e concisão das informações, evitando ambiguidades e retrabalho na descrição dos conteúdos dos *datasets*.

### *Nome de autor pessoal*

Nome do autor quando este é uma pessoa física. Deve ser descrito no formato de sobrenome, nome. Seguem exemplos:

Silva, José

Pereira Júnior, João

### *Nome de autor corporativo*

Nome do autor quando este é uma instituição. O preenchimento, neste caso, deve ser o nome da instituição, sem abreviação.

Exemplo: Empresa Brasileira de Pesquisa Agropecuária

### *E-mail pessoal*

Endereço de e-mail de uma pessoa. Deve ser preenchido por um endereço válido, composto de duas partes: nome do usuário e nome do domínio, ligadas pelo caracter @ - Exemplo: usuario@provedor.com.br

### *E-mail corporativo*

Endereço de e-mail de uma instituição. Deve ser preenchido por um endereço válido, composto de duas partes: nome do usuário e nome do domínio, ligadas pelo caracter @ - Exemplo: usuario@provedor.com.br

### *Afiliação*

Nome de instituição onde está vinculada uma pessoa.

### *Palavras-chave – termos livres*

O campo palavras-chave possibilita que sejam inseridos termos de vocabulários controlados e termos livres. As orientações detalhadas sobre o uso de vocabulários controlados são apresentadas no Apêndice A deste documento.

Já os termos livres são aqueles que não estão descritos em vocabulários controlados. O uso deles possibilita que na descrição temática seja utilizada linguagem mais recente ou mesmo paralela àquela adotada oficialmente por especialistas e que não se encontram nas fontes de informação dos vocabulários controlados. Recomenda-se cautela na escrita, tanto em nível ortográfico, quanto na quantidade utilizada, para que não haja prejuízo na recuperação dos *datasets*.

### *Data*

Para o preenchimento de campos correspondentes a datas, adota-se a recomendação da ISO 8601<sup>15</sup> (*Date and Time Format*), cujo formato é AAAA-MM-DD, em que AAAA é o ano, MM é o mês, e DD o dia.

<sup>14</sup> Disponível em: <<https://www.embrapa.br/equipe>>.

<sup>15</sup> Disponível em: <<https://www.w3.org/TR/NOTE-datetime>>.

## Descrição dos Elementos no Esquema Metadados de Citação

Para a catalogação de *datasets*, o software Dataverse possui esquema de metadados predefinido denominado de Metadados de Citação. Esse conjunto é formado por 33 elementos, constituídos de campos e subcampos. A Tabela 5 apresenta os 33 elementos do esquema Metadados de Citação, que compõem o formulário de catalogação no software Dataverse, em uso no Redape, bem como os respectivos níveis de exigência de preenchimento.

**Tabela 5.** Elementos do esquema Metadados de Citação do formulário de catalogação do Redape.

| Elemento metadado                              | Exigência              |
|--|------------------------|
| Título   | Requerido              |
| URL alternativa                                | Recomendado            |
| Outro identificador                            | Opcional               |
| Autor  | Requerido              |
| Contato  | Requerido              |
| Descrição                                      | Requerido              |
| Assunto  | Requerido              |
| Palavras-chave                                 | Fortemente recomendado |
| Classificação                                  | Fortemente recomendado |
| Publicação relacionada                         | Opcional               |
| Notas  | Opcional               |
| Idioma   | Opcional               |
| Produtor                                       | Recomendado            |
| Data de produção                               | Fortemente recomendado |
| Local de produção                              | Recomendado            |
| Colaborador                                    | Fortemente recomendado |
| Financiamento                                  | Recomendado            |
| Distribuidor                                   | Fortemente recomendado |
| Data de distribuição                           | Recomendado            |
| Depositante                                    | Requerido              |
| Data de depósito                               | Recomendado            |
| Período de cobertura de dados                  | Recomendado            |
| Período de coleta de dados                     | Recomendado            |
| Tipos de dados                                 | Recomendado            |
| Série  | Opcional               |
| Software                                       | Recomendado            |
| Material selecionado                           | Opcional               |
| <i>Datasets</i> relacionados                   | Opcional               |
| Outras referências                             | Opcional               |
| Fontes de dados                                | Opcional               |
| Origem das fontes de dados                     | Opcional               |
| Características das fontes de dados observadas | Opcional               |
| Documentação e acesso a fontes de dados        | Opcional               |

Fonte: Souza et al. (2020).

Para a catalogação de *datasets* foi adotado um conjunto de atributos comuns que orientam a descrição de todos os elementos metadados (campos e subcampos), conforme mostra a Tabela 6.

**Tabela 6.** Atributos orientadores para a descrição dos elementos do esquema Metadados de Citação (campos e subcampos) no Redape.

|                            |   |
|----------------------------|---|
| Nome                       | A denominação atribuída para o elemento de dado.  |
| Identificador              | O identificador único atribuído para o elemento de dado.  |
| Ocorrência                 | Indica qualquer limite para a repetibilidade do elemento de dado.                               |
| Exigência de preenchimento | Indica se o elemento necessita estar presente sempre ou esporadicamente.                        |
| Definição                  | Uma declaração que claramente representa o conceito e a natureza essencial do elemento de dado. |
| Comentário                 | Um comentário relativo à aplicação do elemento de dado.   |
| Exemplo                    | Um exemplo relativo à aplicação do elemento de dado.  |
| Documento orientador       | Um documento que oriente a aplicação do elemento de dado.                                       |

Fonte: Souza et al. (2020).

A seguir são apresentados os 33 elementos do esquema Metadados de Citação com seus campos e subcampos e as respectivas descrições para a catalogação de *dataset* (Tabela 7).

**Tabela 7.** Descrição dos elementos do esquema Metadados Citação (campos e subcampos) no Redape.

| Título*                    |   |
|----------------------------|---|
| Identificador              | Título  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Requerido   |
| Definição                  | Nome atribuído ao <i>dataset</i> , e pelo qual é formalmente conhecido. Título deve ser formado pela estrutura a seguir: <b>Tipo de estudo ômico + Amostra estudada + Espécie estudada (nome comum + nome científico)</b> . |
| Comentário                 | Descrever o título do <i>dataset</i> observando a norma culta da língua, quanto à grafia e acentuação. Iniciais de título, de nome próprio e de nome científico devem ser grafadas em maiúscula.                            |
| Exemplo                    | Transcriptoma da glândula salivar do carrapato bovino ( <i>Rhipicephalus (B) microplus</i> ).   |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.              |
| Subtítulo                  |   |
| Identificador              | Subtítulo   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Nome secundário atribuído ao <i>dataset</i> para generalizar ou especificar a abrangência do título.  |
| Comentário                 | Caso exista subtítulo, descreva-o de modo a ampliar ou restringir o título.   |
| Exemplo                    | Título:<br>Genoma do carrapato montado<br>Subtítulo:<br>Expressão gênica na saliva do carrapato bovino  |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.              |

Continua...

**Título alternativo**

|                            |   |
|----------------------------|---|
| Identificador              | Título alternativo  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Nome alternativo para o <i>dataset</i> , pelo qual também é conhecido ou referido. Pode ser ainda uma forma abreviada do título.  |
| Comentário                 | Caso exista título alternativo, transcrevê-lo de modo a dar conhecimento à forma como o <i>dataset</i> também é conhecido.  |
| Exemplo                    | Genoma do carrapato   |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catalogação de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular. |

**URL alternativa**

|                            |  |
|----------------------------|--|
| Identificador              | URL alternativa  |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Recomendado  |
| Definição                  | URL alternativa válida indica onde o <i>dataset</i> pode também ser visualizado. A URL não pode ser dependente de qualquer tipo de autenticação, uso de <i>captcha</i> , ou qualquer outro meio que impeça o acesso direto ao <i>dataset</i> .   |
| Comentário                 | Transcrever a URL alternativa para acesso ao título do <i>dataset</i> .  |
| Exemplos                   | <a href="https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material">https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material</a><br><a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA596777">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA596777</a> |
| Documento orientador       | Não se aplica.   |

**Outro identificador**

|                            |  |
|----------------------------|--|
| Identificador              | Outro identificador  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | Outro identificador único digital que identifica o <i>dataset</i> , por exemplo, o número que possua em outro repositório. |
| Comentário                 | É composto de dois subcampos: Agência e Identificador.   |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Não se aplica.   |

**Agência**

|                            |  |
|----------------------------|--|
| Identificador              | Agência do identificador digital                                 |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | Nome da agência que gerou o identificador digital.               |
| Comentário                 | Transcrever o nome da agência que gerou o identificador digital. |
| Exemplo                    | NCBI   |
| Documento orientador       | Não se aplica.   |

Continua...

| Identificador              |   |
|----------------------------|---|
| Identificador              | Identificador digital persistente   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Outro Identificador digital que identifica o <i>dataset</i> , atribuído pela Agência descrita no subcampo anterior. |
| Comentário                 | Transcrever outro identificador digital do <i>dataset</i> .   |
| Exemplo                    | PRJNA596777   |
| Documento orientador       | Não se aplica.  |

| Autor                      |  |
|----------------------------|--|
| Identificador              | Autor  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Requerido  |
| Definição                  | Nome da(s) pessoa(s) ou da instituição responsáveis pela criação do <i>dataset</i> .     |
| Comentário                 | É composto por quatro subcampos: Nome, Afiliação, Sistema identificador e Identificador. |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Não se aplica.   |

| Nome*                      |   |
|----------------------------|---|
| Identificador              | Nome  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Requerido   |
| Definição                  | Nome(s) do(s) autor(es) do <i>dataset</i> . Pode(m) ser autor(es) pessoal(is) ou autor corporativo.   |
| Comentário                 | Quando o autor é pessoa física, preenche-se com o(s) nome(s) de pessoa(s), no formato sobrenome(s), seguidos dos prenomes (observar os casos com grau de parentesco, sobrenome hispânico, nome composto e sobrenome com prefixos, conforme os exemplos a seguir). Caso seja um autor corporativo, descreve-se o nome da instituição, sem abreviações.   |
| Exemplos                   | <p>Nomes de autores pessoais:</p> <p>Alves, Eliseu Roberto de Andrade<br/> Pereira Júnior, João (grau de parentesco)<br/> García Marquez, Gabriel (sobrenome hispânico)<br/> Espírito Santo, Miguel Frederico de (nome composto)<br/> D'Ambrosio, Ubiratan (sobrenome com prefixo)</p> <p>Nome de autor corporativo:</p> <p>Embrapa (nome síntese)<br/> Empresa Brasileira de Pesquisa Agropecuária</p> |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.  |

| Afiliação                  |  |
|----------------------------|--|
| Identificador              | Afiliação  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome da instituição de vínculo do(s) autor(es) pessoal(is) durante o período de criação ou produção do <i>dataset</i> .  |
| Comentário                 | Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome.<br>Caso o autor do <i>dataset</i> seja um autor corporativo, não é necessário preencher este campo.             |
| Exemplos                   | <p>Empresa Brasileira de Pesquisa Agropecuária<br/> Embrapa Gado de Corte<br/> Embrapa Agricultura Digital</p>   |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular. |

**Sistema do identificador**

|                            |   |
|----------------------------|---|
| Identificador              | Sistema do identificador digital do autor   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Nome do sistema do identificador digital escolhido para a identificação de cada autor pessoal do <i>dataset</i> . O uso de identificador digital persistente permite diferenciar um autor de outro. Também viabiliza a conexão automática entre o autor e suas atividades profissionais. Um sistema bastante utilizado é o ORCID ( <a href="https://orcid.org/">https://orcid.org/</a> ). |
| Comentário                 | Selecionar o sistema de identificação digital. Caso o autor seja corporativo, este campo não deve ser preenchido.   |
| Exemplos                   | ORCID<br>ResearcherID<br>Scopus ID  |
| Documento orientador       | Não se aplica.  |

**Identificador**

|                            |   |
|----------------------------|---|
| Identificador              | Identificador digital persistente   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Identificador digital persistente do(s) autor(es) pessoal(is) do <i>dataset</i> .   |
| Comentário                 | A partir do sistema identificador selecionado, transcrever o identificador digital persistente de cada autor pessoal do <i>dataset</i> . Caso o autor seja corporativo, este campo não deve ser preenchido. |
| Exemplo                    | No esquema ORCID um exemplo é:<br><a href="http://orcid.org/0000-0002-1825-0097">http://orcid.org/0000-0002-1825-0097</a> (ORCID de Josiah Carberry)  |
| Documento orientador       | Não se aplica.  |

**Contato**

|                            |  |
|----------------------------|--|
| Identificador              | Contato  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Requerido  |
| Definição                  | Nome da(s) pessoa(s) ou da instituição para contato sobre o <i>dataset</i> . |
| Comentário                 | É composto por três subcampos: Nome, Afiliação e E-mail.                     |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Não se aplica.   |

**Nome**

|                            |  |
|----------------------------|--|
| Identificador              | Nome   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Nome pessoal ou corporativo para contato sobre o <i>dataset</i> .  |
| Comentário                 | Nome pessoal deve ser descrito na forma direta, ou seja, prenome(s) seguido(s) de sobrenome(s).<br>Caso o contato seja de uma corporação, preencha com o nome da instituição, sem abreviações.                 |
| Exemplos                   | Nome pessoal:<br>José Silva<br>João Pereira Júnior<br>Nome de corporativo para contato:<br>Embrapa Agricultura Digital   |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular. |

### Afiliação

|                            |  |
|----------------------------|--|
| Identificador              | Afiliação do contato   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome da instituição à qual a pessoa indicada para contato está vinculada.  |
| Comentário                 | Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome.<br>Caso a instituição já seja a indicada para contato sobre o <i>dataset</i> , não é necessário preencher este campo. |
| Exemplos                   | Empresa Brasileira de Pesquisa Agropecuária<br>Embrapa Gado de Corte<br>Embrapa Agricultura Digital  |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.       |

### E-mail do contato\*

|                            |   |
|----------------------------|---|
| Identificador              | E-mail  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Requerido   |
| Definição                  | Endereço de e-mail do contato sobre o <i>dataset</i> , seja pessoal ou corporativo. |
| Comentário                 | Preencher com o e-mail do contato.  |
| Exemplo                    | jose.silva@empresa.com.br   |
| Documento orientador       | Não se aplica.  |

### Descrição

|                            |   |
|----------------------------|---|
| Identificador              | Descrição   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Requerido   |
| Definição                  | Um sumário descritivo e sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> . |
| Comentário                 | É composto por dois subcampos: Texto e Data   |
| Exemplo                    | Não se aplica   |
| Documento orientador       | Não se aplica.  |

### Texto\*

|                            |  |
|----------------------------|--|
| Identificador              | Texto  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Requerido  |
| Definição                  | Texto sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> .              |
| Comentário                 | Texto livre.   |
| Exemplo                    | <i>Dataset</i> réplica 1 da glândula salivar Holandês, dados brutos formato <i>fastq</i> . |
| Documento orientador       | Não se aplica.   |

Continua...

**Data**

|                            |  |
|----------------------------|--|
| Identificador              | Data   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | Data em que a descrição do <i>dataset</i> foi realizada. Um <i>dataset</i> pode ter mais de uma descrição, realizada em datas distintas. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.  |
| Exemplo                    | 2020-06-08   |
| Documento orientador       | ISO 8601 Date and Time Format.   |

**Assunto\***

|                            |   |
|----------------------------|---|
| Identificador              | Assunto   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Requerido   |
| Definição                  | Áreas de domínio abrangentes e relevantes com as quais o <i>dataset</i> está relacionado.                                   |
| Comentário                 | Selecionar a opção de área de domínio com as quais o <i>dataset</i> está relacionado. Pode-se selecionar mais de uma opção. |
| Exemplos                   | Agricultural Sciences<br>Medicine, Health and Life Sciences   |
| Documento orientador       | Apêndice A – Vocabulários Controlados.  |

**Palavras-chave**

|                            |  |
|----------------------------|--|
| Identificador              | Palavras-chave   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Palavras que expressam o conteúdo-chave do <i>dataset</i> , sendo capazes de representar aspectos relevantes para a compreensão do assunto. Palavras-chave podem ser utilizadas na construção de índices, na realização de buscas e na recuperação de informação. Palavras-chaves podem estar vinculadas ou não a um vocabulário controlado. |
| Comentário                 | É composto de três subcampos: Termo-chave, Vocabulário controlado do termo-chave e URL do vocabulário controlado.  |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Apêndice A – Vocabulários Controlados.   |

**Termo-chave**

|                            |  |
|----------------------------|--|
| Identificador              | Termo-chave  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Termos-chave são indicados para representar aspectos importantes de um <i>dataset</i> . Podem ser usados na construção de índices, na realização de busca e na recuperação de informação. Recomenda-se a utilização de vocabulário controlado para atribuição de termos-chave. |
| Comentário                 | Preencher com termos-chave que descrevam aspectos relevantes do <i>dataset</i> . Podem estar vinculados a um vocabulário controlado ou não.  |
| Exemplos                   | Host-parasite relationships<br>Nucleotide sequences<br><i>Rhipicephalus (Boophilus) microplus</i><br>Ticks<br>Transcriptome  |
| Documento orientador       | Apêndice A – Vocabulários Controlados<br>(AGROVOC <i>Thesaurus</i> )<br>(NAL <i>Thesaurus</i> )<br>(NCBI <i>Taxonomy</i> )   |

### Vocabulário controlado do termo-chave

|                            |  |
|----------------------------|--|
| Identificador              | Vocabulário controlado do termo-chave  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Vocabulário controlado utilizado para a atribuição de termos-chave do <i>dataset</i> .                                     |
| Comentário                 | Preencher com nome do vocabulário controlado utilizado para a atribuição de termos-chave para o <i>dataset</i> .           |
| Exemplos                   | AGROVOC <i>Thesaurus</i><br>NAL <i>Thesaurus</i><br>NCBI <i>Taxonomy</i>   |
| Documento orientador       | Apêndice A – Vocabulários Controlados<br>(AGROVOC <i>Thesaurus</i> )<br>(NAL <i>Thesaurus</i> )<br>(NCBI <i>Taxonomy</i> ) |

### URL do vocabulário controlado

|                            |  |
|----------------------------|--|
| Identificador              | URL do vocabulário controlado  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | URL do vocabulário controlado onde o termo poderá ser acessado.  |
| Comentário                 | <b>Atenção:</b> Neste campo não utilizar a URL do vocabulário controlado; em seu lugar preencher o campo com a URL válida do termo-chave (e não a URL do vocabulário controlado).<br><b>Observação:</b> No formulário Dataverse não há campo destinado à descrição da <b>URL do termo-chave</b> , por esta razão, escolheu-se o campo <b>URL do vocabulário controlado</b> para fazer o seu registro.                                  |
| Exemplos                   | URL do termo-chave 'nucleotide sequence' (AGROVOC <i>Thesaurus</i> ):<br><a href="http://aims.fao.org/aos/agrovoc/c_27583">http://aims.fao.org/aos/agrovoc/c_27583</a><br>URL do termo-chave 'nucleotide sequences' (NAL <i>Thesaurus</i> ): <a href="https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2">https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2</a> |
| Documento orientador       | Apêndice A – Vocabulários Controlados<br>(AGROVOC <i>Thesaurus</i> )<br>(NAL <i>Thesaurus</i> )<br>(NCBI <i>Taxonomy</i> )   |

### Classificação

|                            |   |
|----------------------------|---|
| Identificador              | Classificação   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Fortemente recomendado  |
| Definição                  | Refere-se à classificação do assunto do <i>dataset</i> , por meio da atribuição de uma ou mais categorias, sob as quais podem ser agrupados <i>datasets</i> de natureza semelhante. Recomenda-se o uso de categorias vinculadas a vocabulário controlado. |
| Comentário                 | É composto de três subcampos: Termo de classificação, Vocabulário controlado da classificação e URL do vocabulário controlado.  |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Apêndice A – Vocabulários Controlados<br>(AGROVOC <i>Thesaurus</i> )<br>(NAL <i>Thesaurus</i> )<br>(NCBI <i>Taxonomy</i> )  |

Continua...

**Termo de classificação**

|                            |  |
|----------------------------|--|
| Identificador              | Termo de classificação   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Indicação da classificação de assunto (categoria) de maior abrangência, na qual o <i>dataset</i> se classifica.  |
| Comentário                 | Preencher com a expressão que indique a classificação de assunto mais abrangente para representar o <i>dataset</i> . O uso termos associados a vocabulários controlados é recomendado. |
| Exemplos                   | Animal Science and Animal Products (NAL <i>Thesaurus</i> )<br>Phenomena (AGROVOC <i>Thesaurus</i> )<br>Ixodida (NCBI <i>Taxonomy</i> )   |
| Documento orientador       | Apêndice A – Vocabulários Controlados (AGROVOC <i>Thesaurus</i> ) (NAL <i>Thesaurus</i> ) (NCBI <i>Taxonomy</i> )  |

**Vocabulário controlado da classificação**

|                            |  |
|----------------------------|--|
| Identificador              | Vocabulário controlado da classificação  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Vocabulário controlado utilizado para a classificação de assunto do <i>dataset</i> .   |
| Comentário                 | Preencher com nome do vocabulário controlado utilizado para a atribuição da classificação de assunto para representar o <i>dataset</i> . |
| Exemplos                   | NAL <i>Thesaurus</i><br>NCBI <i>Taxonomy</i><br>AGROVOC <i>Thesaurus</i>   |
| Documento orientador       | Apêndice A – Vocabulários Controlados (AGROVOC <i>Thesaurus</i> ) (NAL <i>Thesaurus</i> ) (NCBI <i>Taxonomy</i> )                        |

**URL do vocabulário controlado**

|                            |  |
|----------------------------|--|
| Identificador              | URL do vocabulário controlado  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | URL do vocabulário controlado onde o termo de classificação de assunto (categoria) do <i>dataset</i> poderá ser acessado.  |
| Comentário                 | <b>Atenção:</b> Neste campo não utilizar a URL do vocabulário controlado; em seu lugar preencher o campo com a URL válida do termo de classificação (e não a URL do vocabulário controlado).<br><b>Observação:</b> No formulário Dataverse não há campo destinado à descrição da URL do termo de classificação, por esta razão, escolheu-se o campo URL do vocabulário controlado para fazer o seu registro.   |
| Exemplos                   | URL do termo de classificação 'phenomena' (AGROVOC <i>Thesaurus</i> ):<br><a href="http://aims.fao.org/aos/agrovoc/c_330704">http://aims.fao.org/aos/agrovoc/c_330704</a><br>URL do termo de classificação 'L Animal Science and Animal Products' (NAL <i>Thesaurus</i> ):<br><a href="https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100">https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100</a><br>URL do termo de classificação 'Ixodida' (NCBI <i>Taxonomy</i> ):<br><a href="https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&amp;id=6935&amp;lvl=3&amp;p=nucore&amp;lin=f&amp;keep=1&amp;srchmode=1&amp;unlock">https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&amp;id=6935&amp;lvl=3&amp;p=nucore&amp;lin=f&amp;keep=1&amp;srchmode=1&amp;unlock</a> |
| Documento orientador       | Apêndice A – Vocabulários Controlados (AGROVOC <i>Thesaurus</i> ) (NAL <i>Thesaurus</i> ) (NCBI <i>Taxonomy</i> )  |

Continua...

### Publicação relacionada

|                            |  |
|----------------------------|--|
| Identificador              | Publicação relacionada   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Refere-se a publicação que usa dados do <i>dataset</i> que está sendo catalogado.                            |
| Comentário                 | É composto de quatro subcampos: Citação, Tipo de identificador, Número do identificador e URL da publicação. |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Não se aplica.   |

### Citação

|                            |   |
|----------------------------|---|
| Identificador              | Citação   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Refere-se à referência completa da publicação relacionada ao <i>dataset</i> que está sendo catalogado.  |
| Comentário                 | Preencher em formato de referência bibliográfica a publicação relacionada ao <i>dataset</i> . Recomenda-se adotar o padrão ABNT 6023.   |
| Exemplo                    | GIACHETTO, P. F. Transcriptoma do carrapato dos bovinos. In: GARCIA, M. V.; KOLLER, W. W. (Ed.). <b>Carrapatos na cadeia produtiva de bovinos</b> . Brasília, DF: Embrapa, 2019. cap. 16. p. 205-222. |
| Documento orientador       | ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. <b>NBR 6023</b> : informação e documentação – referências: elaboração. 2. ed. Rio de Janeiro, 2020. 68 p. Versão corrigida 2.                               |

### Tipo de Identificador

|                            |  |
|----------------------------|--|
| Identificador              | Tipo de identificador  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Tipo de identificador digital usado pela publicação.   |
| Comentário                 | Indicar a opção que corresponda ao tipo de identificador digital utilizado pela publicação relacionada ao <i>dataset</i> . |
| Exemplo                    | DOI  |
| Documento orientador       | Não se aplica  |

### Número do Identificador

|                            |   |
|----------------------------|---|
| Identificador              | Número do identificador   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Identificador digital único que remete para a publicação relacionada ao <i>dataset</i> .  |
| Comentário                 | Transcrever o identificador digital único da publicação relacionada ao <i>dataset</i> .   |
| Exemplo                    | Número do identificador digital:<br><a href="https://doi.org/10.3389/fcimb.2019.00477">https://doi.org/10.3389/fcimb.2019.00477</a> |
| Documento orientador       | Não se aplica.  |

Continua...

**URL da publicação**

|                            |   |
|----------------------------|---|
| Identificador              | URL da publicação   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | URL da página web onde se encontra a publicação. Pode ser a página de um periódico indexado, revista etc.   |
| Comentário                 | Transcrever a URL da página web da publicação relacionada ao <i>dataset</i> .   |
| Exemplo                    | URL da publicação:<br><a href="https://www.frontiersin.org/journals/cellular-and-infection-microbiology">https://www.frontiersin.org/journals/cellular-and-infection-microbiology</a> |
| Documento orientador       | Não se aplica.  |

**Notas**

|                            |   |
|----------------------------|---|
| Identificador              | Notas   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Informações adicionais sobre o <i>dataset</i> .   |
| Comentário                 | Texto livre.  |
| Exemplo                    | Réplica de dados de glândula salivar do carrapato bovino infestando gado Holandês. Ensaio envolve dados de gado Nelore e mestiço Nelore x Holandês. |
| Documento orientador       | Não se aplica.  |

**Idioma**

|                            |   |
|----------------------------|---|
| Identificador              | Idioma  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Idioma do <i>dataset</i> .  |
| Comentário                 | Selecionar um ou mais idiomas a partir da lista de opções oferecidas. |
| Exemplos                   | Inglês<br>Inglês, Português<br>Inglês, Português, Espanhol            |
| Documento orientador       | Não se aplica.  |

**Produtor**

|                            |  |
|----------------------------|--|
| Identificador              | Produtor   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome pessoal ou nome corporativo do responsável financeiro ou administrativo pelo <i>dataset</i> .   |
| Comentário                 | É composto por cinco subcampos: Nome do produtor, Afiliação do produtor, Abreviatura do nome do produtor, URL do produtor e URL do logotipo do produtor. |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Não se aplica.   |

Continua...

### Nome do produtor

|                            |  |
|----------------------------|--|
| Identificador              | Nome do produtor   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome pessoal ou corporativo do produtor do <i>dataset</i> .  |
| Comentário                 | Nome pessoal deve ser descrito na forma direta, ou seja, prenomes (s) seguidos (s) de sobrenomes (s).<br>Caso o produtor seja um nome corporativo, preencher com nome da instituição, sem abreviações.         |
| Exemplos                   | Nome pessoal:<br>José Silva<br>Nome corporativo:<br>Empresa Brasileira de Pesquisa Agropecuária  |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular. |

### Afiliação do produtor

|                            |  |
|----------------------------|--|
| Identificador              | Afiliação do produtor  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome da instituição à qual o produtor do <i>dataset</i> está vinculado.  |
| Comentário                 | Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome.<br>Caso a instituição seja aquela indicada como o produtor do <i>dataset</i> , não é necessário preencher este campo. |
| Exemplos                   | Embrapa Gado de Corte<br>Empresa Brasileira de Pesquisa Agropecuária   |
| Documento orientador       | RIBEIRO, A. M. de C. M. <b>Catálogo de recursos bibliográficos pela AACR2 2002</b> : Anglo-American Cataloguing Rules, 2nd edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.       |

### Abreviatura do nome do produtor

|                            |   |
|----------------------------|---|
| Identificador              | Abreviatura do nome do produtor   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Abreviatura do nome do produtor do <i>dataset</i> .   |
| Comentário                 | Indicar a abreviatura ou sigla do nome do produtor quando se tratar de um nome corporativo. |
| Exemplos                   | Embrapa<br>LMB  |
| Documento orientador       | Não se aplica.  |

### URL do produtor

|                            |   |
|----------------------------|---|
| Identificador              | URL do produtor   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | URL válida de uma página na web do produtor do <i>dataset</i> .   |
| Comentário                 | Transcrever a URL válida da página web do produtor do <i>dataset</i> .  |
| Exemplos                   | URL Empresa Brasileira de Pesquisa Agropecuária:<br><a href="https://www.embrapa.br/">https://www.embrapa.br/</a><br>URL do Laboratório Multiusuário de Bioinformática da Embrapa:<br><a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a> |
| Documento orientador       | Não se aplica.  |

**URL do logotipo do produtor**

|                            |  |
|----------------------------|--|
| Identificador              | URL do logotipo do produtor  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Comentário                 | Transcrever uma URL válida da página web onde o logotipo do produtor pode ser visualizado.   |
| Exemplos                   | URL do logotipo da Embrapa:<br><a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a><br>URL do logotipo do Laboratório Multiusuário de Bioinformática da Embrapa:<br><a href="https://www.embrapa.br/documents/1355331/34714723/Banner+lmb/e20c531f-d6ea-2d2c-2acd-ff1123a9a979?t=1528312726048">https://www.embrapa.br/documents/1355331/34714723/Banner+lmb/e20c531f-d6ea-2d2c-2acd-ff1123a9a979?t=1528312726048</a> |
| Documento orientador       | Não se aplica.   |

**Data de produção**

|                            |   |
|----------------------------|---|
| Identificador              | Data de produção                            |
| Ocorrência                 | Limitada                                    |
| Exigência de preenchimento | Fortemente recomendado                      |
| Definição                  | Data em que o <i>dataset</i> foi produzido. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.                 |
| Exemplo                    | 2015-05-01                                  |
| Documento orientador       | ISO 8601 Date and Time Format.              |

**Local de produção**

|                            |   |
|----------------------------|---|
| Identificador              | Local de produção   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Local onde o <i>dataset</i> e quaisquer outros materiais relacionados foram produzidos. |
| Comentário                 | Texto livre.  |
| Exemplo                    | Laboratório de Biotecnologia Animal. Centro de Genômica Funcional                       |
| Documento orientador       | Não se aplica.  |

**Colaborador**

|                            |   |
|----------------------------|---|
| Identificador              | Colaborador   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Fortemente recomendado  |
| Definição                  | Pessoa ou instituição que colabora com a coleta, gestão ou qualquer outra contribuição relevante para o desenvolvimento do <i>dataset</i> . |
| Comentário                 | É composto por dois subcampos: Tipo de colaborador e Nome do colaborador.   |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Não se aplica.  |

Continua...

### Tipo de colaborador

|                            |   |
|----------------------------|---|
| Identificador              | Tipo de colaborador   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Tipo de colaborador do <i>dataset</i> .   |
| Comentário                 | Selecionar um tipo de colaborador a partir da lista de opções oferecidas. Pode-se selecionar mais de um tipo de colaborador, mas para isso, é necessário adicionar mais uma ocorrência. |
| Exemplos                   | Coletor de dados<br>Curador de dados<br>Líder do projeto<br>Pesquisador   |
| Documento orientador       | Não se aplica.  |

### Nome do colaborador

|                            |  |
|----------------------------|--|
| Identificador              | Nome do colaborador  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Nome pessoal ou corporativo do colaborador do <i>dataset</i> .   |
| Comentário                 | Nomes pessoais devem ser indicados na forma direta, ou seja, prenomes seguidos de sobrenomes.<br>Caso o nome do colaborador seja um nome corporativo, preencher o nome completo da instituição; não usar abreviaturas ou suprimir partes do nome da instituição. |
| Exemplos                   | Nome pessoal de colaborador:<br>Paula Regina Kuser Falcão<br>Nome corporativo (instituição) de colaborador:<br>Laboratório de Biotecnologia Animal Centro de Genômica Funcional  |
| Documento orientador       | Não se aplica.   |

### Financiamento

|                            |   |
|----------------------------|---|
| Identificador              | Financiamento   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Informações sobre instituição, agência de fomento, organização e/ou projeto responsável pela concessão de suporte e/ou apoio financeiro para a produção do <i>dataset</i> . |
| Comentário                 | É composto pelos subcampos: Nome do órgão financiador e Número do contrato  |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Não se aplica.  |

### Nome do órgão financiador

|                            |   |
|----------------------------|---|
| Identificador              | Nome do órgão financiador   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Nome da instituição, agência de fomento ou organização que concedeu suporte e/ou apoio financeiro para a produção do <i>dataset</i> .   |
| Comentário                 | Texto livre.  |
| Exemplos                   | Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)<br>Empresa Brasileira de Pesquisa Agropecuária (Embrapa) Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) |
| Documento orientador       | Não se aplica.  |

**Número do contrato**

|                            |   |
|----------------------------|---|
| Identificador              | Número do contrato  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Número do contrato de concessão de financiamento dos recursos para a produção do <i>dataset</i> . |
| Comentário                 | Transcrever o número do contrato, caso exista.  |
| Exemplo                    | 01.11.07.002.03.00 (projeto SEG, da Embrapa)  |
| Documento orientador       | Não se aplica.  |

**Distribuidor**

|                            |   |
|----------------------------|---|
| Identificador              | Distribuidor  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Fortemente recomendado  |
| Definição                  | Pessoa ou instituição designada pelo autor ou produtor do <i>dataset</i> para gerar cópias do trabalho, incluindo quaisquer edições ou revisões necessárias.            |
| Comentário                 | É composto de cinco subcampos: Nome do distribuidor, Afiliação do distribuidor, Abreviatura do nome distribuidor, URL do distribuidor e URL do logotipo do distribuidor |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Não se aplica.  |

**Nome do distribuidor**

|                            |  |
|----------------------------|--|
| Identificador              | Nome do distribuidor   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Fortemente recomendado   |
| Definição                  | Nome pessoal ou corporativo do distribuidor do <i>dataset</i> .  |
| Comentário                 | Nome pessoal deve ser descrito na forma direta, ou seja, prenome(s) seguido(s) de sobrenome(s).<br>Caso o distribuidor seja uma instituição, preencha com o nome por extenso, sem abreviações. |
| Exemplos                   | Nome pessoal:<br>José Silva<br>Nome corporativo:<br>Empresa Brasileira de Pesquisa Agropecuária  |
| Documento orientador       | Não se aplica.   |

**Afiliação do distribuidor**

|                            |  |
|----------------------------|--|
| Identificador              | Afiliação do distribuidor  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome da instituição à qual o distribuidor do <i>dataset</i> está vinculado.  |
| Comentário                 | Preencher com o nome completo da instituição. Não usar abreviações ou suprimir partes do nome.<br>Caso a instituição seja aquela indicada como o distribuidor do <i>dataset</i> , não é necessário preencher este campo. |
| Exemplos                   | Embrapa (nome síntese)<br>Empresa Brasileira de Pesquisa Agropecuária  |
| Documento orientador       | Não se aplica.   |

Continua...

### Abreviatura do nome do distribuidor

|                            |   |
|----------------------------|---|
| Identificador              | Abreviatura do nome do distribuidor   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Abreviatura do nome distribuidor do <i>dataset</i> .  |
| Comentário                 | Indicar a abreviatura ou sigla do nome do distribuidor quando se tratar de um nome corporativo. |
| Exemplos                   | ESALQ<br>LMB  |
| Documento orientador       | Não se aplica.  |

### URL do distribuidor

|                            |   |
|----------------------------|---|
| Identificador              | URL do distribuidor   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | URL válida de uma página na web do distribuidor do <i>dataset</i> .   |
| Comentário                 | Transcrever a URL válida da página web do distribuidor do <i>dataset</i> .  |
| Exemplos                   | URL da Empresa Brasileira de Pesquisa Agropecuária: <a href="https://www.embrapa.br/URL">https://www.embrapa.br/URL</a> do Laboratório Multiusuário de Bioinformática da Embrapa: <a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a> |
| Documento orientador       | Não se aplica.  |

### URL do logotipo do distribuidor

|                            |  |
|----------------------------|--|
| Identificador              | URL do logotipo do distribuidor  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | URL válida do logotipo do distribuidor do <i>dataset</i> .   |
| Comentário                 | Transcrever a URL válida da página web onde o logotipo do distribuidor pode ser visualizado.   |
| Exemplos                   | URL do logotipo da Embrapa:<br><a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a><br>URL do logotipo do Laboratório Multiusuário de Bioinformática da Embrapa:<br><a href="https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png">https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png</a> |
| Documento orientador       | Não se aplica.   |

### Data de distribuição

|                            |   |
|----------------------------|---|
| Identificador              | Data de distribuição  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Data em que o <i>dataset</i> foi disponibilizado para distribuição. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.   |
| Exemplo                    | 2020-01-15  |
| Documento orientador       | ISO 8601 Date and Time Format                                       |

Continua...

**Depositante\***

|                            |  |
|----------------------------|--|
| Identificador              | Depositante  |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Requerido<br><b>Observação:</b> Este campo é preenchido automaticamente pelo sistema, com o nome do usuário que fez o login e depositou o <i>dataset</i> .   |
| Definição                  | Nome pessoal ou corporativo do depositante do <i>dataset</i> .   |
| Comentário                 | Nome pessoal deve ser descrito na forma direta, ou seja, prenome(s) seguido(s) de sobrenome(s).<br>Caso o depositante seja um nome corporativo, preencha com nome da instituição, sem abreviações. |
| Exemplos                   | Nome pessoal:<br>José Silva<br>Nome corporativo:<br>Empresa Brasileira de Pesquisa Agropecuária  |
| Documento orientador       | Não se aplica.   |

**Data do depósito**

|                            |   |
|----------------------------|---|
| Identificador              | Data do depósito  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Data em que o <i>dataset</i> foi depositado no repositório. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.                                 |
| Exemplo                    | 2020-06-11  |
| Documento orientador       | ISO 8601 Date and Time Format.                              |

**Período da cobertura de dados**

|                            |   |
|----------------------------|---|
| Identificador              | Período da cobertura de dados   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Período de tempo ao qual os dados se referem ou período de tempo abrangido pelos dados. |
| Comentário                 | É composto por dois subcampos: Data de início e Data de término                         |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Não se aplica.  |

**Data de início**

|                            |   |
|----------------------------|---|
| Identificador              | Data de início                                    |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado                                       |
| Definição                  | Data de início do período de cobertura dos dados. |
| Exemplo                    | 2015-05-01  |
| Documento orientador       | ISO 8601 Date and Time Format.                    |

**Data de término**

|                            |  |
|----------------------------|--|
| Identificador              | Data de término                                    |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Data de término do período da cobertura dos dados. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.                        |
| Exemplo                    | 2015-06-22   |
| Documento orientador       | ISO 8601 Date and Time Format.                     |

### Período da coleta de dados

|                            |   |
|----------------------------|---|
| Identificador              | Período da coleta de dados                                      |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Período de tempo em que os dados foram coletados.               |
| Comentário                 | É composto por dois subcampos: Data de início e Data de término |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Não se aplica.  |

### Data de início

|                            |  |
|----------------------------|--|
| Identificador              | Data de início                                 |
| Ocorrência                 | Ilimitada                                      |
| Exigência de preenchimento | Recomendado                                    |
| Definição                  | Data de início do período da coleta dos dados. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.                    |
| Exemplo                    | 2015-05-01                                     |
| Documento orientador       | ISO 8601 Date and Time Format.                 |

### Data de término

|                            |  |
|----------------------------|--|
| Identificador              | Data de término                                |
| Ocorrência                 | Ilimitada                                      |
| Exigência de preenchimento | Recomendado                                    |
| Definição                  | Data de término do período da coleta de dados. |
| Comentário                 | Seguir o padrão AAAA-MM-DD.                    |
| Exemplo                    | 2015-06-22                                     |
| Documento orientador       | ISO 8601 Date and Time Format.                 |

### Tipo de dados

|                            |   |
|----------------------------|---|
| Identificador              | Tipos de dados  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Recomendado   |
| Definição                  | Tipos de dados do <i>dataset</i> , que podem ser identificados como: <i>survey</i> , dados de censo, dados agregados, dados clínicos, dados de eventos, dados de transações, código-fonte do programa, texto legível por máquina, dados de transações administrativas, dados experimentais, dados de teste psicológico, dados textuais, documentos codificados, dados/avaliações de observação, dados produzidos pelo processo etc. |
| Comentário                 | Texto livre.<br>Indique textualmente o tipo de dado que corresponde ao <i>dataset</i> . Pode ser inserido mais de um tipo.  |
| Exemplos                   | Dados de pesquisa<br>Dados experimentais<br>Dados biológicos<br>Dados genômicos<br>Dados de sequência RNA   |
| Documento orientador       | Não se aplica.  |

Continua...

**Série**

|                            |   |
|----------------------------|---|
| Identificador              | Série   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Informação sobre a série do <i>dataset</i> , caso exista.                               |
| Comentário                 | Texto livre.<br>É composto de dois subcampos: Nome da série e Informação sobre a série. |
| Exemplo                    | Não se aplica.  |
| Documento orientador       | Não se aplica.  |

**Nome da série**

|                            |   |
|----------------------------|---|
| Identificador              | Nome da série   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Nome de série ao qual o <i>dataset</i> está vinculado.    |
| Comentário                 | Texto livre.  |
| Exemplo                    | Portfólio de Sanidade Animal e Manejo de Pragas e Doenças |
| Documento orientador       | Não se aplica.  |

**Informação sobre série**

|                            |   |
|----------------------------|---|
| Identificador              | Informação sobre série  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Informação sucinta sobre a série à qual o <i>dataset</i> pertence.  |
| Comentário                 | Texto livre.  |
| Exemplo                    | A Série Portfólio Sanidade Animal e Manejo de Pragas e Doenças abarca o conjunto de <i>dataset</i> produzidos no âmbito do Projeto RmVAC. |
| Documento orientador       | Não se aplica.  |

**Software**

|                            |  |
|----------------------------|--|
| Identificador              | Software   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Informações sobre o software usado para gerar o <i>dataset</i> .                     |
| Comentário                 | Texto livre.<br>É composto de dois subcampos: Nome do software e Versão do software. |
| Exemplo                    | Não se aplica.   |
| Documento orientador       | Não se aplica.   |

**Nome do software**

|                            |  |
|----------------------------|--|
| Identificador              | Nome do software   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Nome do software utilizado para gerar o <i>dataset</i> . |
| Comentário                 | Texto livre.   |
| Exemplos                   | SAS<br><i>Scripts</i><br>Software R<br><i>Trinity</i>    |
| Documento orientador       | Não se aplica.   |

### Versão do software

|                            |  |
|----------------------------|--|
| Identificador              | Versão   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Recomendado  |
| Definição                  | Versão do software utilizado para gerar o <i>dataset</i> . |
| Comentário                 | Texto livre.   |
| Exemplos                   | Versão 2.8.4<br>Versão 0.01                                |
| Documento orientador       | Não se aplica.   |

### Material relacionado

|                            |  |
|----------------------------|--|
| Identificador              | Material relacionado   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | Outro material relacionado ao <i>dataset</i> que está sendo descrito.      |
| Comentário                 | Texto livre.   |
| Exemplo                    | Projetos anteriores e/ou outros materiais relacionados ao <i>dataset</i> . |
| Documento orientador       | Não se aplica  |

### Datasets relacionados

|                            |  |
|----------------------------|--|
| Identificador              | <i>Datasets</i> relacionados   |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | Quaisquer <i>datasets</i> relacionados ao <i>dataset</i> que está sendo descrito, tais como <i>datasets</i> provenientes de pesquisas anteriores sobre esse assunto. |
| Comentário                 | Texto livre.   |
| Exemplo                    | Não há   |
| Documento orientador       | Não se aplica.   |

### Outras referências

|                            |   |
|----------------------------|---|
| Identificador              | Outras referências  |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Quaisquer outras referências que podem ter servido de material de apoio ao <i>dataset</i> .                 |
| Comentário                 | Texto livre.  |
| Exemplo                    | Acesso e descoberta de conhecimento produzido por pesquisadores de instituições parceiras do Projeto RmVAC. |
| Documento orientador       | Não se aplica.  |

### Fontes de dados

|                            |   |
|----------------------------|---|
| Identificador              | Fontes de dados   |
| Ocorrência                 | Ilimitada   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Referem-se a materiais que serviram de fontes para o <i>dataset</i> , como: lista de livros, artigos, publicações em série ou arquivos de dados legíveis por máquina.   |
| Comentário                 | Texto livre.  |
| Exemplo                    | GIACHETTO, P. F.; CUNHA, R. C.; NHANI JUNIOR, A.; GARCIA, M. V.; FERRO, J. A.; ANDREOTTI, R. Gene expression in the salivary gland of <i>Rhipicephalus (Boophilus) microplus</i> fed on tick-susceptible and tick-resistant hosts. <b>Frontiers in Cellular and Infection Microbiology</b> , v. 9, p. 1-15, Jan. 2020. DOI: 10.3389/fcimb.2019.00477. |

|                      |   |
|----------------------|---|
| Documento orientador | ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. <b>NBR 6023</b> : informação e documentação – referências: elaboração. 2. ed. Rio de Janeiro, 2020. 68 p. Versão corrigida 2. |
|----------------------|---|

### Origem das fontes de dados

|                            |  |
|----------------------------|--|
| Identificador              | Origem das fontes de dados   |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Opcional   |
| Definição                  | Para materiais históricos, informar a origem das fontes e as regras seguidas no estabelecimento das fontes do <i>dataset</i> . |
| Comentário                 | Texto livre.   |
| Exemplo                    | Literatura que retrata o estado da arte.   |
| Documento orientador       | Não se aplica  |

### Características das fontes de dados observadas

|                            |  |
|----------------------------|--|
| Identificador              | Características das fontes de dados  |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Opcional   |
| Definição                  | Avaliação das características e material de origem do <i>dataset</i> . Descrever aspecto digno de nota para o dado coletado. |
| Comentário                 | Texto livre  |
| Exemplo                    | Não há   |
| Documento orientador       | Não se aplica.   |

### Documentação e acesso a fontes de dados

|                            |  |
|----------------------------|--|
| Identificador              | Documentação e acesso a fontes de dados  |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Opcional   |
| Definição                  | Nível de acesso à documentação do material de origem. Informar quaisquer restrições de acesso aos dados de origem. |
| Comentário                 | Texto livre.   |
| Exemplo                    | Não há.  |
| Documento orientador       | Não se aplica.   |

Fonte: Adaptado de Souza et al. (2020).

## Descrição dos Elementos do Esquema Metadados Ciência da Vida

Para atender diretamente ao domínio de dados biológicos, o software Dataverse oferece também o esquema específico de metadados - o *Life Science Metadata* -, aqui denominado esquema Metadados Ciência da Vida. Este esquema abarca as especificidades e os interesses da comunidade de dados biológicos, tornando-o apropriado para descrever *datasets* ômicos.

Desse modo, os esquemas Metadados de Citação e Metadados Ciência da Vida podem ser utilizados concomitantemente. Essa recomendação está apoiada nos estudos realizados no LMB para a escolha de esquemas para catalogar dados ômicos. A utilização simultânea dos dois esquemas também está condicionada à estruturação do *dataverse*<sup>16</sup> ao qual o *dataset* estiver vinculado (o *dataverse* quando criado deve prever o uso de ambos os esquemas).

<sup>16</sup> *Dataverse* é aqui entendido como um projeto que reúne conjuntos de dados (*datasets*).

O esquema de Metadados Ciência da Vida é formado por nove elementos de preenchimento opcional, listados a seguir na Tabela 8.

**Tabela 8.** Elementos do esquema Metadados Ciência da Vida do formulário de catalogação do Redape.

| Elemento metadado      | Exigência |
|------------------------|-----------|
| Tipo de design         | Opcional  |
| Tipo de fator          | Opcional  |
| Organismo              | Opcional  |
| Outro organismo        | Opcional  |
| Tipo de medida         | Opcional  |
| Outro tipo de medida   | Opcional  |
| Tipo de tecnologia     | Opcional  |
| Plataforma tecnológica | Opcional  |
| Tipo de célula         | Opcional  |

Fonte: Adaptado de Souza et al. (2020).

Para a catalogação de *datasets* de acordo com os elementos do esquema Metadados Ciência da Vida foi adotado um conjunto de atributos comuns que orientam a descrição de todos os seus elementos (campos e subcampos), conforme mostra a Tabela 9.

**Tabela 9.** Atributos orientadores para a descrição dos elementos do esquema Metadados Ciência da Vida do Redape.

| Nome                       | O rótulo atribuído para o elemento de dado.   |
|----------------------------|---|
| Identificador              | O único identificador atribuído para o elemento de dado.  |
| Ocorrência                 | Indica qualquer limite para a repetibilidade do elemento de dado.                               |
| Exigência de preenchimento | Indica se o elemento de dado necessita estar presente sempre ou esporadicamente.                |
| Definição                  | Uma declaração que claramente representa o conceito e a natureza essencial do elemento de dado. |
| Comentário                 | Um comentário relativo à aplicação do elemento de dado.   |
| Exemplo                    | Um exemplo relativo à aplicação do elemento de dado.  |
| Documento orientador       | Um documento que oriente a aplicação do elemento de dado.                                       |

Fonte: Souza et al. (2020).

A seguir são apresentados os nove elementos do esquema Metadados Ciência da Vida e respectivas descrições para a catalogação de *datasets* (Tabela 10).

**Tabela 10.** Descrição dos elementos do esquema Metadados Ciência da Vida no Redape.

| Tipo de design             |  |
|----------------------------|--|
| Identificador              | Tipo de design   |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Opcional   |
| Definição                  | Definição geral do desenho experimental.   |
| Comentário                 | Selecionar um ou mais tipos de design.   |
| Exemplos                   | Case control; Cross sectional; Cohort study; Nested case control design; Not specified; Parallel group design; Perturbation design; Randomized controlled trial; Technological design. |
| Documento orientador       | Não se aplica.   |

Continua...

**Tipo de fator**

|                            |   |
|----------------------------|---|
| Identificador              | Tipo de fator   |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Fatores utilizados no <i>dataset</i> . Fator experimental; estresse hídrico; estágio de desenvolvimento; tipo de célula.      |
| Comentário                 | Selecionar um ou mais tipos de fator.   |
| Exemplos                   | Age; Biomarkers; Cell surface markers; Cell type/cell line; Developmental stage; Genetic characteristics; Treatment type etc. |
| Documento orientador       | Não se aplica.  |

**Organismo**

|                            |   |
|----------------------------|---|
| Identificador              | Organismo   |
| Exigência de preenchimento | Opcional  |
| Definição                  | Nome taxonômico do organismo usado nos <i>datasets</i> ou de onde o material biológico se originou.               |
| Comentário                 | Selecionar um ou mais organismos. Para indicar organismos não constantes na lista apresentada, selecione 'Other'. |
| Exemplos                   | <i>Bos taurus</i><br><i>Escherichia coli</i><br><i>Rhipicephalus microplus</i><br>Carrapato do boi                |
| Documento orientador       | Não se aplica.  |

**Outro organismo**

Não se aplica

**Tipo de medida**

|                            |   |
|----------------------------|---|
| Identificador              | Tipo de medida  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Um termo para qualificar o que está sendo medido.   |
| Comentário                 | Selecionar um ou mais tipos de medida. Para indicar algum tipo de medida não constante na lista apresentada, selecione 'Other'. |
| Exemplos                   | Cell counting; DNA methylation profiling; Genome sequencing; Protein=RNA binding (RIP-Seq) etc.                                 |
| Documento orientador       | Não se aplica.  |

**Outro tipo de medida**

Não se aplica

**Tipo de tecnologia**

|                            |   |
|----------------------------|---|
| Identificador              | Tipo de tecnologia  |
| Ocorrência                 | Limitada  |
| Exigência de preenchimento | Opcional  |
| Definição                  | Um termo usado para identificar a tecnologia usada para coletar os dados.   |
| Comentário                 | Selecionar um ou mais tipos de tecnologia. Para indicar algum tipo de tecnologia não constante na lista apresentada, selecione 'Other'. |
| Exemplos                   | DNA microarray; Mass spectrometry; Nucleotide sequencing; Protein microarray etc.   |
| Documento orientador       | Não se aplica.  |

Continua...

### Plataforma tecnológica

|                            |  |
|----------------------------|--|
| Identificador              | Plataforma tecnológica   |
| Ocorrência                 | Limitada   |
| Exigência de preenchimento | Opcional   |
| Definição                  | O fabricante e o nome da plataforma usada no experimento para coleta de dados.   |
| Comentário                 | Selecionar um ou mais plataformas tecnológicas. Para indicar plataforma tecnológica não constante na lista apresentada, selecione 'Other'. |
| Exemplos                   | Illumina<br>HiSeq2500  |
| Documento orientador       | Não se aplica.   |

### Tipo de célula

|                            |  |
|----------------------------|--|
| Identificador              | Linhagem de células  |
| Ocorrência                 | Ilimitada  |
| Exigência de preenchimento | Opcional   |
| Definição                  | Nome da linhagem de células das quais as amostras derivam.   |
| Comentário                 | Texto livre.<br>Indicar os nomes dos tipos de célula presentes nos tecidos, por exemplo, do carrapato. |
| Exemplos                   | Larvas; Ninfas; Ovários; Intestinos; Glândula salivar etc.   |
| Documento orientador       | Não se aplica.   |

Fonte: Souza et al. (2020).

## Referências

ANTONELLI, L.; GUARRACINO, M. R.; MADDALENA, L.; SANGIOVANNI, M. Integrating imaging and omics data: a review. **Biomedical Signal Processing and Control**, Oxford, v. 52, p. 264-280, July 2019. DOI: 10.1016/j.bspc.2019.04.032.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação – referências: elaboração. 2. ed. Rio de Janeiro, 2020. 68 p. Versão corrigida 2.

BRASILEIRO, N. T. R. S. **Uma abordagem para geração de perfil de conjunto de dados com metadados enriquecidos semanticamente**. 2018. 119 p. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/31430>>. Acesso em: 11 out. 2021.

CESAR JUNIOR, R. M. Apresentação à edição brasileira. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Org.). **O quarto paradigma**: descobertas científicas na era da *eScience*. São Paulo: Oficina de Textos, 2011. p. 7.

CHAN, L.; OKUNE, A.; SAMBULI, N. O que é ciência aberta e colaborativa, e que papéis ela poderia desempenhar no desenvolvimento? In: ALBABLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília, DF: Ibict; Rio de Janeiro: Unirio, 2015. cap. 5, p. 91-119. Disponível em: <[http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta\\_questoes%20abertas\\_PORTUGUES\\_DIGITAL%20\(5\).pdf](http://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf)>. Acesso em: 28 ago. 2020.

CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004. 1 v. Paginação irregular.

CURTY, R. Abordagens de reuso e a questão de reusabilidade dos dados científicos. **Liinc em Revista**, v. 15, n. 2, p. 177-193, nov. 2019. DOI: 10.18617/liinc.v15i2.477.

DATAVERSE. **Harvard Dataverse**. [Cambridge, MA: Harvard College, 2020]. Disponível em: <<https://dataverse.harvard.edu>>. Acesso em: 2 maio 2020.

DATAVERSE + file management. In: DATAVERSE PROJECT. **User guide**. [Cambridge: Harvard College], 2016. Disponível em: <<https://guides.dataverse.org/en/latest/user/dataset-management.html#>>. Acesso em: 31 ago. 2021.

EMBRAPA. **Redape (Repositório de Dados de Pesquisa da Embrapa)**. Brasília, DF, 2021. Disponível em: <<https://www.dataverse-h.cnptia.embrapa.br/>>. Acesso em: 3 dez. 2021.

EMBRAPA AGRICULTURA DIGITAL. **Laboratório Multiusuário de Bioinformática**. Disponível em: <<https://www.embrapa.br/agricultura-digital/lmb>>. Acesso em: 17 dez. 2021.

GO FAIR. **FAIR principles**. [2018]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 27 ago. 2020.

GORDANA, R.; DRAGAN, I. Cataloguing dataset in library information systems using the MARC 21 format. In: INTERNATIONAL CONFERENCE ON INFORMATION SOCIETY AND TECHNOLOGY, 7., 2017, Kopaonik. **Proceedings...** Belgrade: Society for Information Systems and Computer Networks Belgrade, 2017. p. 395-399. Disponível em: <<https://pdfs.semanticscholar.org/407a/9d09ecb1261db980ef03b8025fe5dc1939d5.pdf>>. Acesso em: 28 ago. 2020.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, J. L. R.; SANTOS, L. O. B. da S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *Pesquisa Brasileira em Ciência da Informação e Biblioteconomia*, v. 14, n. 3, p. 175-192, 2019. DOI: 10.22478/ufpb.1981-0695.2019v14n3.46969.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília, DF: Briquet de Lemos, 2004. 452 p.

LE COADIC, Y. F. **A ciência da informação**. Brasília, DF: Briquet de Lemos, 1996. 119 p.

MEY, E. S. A.; SILVEIRA, N. C. **Catálogo no plural**. Brasília, DF: Briquet de Lemos, 2009. 217 p.

ORTEGA, C. D. Relações históricas entre Biblioteconomia, Documentação e Ciência da Informação. **DataGramZero**, v. 5, n. 5, 2004. Não paginado.

RIBEIRO, A. M. de C. M. **AACR2**: Anglo-American cataloguing rules, 2nd edition: descrição e pontos de acesso. Brasília, DF: Ed. do Autor, 1995. 577 p.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pela AACR2 2002**: Anglo-American Cataloguing Rules, 2<sup>nd</sup> edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

ROCHA, R. P. da; AZAMBUJA, L. A. B.; BORGES, N. B.; GABRIEL JUNIOR, R. F.; CAREGNATO, S. E.; PAVÃO, C. G.; VANZ, S. A. de; PASSOS, P. C. S. J. **Acesso aberto a dados de pesquisa no Brasil**: soluções tecnológicas: relatório 2018. Porto Alegre: UFRGS, 2018. 74 p. Disponível em: <<https://lume.ufrgs.br/bitstream/handle/10183/185126/001082284.pdf?sequence=1&isAllowed=y>>. Acesso em: 2 set. 2021.

ROCHA, R. P. da; GABRIEL JUNIOR, R. F.; VANZ, S. A. de S.; BORGES, E. N.; AZAMBUJA, L. A. B.; CAREGNATO, S. E.; PAVÃO, C. G.; PASSOS, P. C. S. J.; FELICISSIMO, C. H. Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto. **Revista Brasileira de Biblioteconomia e Documentação**, v. 17, p. 1-25, 2021. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/160963>>. Acesso em: 1 set. 2021.

RUSSO, M. **Fundamentos de Biblioteconomia e Ciência da Informação**. Rio de Janeiro: e-papers, 2010. 177 p. (Coleção biblioteconomia e gestão de unidades de informação. Série didáticos, n. 1).

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa**: [minicurso]. [Rio de Janeiro: CNEN, 2016]. 196 slides.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, v. 48, n. 1, p. 113-129, jan./abr. 2019.

SOUZA, M. I. F.; VISOLI, M. C.; TORRES, T. Z. **Catálogo de dataset no Repositório de Dados da Embrapa**: a experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa. Campinas: Embrapa Informática Agropecuária, 2020. 116 p. (Embrapa Informática Agropecuária. Documentos, 172). Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1127941/1/Doc172-2020.pdf>>. Acesso em 23 ago. 2021.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J. W.; SANTOS, L. B. da S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.;

LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAIK, R. van; SANSONE, S. A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; VAN DER LEI, J.; VAN MULLIGEN, E.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 13, p. 1-8, Mar. 2016. DOI: 10.1038/sdata.2016.18.

## APÊNDICE A - Vocabulários Controlados

### Introdução

Neste Apêndice há orientações para o uso de vocabulários controlados no Repositório de Dados de Pesquisa da Embrapa (Redape), oriundas da experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB), no escopo do projeto de pesquisa “Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus microplus* e espécies relacionadas” (Projeto RmVAC).

Vocabulário controlado consiste em uma lista de termos elaborada por especialistas de diversos domínios para indexação de assunto de diferentes objetos de informação, dentre os quais, os *datasets*. Vocabulário controlado é comumente definido como uma lista autorizada de assuntos sobre determinado domínio. (Souza et al., 2020). Cunha e Cavalcanti (2008, p. 378) assim definem vocabulários controlados:

Conjunto de termos que, nos sistemas de informação, devem ser empregados tanto no momento da indexação como no da recuperação. A finalidade principal desse controle é fazer coincidir a linguagem do pesquisador com a do indexador. Nos vocabulários controlados são feitas remissivas dos sinônimos e quase-sinônimos para o termo selecionado como descritor.

Um de seus objetivos é compatibilizar a linguagem utilizada por pesquisadores à linguagem adotada em sistemas de informação. Em processos de catalogação, os vocabulários controlados são utilizados para representar, por meio de termos ou palavras-chave, o conteúdo do recurso de informação que está sendo catalogado. (Souza et al., 2020).

A adoção de um vocabulário controlado para descrever e classificar assuntos é uma medida necessária para prover serviços de busca e de recuperação de dados e informações com maior garantia de qualidade (Souza et al., 2020). No contexto deste trabalho, o uso de vocabulário controlado tem a função de:

- Atribuir palavras-chave e classificação de assuntos na catalogação de *datasets*.
- Indicar relacionamentos e associações hierárquicas e de equivalência entre termos atribuídos pelo catalogador a palavras-chave e classificação.

- Eliminar polissemia e ambiguidade na atribuição de termos nos campos Palavras-chave e Classificação.
- Possibilitar a identificação e a recuperação de *datasets* descritos sob determinados termos.
- Explorar a riqueza terminológica existente nos diferentes vocabulários, sob a perspectiva da análise semântica e da mineração de textos.
- Oferecer possibilidade de análise conceitual dos termos em diferentes idiomas, subsidiando a descrição de assunto e de classificação de *datasets*.

Vocabulários controlados são altamente recomendados para a representação temática de *datasets*, sobretudo quando são compatíveis com os princípios FAIR, especialmente o princípio *Interoperable* (Interoperabilidade) (Henning et al., 2019; Veiga, 2019). Metadados bem descritos, com vocabulários controlados, e indexados em catálogos e repositórios de dados são facilmente ‘encontráveis’ por mecanismos de busca, de modo que tanto os dados como os metadados estarão ao alcance dos usuários.

Além disso, metadados descritos com vocabulários controlados permitem a representação de assuntos e de classificação de forma padronizada, garantindo, assim, a acessibilidade e a interoperabilidade semântica (Veiga, 2019).

Nos campos e subcampos do formulário de catalogação do software Dataverse, é possível o uso de termos de vocabulário controlado em **Palavras-chave** e também em **Classificação**. Em **Palavras-chave**, recomenda-se usar o termo mais específico possível para classificar o *dataset*. Em **Classificação**, recomenda-se utilizar um termo mais abrangente.

Os principais vocabulários controlados estão disponíveis na Internet com acesso gratuito, consulta e uso on-line. No entanto, o uso de vocabulário controlado para identificar um termo ou classificação de assunto quase sempre demanda o auxílio de profissional da área de Biblioteconomia ou Ciência da Informação. Esta ajuda poderá ocorrer durante a catalogação e também na revisão, que ocorre antes da publicação dos *datasets*. (Souza et al., 2020).

Algumas características e boas práticas para a escolha de vocabulários controlados são indicadas pela *The World Wide Web Consortium* (W3C), organização que é referência no estabelecimento de padrões

para a Internet, tais como mencionadas a seguir (The World Wide Web Consortium, 2012, tradução nossa):

- Escolher vocabulários que preferencialmente tenham também termos em inglês.
- Escolher vocabulários que já são utilizados por outras instituições nas áreas de domínio correlatas.
- Escolher vocabulários autoexplicativos.
- Escolher vocabulários acessíveis por um período longo.
- Identificar o mantenedor e qual infraestrutura está disponível para sustentar os vocabulários escolhidos.
- Escolher vocabulários que sejam publicados por instituições ou grupos confiáveis.
- Verificar se as instituições que mantêm os vocabulários escolhidos possuem esta atribuição definida formalmente.
- Escolher vocabulários que sejam estruturados com URI permanentes.
- Escolher vocabulários que proveem políticas de versionamento para novas versões, bem como para a manutenção versões anteriores.

A seguir estão apresentadas as orientações para o uso dos vocabulários controlados *AGROVOC Thesaurus*, *NAL Thesaurus* e *NCBI Taxonomy*, identificados como apropriados à catalogação de *datasets* (aderentes aos princípios FAIR: encontráveis, interoperáveis, acessíveis e reusáveis), junto à equipe do Projeto RmVAC e do Projeto-piloto de Implantação de Processos de Gestão de Dados de Pesquisa no LMB.

Tais orientações, a seguir descritas, são destinadas especificamente para o preenchimento do elemento **Palavras-chave** nos campos **Termo-chave**, **Vocabulário controlado do termo-chave** e **URL do vocabulário controlado**; bem como do elemento **Classificação** nos campos **Termo de classificação**, **Vocabulário controlado da classificação** e **URL do vocabulário controlado**.

## **AGROVOC *Thesaurus***

### **O que é AGROVOC *Thesaurus*?**

AGROVOC *Thesaurus* é um vocabulário controlado desenvolvido e mantido pela *Food and Agriculture Organization of the United Nations* (FAO), desde a década de 1980. Contém mais de 39 mil conceitos organizados de forma hierárquica, associativa e equivalente em 41 idiomas, incluindo o português europeu. Sua cobertura abrange extensa área de conhecimento como: alimentação, nutrição, agricultura, florestas, pesca, nomes científicos e comuns de animais e plantas, meio ambiente, aspectos biológicos e técnicas de cultivo de plantas e de criação de animais. (FAO, 2021).

AGROVOC *Thesaurus* é amplamente usado por bibliotecas e repositórios digitais para indexar conteúdos, bem como para fins de organização de conhecimento. No âmbito do Repositório de Dados de Pesquisa da Embrapa, o AGROVOC *Thesaurus* foi um dos vocabulários controlados escolhidos para descrever assunto (palavras-chave e classificação) de *datasets*.

### **AGROVOC *Thesaurus* para Palavras-chave**

A atribuição de palavras-chave é uma etapa da catalogação de *datasets* no Redape, especificamente no preenchimento dos campos do Elemento **Palavras-chave**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e preenchimento dos campos **Termo-chave**, **Vocabulário controlado do termo-chave** e **URL do vocabulário controlado**.

Segue passo a passo:

- **Etapa 1. Acesso ao site do AGROVOC *Thesaurus*:** acesse o endereço: <http://aims.fao.org/standards/agrovoc/functionalities/search>. A página é apresentada como mostra a Figura 1.

## Search

SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for "regis" you will find, for example "registration" and "registered designation of origin", but not "deed registration". If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results.



AGROVOC Multilingual Thesaurus
Content language: English x Search

Alphabetical Hierarchy

A B C Ç D E F G H I J K  
L M N O P Q R S Ş T U V  
W X Y Z 0-9

A horizons  
Aaptosyax grypus  
Aaron's rod → Verbascum  
ABA  
Abaca  
abachi → Triplochiton scleroxylon  
Abalistes stellaris  
abalone culture  
abalone fisheries → gastropod fisheries  
abalones  
abamectin  
abandoned land  
abattoir byproducts  
abattoirs  
Abbottina rivularis  
abdomen  
abdominal cavity

### Vocabulary information

**TITLE** AGROVOC Multilingual Thesaurus

**LAST MODIFIED** Wednesday, July 8, 2020 09:42:42

**TYPE** <http://www.w3.org/2004/02/skos/core#ConceptScheme>

**VOID:INDATASET** <http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc>

**URI** <http://aims.fao.org/aos/agrovoc>

**Resource counts by type**

| Type    | Count |
|---------|-------|
| Concept | 37436 |

**Term counts by language**

Language: English Preferred terms: Alternata terms: Hidden terms:

**Figura 1.** Tela de busca do AGROVOC *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 2. Realização da busca:** no menu à esquerda clique em 'Alphabetical', e em seguida no campo 'Search' digite o termo desejado, por exemplo 'nucleotide sequence', e execute a busca clicando em 'Search'. Após a realização da busca, serão apresentados os resultados, como mostra a Figura 2.

## Search

SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for "regis" you will find, for example "registration" and "registered designation of origin", but not "deed registration". If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results.


 Food and Agriculture Organization  
of the United Nations

Vocabularies About Feedback Help

AGROVOC Multilingual Thesaurus

Content language

English

nucleotide sequence

Search

 Search  
options

By subvocabulary

By parent

Limit search

2 results for 'nucleotide sequence'

**nucleotide sequence** (en)

↳ genetic structures (en)

↳ codons (en), DNA sequence (en), open reading frames (en), regulatory sequenc... (5)

↳ multilocus sequence typing (en)

↳ gene sequence, nucleic sequence

 ↳ تتابع النيوكليوتيد (ar), 核苷酸序列 (zh), 基因序列 (zh), 核酸序列 (zh), **sekvence nukleotidú**

 (cs), *sekvence genů* (cs), *nukleová sekvence* (cs), **Séquence nucléotidique** (fr),

*Séquence génique* (fr), *Séquence de protéines* (fr), *Séquence nucléique* (fr),

**Nucleotidsequenz** (de), *Gensequenz* (de), *Nukleosequenz* (de), न्यूक्लियोटाइड अनुक्रम

 (hi), न्यूक्लिनक क्रममाक (hi), जीन क्रम (hi), **nukleotid szekvencia** (hu), *génszekvencia* (hu),

*nukleinszekvencia* (hu), **Sequenza nucleotidica** (it), *Sequenza genica* (it), *Sequenza*
*nucleica* (it), **ヌクレオチド配列**, **核酸配列** (ja), **ヌクレオチド配列** (ja), **遺伝子配列**

 (ja), **မာဲတီပေဝေ့ပီဝဲဝဲမီဝဲမီဝဲ** (lo), **မာဲတီပေဝေ့ပီဝဲမီဝဲမီဝဲ** (lo), **မာဲတီပေဝေ့ပီဝဲ** (lo),

**nukleotidsekvenser** (nb), **توالى نوكلوتيد** (fa), **توالى هسته‌اي** (fa), **توالى ژن** (fa), **Sekwencja**
**nukleotydów** (pl), **Sekwencja nukleinowa** (pl), **Sekwencja genów** (pl), **Sequência**
**nucleotídica** (pt), **Sequência gênica** (pt), **Sequência nucleica** (pt), **нуклеотидная**

**Figura 2.** Tela de resultado de busca do AGROVOC *Thesaurus* para o termo 'nucleotide sequence'.

Fonte: Souza et al. (2020).

- **Etapas 3. Acesso a página detalhada do termo-chave:** clique no termo 'nucleotide sequence' no resultado da busca. A Figura 3 apresenta os detalhes do termo pesquisado.

The screenshot shows the AIMS website interface. At the top, there is a navigation menu with options: GET INVOLVED, ACCESS TO INFORMATION, AGROVOC, SEMANTICS, ACTIVITIES, THEMES, and ABOUT. Below this is a search bar with the text 'Search'. The main content area displays the 'AGROVOC Multilingual Thesaurus' with a search bar set to 'English'. The search results for 'nucleotide sequence' are shown, including a list of related terms on the left and detailed information on the right. The detailed information includes the preferred term 'nucleotide sequence', its definition, broader concept 'genetic structures (en)', narrower concepts 'codons (en)', 'DNA sequence (en)', 'open reading frames (en)', 'regulatory sequences (en)', and 'RNA sequence (en)', and related concepts 'multilocus sequence typing (en)'. Entry terms include 'gene sequence (en)' and 'nucleic sequence (en)'.

**Figura 3.** Tela do termo do 'nucleotide sequence' contendo informações detalhadas.

Fonte: Souza et al. (2020).

- **Etapa 4. Obtenção da URL do termo-chave:** na página com os detalhes do termo, apresentado na Figura 3, deslize o mouse para o fim da tela, e procure o item URI, como mostrado na Figura 4. Copie o link apresentado, ou então, clique no ícone  para copiar e levar o endereço para o formulário de catalogação no campo **URL do vocabulário controlado**. **Observação:** No formulário Dataverse não há campo destinado para a descrição da **URL do termo-chave**, por esta razão, escolheu-se o campo **URL do vocabulário controlado** para fazer o seu registro.

The screenshot shows the search results for 'nucleotide sequence' in the AGROVOC Thesaurus. The term is displayed in Ukrainian: 'послідовність'. Below the term, the URI is shown: [http://aims.fao.org/aos/agrovoc/c\\_27583](http://aims.fao.org/aos/agrovoc/c_27583). There are also options to download the concept in various formats: RDF/XML, Turtle, and JSON-LD. The page also shows the creation and modification dates: 'Created 11/20/11, last modified 5/21/20'. At the bottom, there is a section for 'EXACTLY MATCHING CONCEPTS' with a link to <http://d-nb.info>.

**Figura 4.** Resultado da busca do termo 'nucleotide sequence', no AGROVOC *Thesaurus*, com destaque para o item URI.

Fonte: Souza et al. (2020).

- **Etapa 5. Preenchimento dos campos do elemento Palavras-chave:** no formulário de catalogação de *datasets* no Redape, preencha os campos conforme o exemplo:

Termo-chave: nucleotide sequence

Vocabulário controlado do termo-chave: AGROVOC *Thesaurus*

URL do vocabulário controlado: [http://aims.fao.org/aos/agrovoc/c\\_27583](http://aims.fao.org/aos/agrovoc/c_27583).

## **AGROVOC *Thesaurus* para Classificação**

A classificação do assunto de *datasets* é feita por meio da atribuição de uma ou mais categorias de assunto. Tais categorias representam estruturas de assunto mais abrangentes sob as quais um assunto específico, como por exemplo, 'nucleotide sequence', está inserido. A atribuição de classificação de assunto, por meio de categorias, é uma etapa da catalogação de *datasets*, especificamente no preenchimento do campo **Termo de classificação**, do Elemento **Classificação**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e preenchimento dos campos **Termo de classificação**, **Vocabulário controlado da classificação** e **URL do vocabulário controlado**. Segue o passo a passo:

- **Etapa 1. Acesso ao site do AGROVOC *Thesaurus*:** acesse o endereço: <http://aims.fao.org/standards/agrovoc/functionalities/search>. A página é apresentada como mostra a Figura 5.

AIMS

SEARCH CONNECT SIGN UP LOGIN

GET INVOLVED ACCESS TO INFORMATION AGROVOC SEMANTICS ACTIVITIES THEMES ABOUT

## Search

SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for "regis" you will find, for example "registration" and "registered designation of origin", but not "deed registration". If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results.

Food and Agriculture Organization of the United Nations

Vocabularies About Feedback Help

AGROVOC Multilingual Thesaurus Content language English Search

Alphabetical Hierarchy

A B C Ç D E F G H I J K  
L M N O P Q R S Ş T U V  
W X Y Z 0-9

A horizons  
Aaptosyax grypus  
Aaron's rod → Verbascum  
ABA  
Abaca  
abachi → Triplochiton scleroxylon  
Abalistes stellaris  
abalone culture  
abalone fisheries → gastropod fisheries  
abalones  
abamectin  
abandoned land  
abattoir byproducts  
abattoirs  
Abbottina rivularis  
abdomen  
abdominal cavity

### Vocabulary information

TITLE AGROVOC Multilingual Thesaurus

LAST MODIFIED Wednesday, July 8, 2020 09:42:42

TYPE <http://www.w3.org/2004/02/skos/core#ConceptScheme>

VOID:INDATASET <http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc>

URI <http://aims.fao.org/aos/agrovoc>

#### Resource counts by type

| Type    | Count |
|---------|-------|
| Concept | 37436 |

#### Term counts by language

Language Preferred terms Alternate terms Hidden terms

Figura 5. Tela de busca do AGROVOC *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 2. Realização da busca:** no menu à esquerda clique em 'Hierarchy', em seguida no campo 'Search' digite o termo desejado para o qual necessita fazer a classificação de assunto do *dataset*, por exemplo, 'nucleotide sequence', e execute a busca clicando em 'Search'. Após a realização da busca, serão apresentados os resultados, como mostra a Figura 6.

GET INVOLVED ACCESS TO INFORMATION AGROVOC SEMANTICS ACTIVITIES THEMES ABOUT

## Search

SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for "regis" you will find, for example "registration" and "registered designation of origin", but not "deed registration". If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results.

Food and Agriculture Organization of the United Nations

Vocabularies About Feedback Help

AGROVOC Multilingual Thesaurus Content language English Search

Alphabetical Hierarchy

- adaptation
- additive effect
- aerobiosis
- anaerobiosis
- behaviour
- bioavailability
- biological activity in soil
- biological development
- biological differences
- biological interaction
- biological rhythms
- cell cycle
- death
- diseases
- disorders
- genetic phenomena
- allele frequency
- gene frequency

phenomena > biological phenomena > genetic  
phenomena > genetic structures > nucleotide sequence

PREFERRED TERM **nucleotide sequence**

DEFINITION It is the order (or sequence) of alignment of nucleotides of nucleic acid molecules. (en)

BROADER CONCEPT genetic structures (en)

NARROWER CONCEPTS codons (en)  
DNA sequence (en)  
open reading frames (en)  
regulatory sequences (en)  
RNA sequence (en)

RELATED CONCEPTS multilocus sequence typing (en)

ENTRY TERMS

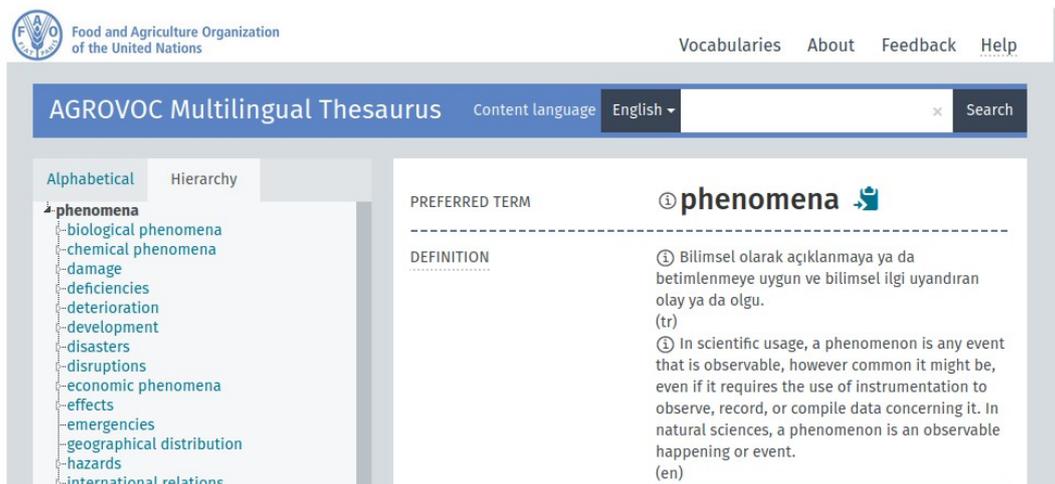
**Figura 6.** Tela de resultado de busca do AGROVOC *Thesaurus* para o termo ‘nucleotide sequence’.

Fonte: Souza et al. (2020).

- **Etapa 3. Busca das categorias:** na tela de resultados apresentada na Figura 6, observe logo acima do termo ‘nucleotide sequence’, uma sequência de termos que representam as suas categorias hierárquicas, sendo elas: ‘phenomena’ > ‘biological phenomena’ > ‘genetic phenomena’ > ‘genetic structures’ > ‘nucleotide sequence’. Essas categorias são as classificações de assunto encontradas para o termo ‘nucleotide sequence’. Caberá ao especialista do domínio orientar a escolha da(s) categoria(s) que melhor classifica(m) o assunto do *dataset*. Ao escolher a categoria ‘phenomena’, por exemplo, clique neste mesmo termo para abrir sua página, como mostra a Figura 7.

## Search

SKOSMOS searches only in concepts having a label starting with the input value. So, if you search for "regis" you will find, for example "registration" and "registered designation of origin", but not "deed registration". If you want to search for all concepts having a label containing such input value, you need to put \* before and after, so, for example, you can input \*regis\* and you will get many more results.



The screenshot shows the AGROVOC Multilingual Thesaurus interface. At the top, there is a search bar with the text "AGROVOC Multilingual Thesaurus" and a dropdown menu for "Content language" set to "English". Below the search bar, there are two tabs: "Alphabetical" and "Hierarchy". The "Alphabetical" tab is selected, and a list of terms is shown, with "phenomena" highlighted. To the right of the list, the details for the term "phenomena" are displayed. The details include the "PREFERRED TERM" (phenomena) and the "DEFINITION" in two languages: Turkish (tr) and English (en). The Turkish definition is "Bilimsel olarak açıklanmaya ya da betimlenmeye uygun ve bilimsel ilgi uyandıran olay ya da olgu." and the English definition is "In scientific usage, a phenomenon is any event that is observable, however common it might be, even if it requires the use of instrumentation to observe, record, or compile data concerning it. In natural sciences, a phenomenon is an observable happening or event."

Figura 7. Tela com os detalhes do termo 'phenomena'.

Fonte: Souza et al. (2020).

- Etapa 4. Obtenção da URL para o termo de classificação:** na página com os detalhes do termo, apresentado na Figura 7, deslize o mouse para o fim da tela, e procure o item URI, como mostrado na Figura 8. Copie o link apresentado, ou então clique ícone  para copiar e levar o endereço para o formulário de catalogação. **Observação:** No formulário do Dataverse não há campo destinado para a descrição da URL do termo de classificação. Por esta razão, escolheu-se o campo URL do vocabulário controlado para fazer o seu registro.

URI

[http://aims.fao.org/aos/agrovoc/c\\_330704](http://aims.fao.org/aos/agrovoc/c_330704) 

Download this concept:

[RDF/XML](#) [TURTLE](#) [JSON-LD](#)

Figura 8. Resultado da busca do termo 'phenomena', no AGROVOC *Thesaurus*, com destaque para o item URI.

Fonte: Souza et al. (2020).

- Etapa 5. Preenchimento dos campos do elemento Classificação:** no formulário de catalogação de *dataset* no Redape, preencha os campos conforme o exemplo:

Termo de classificação: phenomena

Vocabulário controlado da classificação: AGROVOC *Thesaurus*

URL do vocabulário controlado: [http://aims.fao.org/aos/agrovoc/c\\_330704](http://aims.fao.org/aos/agrovoc/c_330704).

## **NAL *Thesaurus***

### **O que é NAL *Thesaurus*?**

NAL *Thesaurus* é um vocabulário controlado desenvolvido e mantido pela *National Agricultural Library* (NAL), dos Estados Unidos. Lançado em 2002, este vocabulário bilíngue (inglês e espanhol) tem abrangência ampla em agricultura, biologia e disciplinas correlatas. Contém mais 260 mil termos, que são organizados em 17 categorias de assunto. Além disso, inclui relações hierárquicas, associativas e de equivalência entre conceitos. (Estados Unidos, 2021).

NAL *Thesaurus* é usado principalmente para indexação de assunto de diferentes objetos de informação, dentre os quais, os *datasets*. No âmbito do Redape, o NAL *Thesaurus* é um dos vocabulários controlados escolhidos para descrever as palavras-chave e a classificação de *datasets*.

### **NAL *Thesaurus* para Palavras-chave**

A atribuição de palavras-chave é uma etapa da catalogação de *datasets* no Redape, especificamente no preenchimento dos campos do Elemento **Palavras-chave**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e preenchimento dos campos **Termo-chave**, **Vocabulário controlado do termo-chave** e **URL do vocabulário controlado**.

Segue passo a passo:

- **Etapa 1. Acesso ao site do NAL *Thesaurus*:** acesse o endereço: <https://agclass.nal.usda.gov/dne/search.shtml>. A página é apresentada como mostra a Figura 9.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

English | Español

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

[Home](#) / thesaurus search

# Thesaurus Search

Search term or text to match  
Enter the terms you wish to search for.

**Language**  
English

**Search method**  
Terms which contain this character string

**Number of terms to display**  
200

[Search the Thesaurus](#)

## A-Z List of Terms

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

## Subject Categories

The thesaurus is organized into 17 Subject Categories to facilitate browsing the vocabulary. Please note that many topic hierarchies are very deep and will require more time to display in your browser.

- [Animal Science and Animal Products](#)
- [Biological Sciences](#)
- [Breeding and Genetic Improvement](#)
- [Economics, Business and Industry](#)
- [Farms and Farming Systems](#)
- [Food and Human Nutrition](#)
- [Forest Science and Forest Products](#)
- [Geographical Locations](#)
- [Government, Law and Regulations](#)
- [Health and Pathology](#)
- [Insects and Entomology](#)
- [Natural Resources, Earth and Environment](#)

**Figura 9.** Tela de busca do NAL *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 2. Realização da busca:** no campo 'Search term' digite o termo desejado. O termo 'mites' foi escolhido como exemplo. No campo 'Select a search method' escolha a opção mais adequada. No exemplo mostrado foi escolhida a opção 'Terms start with'. Clique no botão 'Search the Thesaurus'. A tela de busca é apresentada na Figura 10 e os resultados na Figura 11.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

English | Español

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

[Home](#) / thesaurus search

# Thesaurus Search

mites

Enter the terms you wish to search for.

**Language**  
English

**Search method**  
Terms start with ...

**Number of terms to display**  
200

**Search the Thesaurus**

## A-Z List of Terms

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

## Subject Categories

The thesaurus is organized into 17 Subject Categories to facilitate browsing the vocabulary. Please note that many topic hierarchies are very deep and will require more time to display in your browser.

- [Animal Science and Animal Products](#)
- [Biological Sciences](#)
- [Breeding and Genetic Improvement](#)
- [Economics, Business and Industry](#)
- [Farms and Farming Systems](#)
- [Food and Human Nutrition](#)
- [Forest Science and Forest Products](#)
- [Geographical Locations](#)
- [Government, Law and Regulations](#)
- [Health and Pathology](#)
- [Insects and Entomology](#)
- [Natural Resources, Earth and Environment](#)
- [Physical and Chemical Sciences](#)
- [Plant Science and Plant Products](#)
- [Research, Technology and Engineering](#)
- [Rural and Agricultural Sociology](#)
- Taxonomic Classification of

**Figura 10.** Tela de busca do NAL *Thesaurus* com o termo 'mites'.

Fonte: Souza et al. (2020).

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

Home Search ▾ About the Thesaurus and Glossary ▾ Download Files Help ▾ Contact Us

## Thesaurus Search

Search term or text to match

Enter the terms you wish to search for.

**Language**

English

**Search method**

Terms which contain this charact

**Number of terms to display**

200

**Search the Thesaurus**

## Thesaurus Search Results

Search Results: Click on term to display record

[Next Page](#)

- [mites](#)
- [MITEs \(miniature inverted repeat transposable elements\)](#)
- [mithans](#)
- [Mithridate mustard](#)
- [mithuns](#)
- [miticides](#)
- [mitigation \(pollution\)](#)
- [mitigation banking](#)
- [mitochondria](#)
- [mitochondrial ATPase](#)
- [mitochondrial complex III](#)
- [mitochondrial degradation](#)
- [mitochondrial DNA](#)
- [mitochondrial electron transport complex I](#)
- [mitochondrial electron transport complex II](#)
- [mitochondrial electron transport complex III](#)
- [mitochondrial electron transport complex IV](#)
- [mitochondrial genes](#)
- [mitochondrial genetics](#)
- [mitochondrial genome](#)
- [mitochondrial genomics](#)
- [mitochondrial inheritance](#)
- [mitochondrial membrane](#)
- [mitochondrial proteins](#)
- [mitochondrial RNA](#)
- [mitochondrial thioredoxins](#)
- [mitochondrial turnover](#)
- [mitogen-activated protein kinase](#)
- [mitogen-activated protein kinase kinase](#)

## Browse Alphabetically

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

**Figura 11.** Resultado da busca do termo no NAL *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 3. Acesso a página detalhada do termo-chave:** clique no termo 'mites' no resultado da busca. Um conjunto de informações adicionais será apresentado, conforme a Figura 12.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

## Thesaurus Search Results

### Thesaurus Search

Search term or text to match

Enter the terms you wish to search for.

Language  
English

Search method  
Terms which contain this charact

Number of terms to display  
200

[Search the Thesaurus](#)

### Browse Alphabetically

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

### mites

**Subject Category**  
[L Animal Science and Animal Products](#)  
[S Biological Sciences](#)

**Definition**  
 Tiny eight-legged animals that live off plants, animals or stored food.

**Definition Source**  
 US, Environmental Protection Agency

**RDF/XML Format:**  
<http://lod.nal.usda.gov/nalt/2069.rdf>

**Persistent URI:**  
<http://lod.nal.usda.gov/nalt/2069>

**Broader Term**  
[arthropods](#)

**Narrower Term**  
[chiggers](#)  
[dust mites](#)  
[feather mites](#)  
[honey bee mites](#)  
[parasitic mites](#)  
[phytophagous mites](#)  
[predatory mites](#)  
[storage mites](#)  
[water mites](#)

**Related Term**  
[Acari](#)  
[acarology](#)

**Change Display**  
[Show Term Hierarchy](#)

**Search for this Term**  
[Google Scholar](#)  
[AGRICOLA Articles](#)  
[AGRICOLA Books](#)

**Figura 12.** Tela do termo 'mites' contendo informações sobre classificações, relações, associações, formato, identificador persistente, tradução do termo para o espanhol e número do termo.

Fonte: Souza et al. (2020).

- **Etapa 4. Obtenção da URL do termo-chave:** na página do termo 'mites', apresentada na Figura 12 copie a URL do conteúdo do campo **Persistent URI**, e cole-o no campo URL termo-chave do Elemento Palavras-chave, do formulário de catalogação no campo **URL do vocabulário controlado**. **Observação:** No formulário Dataverse não há campo destinado para a descrição da **URL do termo-chave**, por esta razão, escolheu-se o campo **URL do vocabulário controlado** para fazer o seu registro.
- **Etapa 5. Preenchimento dos campos do elemento Palavras-chave:** no formulário de catalogação de *datasets* no Redape, preencha os campos conforme o exemplo:  
 Termo-chave: mites

Vocabulário controlado do termo-chave: NAL *Thesaurus*

URL do vocabulário controlado: <http://lod.nal.usda.gov/nalt/2069>.

## **NAL *Thesaurus* para Classificação**

A classificação do assunto de *datasets* é feita por meio da atribuição de uma ou mais categorias de assunto. Tais categorias representam estruturas de assunto mais abrangentes sob as quais um assunto específico, como por exemplo, 'mites', está inserido. A atribuição de classificação de assunto, por meio de categorias, é uma etapa da catalogação de *datasets*, especificamente no preenchimento do campo **Termo de classificação**, do Elemento **Classificação** no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e preenchimento dos campos **Termo de classificação**, **Vocabulário controlado da classificação** e **URL do vocabulário controlado**.

Segue passo a passo:

- **Etapa 1. Acesso ao site do NAL *Thesaurus*** - acesse o endereço: <https://agclass.nal.usda.gov/thesaurus-search>. A página é apresentada como mostra a Figura 13.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

English | Español

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

Home / thesaurus search

## Thesaurus Search

mites

Enter the terms you wish to search for.

**Language**  
English

**Search method**  
Terms start with ...

**Number of terms to display**  
200

**Search the Thesaurus**

### A-Z List of Terms

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

### Subject Categories

The thesaurus is organized into 17 Subject Categories to facilitate browsing the vocabulary. Please note that many topic hierarchies are very deep and will require more time to display in your browser.

- [Animal Science and Animal Products](#)
- [Biological Sciences](#)
- [Breeding and Genetic Improvement](#)
- [Economics, Business and Industry](#)
- [Farms and Farming Systems](#)
- [Food and Human Nutrition](#)
- [Forest Science and Forest Products](#)
- [Geographical Locations](#)
- [Government, Law and Regulations](#)
- [Health and Pathology](#)
- [Insects and Entomology](#)
- [Natural Resources, Earth and Environment](#)
- [Physical and Chemical Sciences](#)
- [Plant Science and Plant Products](#)
- [Research, Technology and Engineering](#)
- [Rural and Agricultural Sociology](#)
- Taxonomic Classification of

**Figura 13.** Tela de busca NAL *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 2. Realização da busca:** no campo ‘Search term or text to match’ digite o termo desejado para o qual pretende realizar a classificação de assunto do *dataset*, por exemplo, ‘mites’. Em seguida, escolha a opção mais adequada em ‘Select a search method’. No exemplo, foi escolhida a opção ‘Terms start with’. Clique no botão ‘Search the Thesaurus’. A tela de busca é apresentada na Figura 14 e os resultados na Figura 15.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

English | Español

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

[Home](#) / thesaurus search

# Thesaurus Search

mites

Enter the terms you wish to search for.

**Language**  
English

**Search method**  
Terms start with ...

**Number of terms to display**  
200

**Search the Thesaurus**

## A-Z List of Terms

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

## Subject Categories

The thesaurus is organized into 17 Subject Categories to facilitate browsing the vocabulary. Please note that many topic hierarchies are very deep and will require more time to display in your browser.

- [Animal Science and Animal Products](#)
- [Biological Sciences](#)
- [Breeding and Genetic Improvement](#)
- [Economics, Business and Industry](#)
- [Farms and Farming Systems](#)
- [Food and Human Nutrition](#)
- [Forest Science and Forest Products](#)
- [Geographical Locations](#)
- [Government, Law and Regulations](#)
- [Health and Pathology](#)
- [Insects and Entomology](#)
- [Natural Resources, Earth and Environment](#)
- [Physical and Chemical Sciences](#)
- [Plant Science and Plant Products](#)
- [Research, Technology and Engineering](#)
- [Rural and Agricultural Sociology](#)
- Taxonomic Classification of

**Figura 14.** Tela de busca do NAL *Thesaurus* para o termo 'mites'.

Fonte: Souza et al. (2020).

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

Home Search ▾ About the Thesaurus and Glossary ▾ Download Files Help ▾ Contact Us

## Thesaurus Search

Search term or text to match

Enter the terms you wish to search for.

**Language**

English

**Search method**

Terms which contain this charact

**Number of terms to display**

200

**Search the Thesaurus**

## Thesaurus Search Results

Search Results: Click on term to display record

[Next Page](#)

- [mites](#)
- [MITEs \(miniature inverted repeat transposable elements\)](#)
- [mithans](#)
- [Mithridate mustard](#)
- [mithuns](#)
- [miticides](#)
- [mitigation \(pollution\)](#)
- [mitigation banking](#)
- [mitochondria](#)
- [mitochondrial ATPase](#)
- [mitochondrial complex III](#)
- [mitochondrial degradation](#)
- [mitochondrial DNA](#)
- [mitochondrial electron transport complex I](#)
- [mitochondrial electron transport complex II](#)
- [mitochondrial electron transport complex III](#)
- [mitochondrial electron transport complex IV](#)
- [mitochondrial genes](#)
- [mitochondrial genetics](#)
- [mitochondrial genome](#)
- [mitochondrial genomics](#)
- [mitochondrial inheritance](#)
- [mitochondrial membrane](#)
- [mitochondrial proteins](#)
- [mitochondrial RNA](#)
- [mitochondrial thioredoxins](#)
- [mitochondrial turnover](#)
- [mitogen-activated protein kinase](#)
- [mitogen-activated protein kinase kinase](#)

## Browse Alphabetically

[0-9](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

**Figura 15.** Resultado da busca do termo 'mites' no NAL *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 3. Acesso a página detalhada do termo-chave:** clique no termo 'mites' no resultado da busca. Um conjunto de informações adicionais será apresentado, conforme a Figura 16.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

## Thesaurus Search

Search term or text to match

Enter the terms you wish to search for.

Language

English

Search method

Terms which contain this charact

Number of terms to display

200

Search the Thesaurus

## Thesaurus Search Results

### mites

**Subject Category**  
[L Animal Science and Animal Products](#)  
[S Biological Sciences](#)

**Definition**  
 Tiny eight-legged animals that live off plants, animals or stored food.

**Definition Source**  
 US, Environmental Protection Agency

**RDF/XML Format:**  
<http://lod.nal.usda.gov/nalt/2069.rdf>

**Persistent URI:**  
<http://lod.nal.usda.gov/nalt/2069>

**Broader Term**  
[arthropods](#)

**Narrower Term**  
[chiggers](#)  
[dust mites](#)  
[feather mites](#)  
[honey bee mites](#)  
[parasitic mites](#)  
[phytophagous mites](#)  
[predatory mites](#)  
[storage mites](#)  
[water mites](#)

**Related Term**  
[Acari](#)  
[acarology](#)

### Change Display

[Show Term Hierarchy](#)

### Search for this Term

[Google Scholar](#)  
[AGRICOLA Articles](#)  
[AGRICOLA Books](#)

## Browse Alphabetically

0-9 [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#)  
[L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#)  
[X](#) [Y](#) [Z](#)

**Figura 16.** Tela do termo 'mites' contendo informações sobre classificações, relações, associações, formato, identificador persistente, tradução do termo para o espanhol e número do termo.

Fonte: Souza et al. (2020).

- Etapa 4. Obtenção da URL para o termo de classificação:** na página do termo 'mites', apresentado na Figura 16, escolha em 'Subject Category' uma categoria de assunto que melhor represente o termo 'mites', como por exemplo 'L Animal Science and Animal Product'. Clique no item selecionado. A página mostrada na 17 será apresentada. Copie a URL da página, que será a URL do termo de classificação. **Observação:** No formulário Dataverse não há campo destinado para a descrição da **URL de termo de classificação**, por esta razão, escolheu-se o campo **URL do vocabulário controlado** para fazer o seu registro.

An official website of the United States government. [Here's how you know.](#)

**USDA** NAL Agricultural Thesaurus and Glossary  
U.S. DEPARTMENT OF AGRICULTURE

Home Search About the Thesaurus and Glossary Download Files Help Contact Us

## Thesaurus Search Results

### Thesaurus Search

Search term or text to match

Enter the terms you wish to search for.

**Language**

English

**Search method**

Terms which contain this character

**Number of terms to display**

200

**Search the Thesaurus**

### Browse Alphabetically

0-9 A B C D E F G H I J K  
L M N O P Q R S T U V W  
X Y Z

L Animal Science and Animal Products

[Next Page](#)

- [abalone](#)
- [abalone meat](#)
- [abdominal ganglia](#)
- [abnormal hemoglobin](#)
- [abomasum](#)
- [acarology](#)
- [accessory sex glands](#)
- [acetabulum](#)
- [acid-base balance](#)
- [acidified milk](#)
- [acidophilus milk](#)
- [acrosome reaction](#)
- [acute phase proteins](#)
- [ad libitum feeding](#)
- [adenoids](#)
- [adipose tissue](#)
- [adrenal cortex](#)
- [adrenal glands](#)
- [adrenal medulla](#)
- [adult development](#)
- [adulterated feeds](#)
- [age at first egg](#)
- [age at slaughter](#)
- [aggregation behavior](#)
- [air sacs](#)

[Next Page](#)

**Figura 17.** Tela da categoria 'L Animal Sciences and Animal Products' do NAL *Thesaurus*.

Fonte: Souza et al. (2020).

- **Etapa 5. Preenchimento dos campos do elemento Classificação.** No formulário de catalogação de *datasets* no Redape, preencha os campos conforme o exemplo:

Termo de classificação: L Animal Science and Animal Products

Vocabulário controlado da classificação: NAL *Thesaurus*

URL do vocabulário controlado: <https://agclass.nal.usda.gov/mtwdk.exe?k=default&l=60&s=cid&w=100>.

## NCBI *Taxonomy*

### O que é NCBI *Taxonomy*?

*National Center for Biotechnology Information (NCBI) Taxonomy* é um sistema de informação que inclui nomes e classificações de organismos de sequências de nucleotídeos e de proteínas. *NCBI Taxonomy* é também o repositório padrão de nomenclatura e classificação para o *International Nucleotide Sequence Database Collaboration (INSDC)*. Contém uma lista de nomes considerados nomenclaturalmente corretos ou válidos, os quais são definidos de acordo com diferentes códigos de nomenclatura. Inclui ainda vários nomes que existem a despeito de códigos. (National Center for Biotechnology Information, 2021). *NCBI Taxonomy* integra o Entrez, um metabuscador mantido pelo NCBI, que coleta, trata e disponibiliza informação relevante para o desenvolvimento da biotecnologia.

*NCBI Taxonomy* é usado principalmente para indexação de assunto de diferentes objetos de informação, entre os quais, os *datasets*. No âmbito do LMB e seus projetos de pesquisa, o NCBI é um dos vocabulários controlados escolhidos para descrever assuntos (palavras-chave e classificação) de *datasets* ômicos no Redape.

### NCBI *Taxonomy* para Palavras-chave

A atribuição de palavras-chave é uma etapa da catalogação de *datasets* no Redape, especificamente no preenchimento dos campos do Elemento **Palavras-chave**, no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e preenchimento dos campos **Termo-chave**, **Vocabulário controlado do termo-chave** e **URL do vocabulário controlado**.

Segue passo a passo:

- **Etapa 1. Acesso ao site do NCBI *Taxonomy*:** busque o endereço: <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>. A página é apresentada como mostra a Figura 18.



The "Token set" option returns longer names that include the search terms "Complete match" option versus the "Set of tokens" option. The "Phonetic : organism name. It tries to find the phonetically closest strings (try "Drozofi

**This is the top level of the taxonomy database maintained by NCBI/GenBank. Yo**

- [Archaea](#)
- [Bacteria](#)
- [Eukaryota](#)
- [Viruses](#)
- [Other](#)
- [Unclassified](#)

**Figura 18.** Tela de busca do NCBI *Taxonomy*.

Fonte: Souza et al. (2020).

- **Etapa 2. Realização da busca:** no campo 'Search for' digite o termo desejado, como por exemplo *Rhipicephalus (Boophilus) microplus*. Clique em 'Go', e veja os resultados obtidos. A Figura 19 apresenta os resultados.

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are logos for Entrez, PubMed, Nucleotide, Protein, and Genome. The search bar contains the text "Rhipicephalus (Boophilus) microplus" and is set to search for "complete name". Below the search bar, there are options for "Display" (set to 3) and "levels using filter" (set to none). The search results are displayed under the heading "Rhipicephalus microplus".

Taxonomy ID: 6941 (for references in articles please use NCBI:txid6941)

current name

***Rhipicephalus microplus*** (Canestrini, 1888)

homotypic genbank synonym: ***Boophilus microplus***

homotypic synonym: ***Rhipicephalus (Boophilus) microplus***

Genbank common name: **southern cattle tick**

NCBI BLAST name: **mites & ticks**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

Other names:

common name(s)

**cattle tick**

**Figura 19.** Tela de resultados da busca do termo *Rhipicephalus (Boophilus) microplus* no NCBI Taxonomy.

Fonte: Souza et al. (2020).

- Etapa 3. Obtenção da URL para o termo-chave:** na página de resultado de busca do termo *Rhipicephalus microplus*, apresentado na Figura 19, no frame à direita, em 'Taxonomy', clique no endereço correspondente na coluna 'Direct links' para abrir a página do referido termo. Em seguida, clique em *Rhipicephalus microplus* para abrir a página deste termo. Copie a URL desta página e cole-a no campo **URL do vocabulário controlado** do Elemento **Palavras-chave**, do formulário de catalogação. **Observação:** No formulário Dataverse não há campo destinado para a descrição da **URL do termo-chave**, por esta razão, escolheu-se o campo **URL do vocabulário controlado** para fazer o seu registro.
- Etapa 4. Preenchimento dos campos do elemento Palavras-chave:** No formulário de catalogação de *datasets* no Redape, preencha os campos conforme o exemplo:

Termo-chave: *Rhipicephalus microplus*

Vocabulário controlado do termo-chave: NCBI *Taxonomy*

URL do vocabulário controlado:

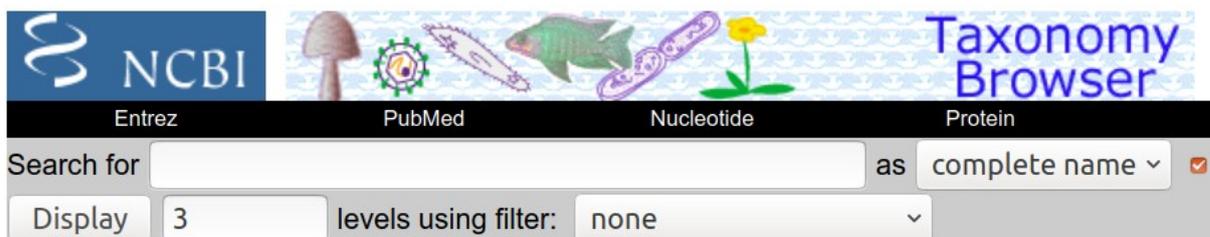
<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=6941>.

## NCBI *Taxonomy* para Classificação

A classificação de assunto de *datasets* é feita por meio da atribuição de uma ou mais categorias de assunto. Tais categorias representam estruturas de assunto mais abrangentes sob as quais um assunto específico, como por exemplo, '*Rhipicephalus (Boophilus) microplus*', está inserido. A atribuição de classificação de assunto, por meio de categorias, é uma etapa da catalogação de *datasets*, especificamente no preenchimento do campo **Termo de classificação**, do Elemento **Classificação** no formulário de catalogação. Abaixo encontra-se o passo a passo para a realização de busca e preenchimento dos campos **Termo de classificação**, **Vocabulário controlado da classificação** e **URL do vocabulário controlado da classificação**.

Segue passo a passo:

- **Etapa 1. Acesso ao site do NCBI *Taxonomy*:** acesse o endereço: <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>. A página é apresentada como mostra a Figura 20.



The "Token set" option returns longer names that include the search terms "Complete match" option versus the "Set of tokens" option. The "Phonetic" option returns the organism name. It tries to find the phonetically closest strings (try "Drozofi

**This is the top level of the taxonomy database maintained by NCBI/GenBank. You**

- [Archaea](#)
- [Bacteria](#)
- [Eukaryota](#)
- [Viruses](#)
- [Other](#)
- [Unclassified](#)

**Figura 20.** Tela de busca do NCBI *Taxonomy*.

Fonte: Souza et al. (2020).

- **Etapa 2. Realização da busca:** no campo 'Search for' digite o termo desejado para o qual pretende fazer a classificação de assunto, como por exemplo *Rhipicephalus (Boophilus) microplus*. Clique em 'Go', e veja os resultados obtidos. A Figura 21 apresenta os resultados.

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are logos for Entrez, PubMed, Nucleotide, and Protein. The search bar contains the text "Rhipicephalus (Boophilus) microplus" and the dropdown menu is set to "complete name". Below the search bar, there are options for "Display" (set to 3) and "levels using filter" (set to none). The main heading is "Rhipicephalus microplus". Below this, the Taxonomy ID is 6941. The current name is "Rhipicephalus microplus (Canestrini, 1888)". There are two homotypic synonyms: "Boophilus microplus" (genbank) and "Rhipicephalus (Boophilus) microplus". The Genbank common name is "southern cattle tick" and the NCBI BLAST name is "mites & ticks". The rank is "species". Genetic codes are provided for standard and mitochondrial DNA. Other names include "cattle tick". A full lineage is listed, starting from cellular organisms down to Boophilus.

NCBI  
Entrez PubMed Nucleotide Protein

Search for **Rhipicephalus (Boophilus) microplus** as complete name ▾

Display 3 levels using filter: none ▾

**Rhipicephalus microplus**

Taxonomy ID: 6941 (for references in articles please use NCBI:txid6941)

current name

***Rhipicephalus microplus*** (Canestrini, 1888)

homotypic genbank synonym: ***Boophilus microplus***

homotypic synonym: ***Rhipicephalus (Boophilus) microplus***

Genbank common name: **southern cattle tick**

NCBI BLAST name: **mites & ticks**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondria\)](#)

Other names:

common name(s)

**cattle tick**

[Lineage\( full \)](#)  
[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Panarthropoda](#); [Arthropoda](#); [Chelicerata](#); [Arachnida](#); [Acari](#); [Parasitiformes](#); [Rhipicephalinae](#); [Rhipicephalus](#); [Boophilus](#)

**Figura 21.** Tela de resultados da busca do termo *Rhipicephalus (Boophilus) microplus* no NCBI *Taxonomy*.

Fonte: Souza et al. (2020).

- **Etapa 3. Seleção do termo de classificação:** no resultado de busca mostrado na Figura 22, escolha em 'Lineage (full)' uma categoria de assunto que melhor classifique o termo *Rhipicephalus microplus*, como por exemplo, 'Ixodida'. A Figura 22 apresenta o conteúdo para o termo escolhido. Selecione item Ixodida, e a tela mostrada na Figura 23 será apresentada.



NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein

Search for  as complete name

Display 3 levels using filter: none

**Ixodida**

Taxonomy ID: 6935 (for references in articles please use NCBI:txid6935)

current name

**Ixodida**

Genbank common name: **ticks**

NCBI BLAST name: **mites & ticks**

Rank: **order**

Genetic code: [Translation table 1 \(Standard\)](#)

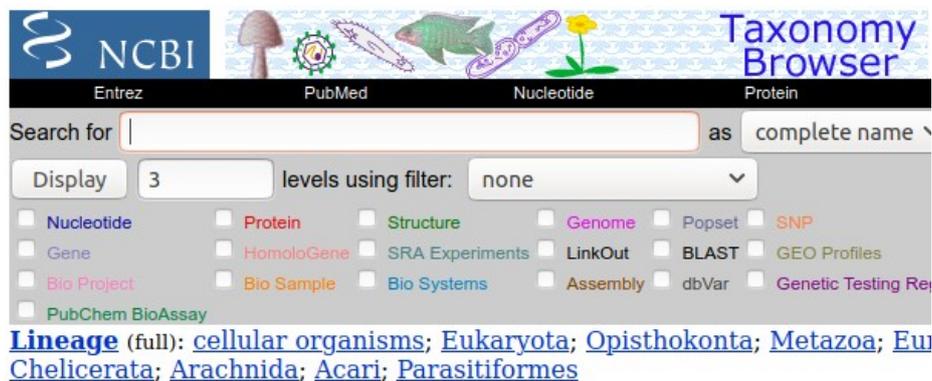
Mitochondrial genetic code: [Translation table 5 \(Invertebrate Mitochondrial\)](#)

**Lineage( full )**

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Ecdysozoa](#); [Panarthropoda](#); [Arthropoda](#); [Chelicerata](#); [Arachnida](#); [Acari](#)

**Figura 22.** Hierarquia do 'Lineage (full)' do termo *Rhipicephalus microplus*.

Fonte: Souza et al. (2020).



NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein

Search for  as complete name

Display 3 levels using filter: none

Nucleotide  Protein  Structure  Genome  Popset  SNP

Gene  HomoloGene  SRA Experiments  LinkOut  BLAST  GEO Profiles

Bio Project  Bio Sample  Bio Systems  Assembly  dbVar  Genetic Testing Re

PubChem BioAssay

**Lineage** (full): [cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Ecdysozoa](#); [Panarthropoda](#); [Arthropoda](#); [Chelicerata](#); [Arachnida](#); [Acari](#); [Parasitiformes](#)

- **Ixodida** (ticks) *Click on organism name to get more information.*
  - **Ixodoidea**
    - **Argasidae** (softbacked ticks)
      - **Antricola**
      - **Argas**
      - **Carios**
      - **Nothoaspis**
      - **Ornithodoros** (relapsing fever ticks)
      - **Otobius**
    - **Ixodidae** (hardbacked ticks)
      - **Amblyomminae**

**Figura 23.** Resultado da escolha do termo 'Ixodida'.

Fonte: Souza et al. (2020).

- **Etapa 4. Obtenção da URL para o termo de classificação:** na página de resultado de busca da categoria 'Ixodida' apresentada na Figura 23, no frame à direita, em 'Taxonomy', clique no endereço correspondente na coluna 'Direct links' para abrir a página da referida categoria. Em seguida, clique em 'Ixodida' para abrir a página desta categoria. Copie a URL desta página e cole-a no campo **URL do vocabulário controlado** do Elemento **Classificação**, do formulário de catalogação. **Observação:** No formulário Dataverse não há campo destinado para a descrição da **URL do termo de classificação**, por esta razão, escolheu-se o campo **URL do vocabulário controlado** para fazer o seu registro.
- **Etapa 5. Preenchimento dos campos do elemento Classificação:** no formulário de catalogação de *datasets* no Redape, preencha os campos conforme o exemplo:  
Termo de classificação: Ixodida  
Vocabulário controlado da classificação: NCBI *Taxonomy*  
URL do vocabulário controlado:  
<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=6935>.

## Referências

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de biblioteconomia e arquivologia**. Brasília, DF: Briquet de Lemos, 2008. 451 p.

ESTADOS UNIDOS. Department of Agriculture. **NAL Agricultural Thesaurus and Glossary**. Disponível em: <<https://agclass.nal.usda.gov/agt.shtml>>. Acesso em: 19 out. 2021.

FAO. **AGROVOC**. Rome, 2021. Disponível em: <<https://agrovoc.fao.org/browse/agrovoc/en/>>. Acesso em: 19 out. 2021.

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, J. L. R.; SANTOS, L. O. B. da S. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 14, n. 3, p. 175-192, 2019. DOI: 10.22478/ufpb.1981-0695.2019v14n3.46969.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **Taxonomy**. Bethesda. Disponível em: <<https://www.ncbi.nlm.nih.gov/taxonomy>>. Acesso em: 19 out. 2021.

SOUZA, M. I. F.; VISOLI, M. C.; TORRES, T. Z. **Catálogo de dataset no Repositório de Dados da Embrapa**: a experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa. Campinas: Embrapa Informática Agropecuária, 2020. 116 p. (Embrapa Informática Agropecuária. Documentos, 172). Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1127941/1/Doc172-2020.pdf>>. Acesso em: 19 out. 2021.

VEIGA, V. Gestão de dados de pesquisa FAIR: dando um JUMP em seus dados. In: ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. **Anais...** Rio de Janeiro: UFRJ, 2019. 59 p. Disponível em: <<https://www.arca.fiocruz.br/handle/icict/33343>>. Acesso em: 19 out. 2021.

THE WORLD WIDE WEB CONSORTIUM. **222 best practices for vocab selection**. [S. l., 2012]. Disponível em: <[https://www.w3.org/2011/gld/wiki/222\\_Best\\_Practices\\_for\\_Vocab\\_Selection](https://www.w3.org/2011/gld/wiki/222_Best_Practices_for_Vocab_Selection)>. Acesso em: 24 ago. 2021.

## APÊNDICE B – Guia Resumido das Orientações para Catalogação de *Datasets* Ômicos no Redape

### Introdução

O processo de catalogação de *datasets* ômicos no ambiente do software Dataverse utiliza um formulário para descrição dos metadados e upload de arquivos de dados. O software Dataverse oferece ao usuário a possibilidade de escolher qual esquema de metadados deseja usar para melhor atender os seus interesses e o contexto dos dados a serem depositados. O *Citation Metadata*, doravante denominado de esquema Metadados de Citação, é predefinido no software Dataverse e exportável para o padrão *Dublin Core*, inclusive. Por conseguinte, este é o esquema genérico de metadados adotado no Laboratório Multiusuário de Bioinformática da Embrapa para a catalogação de *datasets*. No entanto, para atender diretamente ao domínio dos dados biológicos, o Dataverse oferece também o esquema específico de metadados - o *Life Science Metadata* -, aqui denominado de esquema Metadados Ciência da Vida, considerado mais apropriado para descrever *datasets* ômicos. Desse modo, o esquema Metadados de Citação e o esquema Metadados Ciência da Vida podem ser utilizados concomitantemente, desde que o *dataverse*<sup>17</sup> ao qual o *dataset* estiver vinculado, tenha sido criado com essa possibilidade.

Como em quaisquer conjuntos de metadados, o processo de catalogação de *datasets* deve seguir orientações e regras específicas de catalogação<sup>18</sup> para descrever os elementos metadados e respectivos campos e subcampos. O detalhamento desses dois esquemas de metadados - Metadados de Citação e Metadados Ciência da Vida, com elementos, campos e subcampos, é apresentado de forma resumida neste Apêndice B, com breve descrição, indicação de exigência de preenchimento e exemplos.

---

<sup>17</sup> *Dataverse* é aqui entendido como um projeto que reúne conjuntos de dados (*datasets*).

<sup>18</sup> As orientações para catalogação de *dataset* foram elaboradas a partir das regras básicas do Código de Catalogação Anglo-Americano - AACR2 (Ribeiro, 1995, 2003; Código..., 2004).

Este Apêndice é um instrumento sucinto de ajuda rápida ao catalogador no momento da catalogação de *datasets* ônicos, no Redape. As orientações completas estão expostas no corpo principal desta publicação.

No tocante ao preenchimento do formulário on-line de catalogação de *datasets* no Redape, os elementos metadados são classificados em: requeridos, fortemente recomendados, recomendados e opcionais<sup>19</sup>.

A descrição das exigências de preenchimento encontra-se na Tabela 1.

**Tabela 1.** Exigências de preenchimento dos elementos do esquema Metadados de Citação (campos e subcampos).

| <b>Exigência de preenchimento</b> | <b>Descrição</b>   |
|-----------------------------------|--|
| Requerido                         | Elemento e/ou campo e/ou subcampo requerido - deve ser preenchido. Software Dataverse exige o preenchimento.   |
| Fortemente recomendado            | Elemento e/ou campo e/ou subcampo - preenchimento fortemente recomendado. Embora não seja exigência do software, seu preenchimento fornecerá informações para análise de indicadores de qualidade dos metadados. |
| Recomendado                       | Elemento e/ou campo e/ou subcampo – preenchimento recomendado. Embora não seja exigência do software, o seu preenchimento fornecerá informações para aumentar o nível de qualidade dos metadados                 |
| Opcional                          | Elemento e/ou campo e/ou subcampo – preenchimento opcional, ficando a critério do catalogador o fornecimento de informações adicionais. Preencher caso pertinente.   |

Fonte: Adaptado de Souza et al. (2020).

## Descrição dos elementos do esquema Metadados de Citação

**Tabela 2.** Descrição dos elementos do esquema Metadados de Citação (campos e subcampos) no Redape.

| <b>Campos</b> | <b>Definição</b> | <b>Exigência de preenchimento</b> | <b>Exemplos</b> |
|---------------|------------------|-----------------------------------|-----------------|
|---------------|------------------|-----------------------------------|-----------------|

Continua...

<sup>19</sup> Os dois níveis fortemente recomendados e recomendados foram acrescentados para ampliar a capacidade de recuperação, reuso e integração dos *datasets* com outros conjuntos e repositórios, de acordo com os princípios FAIR (Go FAIR, 2018).

|                            |  |                        |   |
|----------------------------|--|------------------------|---|
| <b>Título</b>              | Nome atribuído ao <i>dataset</i> , e pelo qual será formalmente conhecido. Estrutura do título: Tipo de estudo ômico + Amostra estudada + Espécie estudada (nome comum + nome científico). | Requerido              | Transcriptoma da glândula salivar do carrapato bovino ( <i>Rhipicephalus (B.) microplus</i> )   |
| Subtítulo                  | Nome secundário atribuído ao <i>dataset</i> .  | Opcional               | Expressão gênica na saliva do carrapato bovino.   |
| Título alternativo         | Nome alternativo para o <i>dataset</i> ; um nome pelo qual conhecido ou referido; um nome abreviado.   | Opcional               | Genoma do carrapato   |
| <b>URL alternativa</b>     | URL alternativa válida onde o <i>dataset</i> pode ser visualizado.   | Recomendado            | <a href="https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material">https://www.frontiersin.org/articles/10.3389/fcimb.2019.00477/full#supplementary-material</a> |
| <b>Outro identificador</b> | Outro identificador digital único existente para identificar o <i>dataset</i> .  | Não se aplica          | Não se aplica   |
| Agência                    | Nome da agência geradora do identificador digital.   | Opcional               | NCBI  |
| Identificador              | Outro identificador digital persistente que identifica o <i>dataset</i> .  | Opcional               | PRJNA596777   |
| <b>Autor</b>               | Nome de pessoas, instituições ou agências responsáveis pela criação do <i>dataset</i> .  | Não se aplica          | Não se aplica   |
| Nome                       | Nome(s) do autor(es) do <i>dataset</i> . Pode(m) ser autor(es) pessoal(is) ou autor corporativo.   | Requerido              | Pereira Júnior, João<br>Silva, José   |
| Afiliação                  | Nome da instituição de vínculo do(s) autor(es) pessoal(is).  | Recomendado            | Embrapa Gado de Corte   |
| Sistema do Identificador   | Sistema do identificador digital do autor.   | Recomendado            | ORCID<br>ResearcherID<br>Scopus ID  |
| Identificador              | Identificador digital persistente do(s) autor(es) pessoal(is) do <i>dataset</i> .  | Recomendado            | <a href="http://orcid.org/0000-0002-1825-0097">http://orcid.org/0000-0002-1825-0097</a>   |
| <b>Contato</b>             | Nome de pessoas, instituições ou agências para contato sobre o <i>dataset</i> .  | Não se aplica          | Não se aplica   |
| Nome                       | Nome pessoal ou corporativo para contato sobre o <i>dataset</i> .  | Fortemente recomendado | José Silva  |
| Afiliação                  | Nome da instituição à qual a pessoa está vinculada.  | Recomendado            | Embrapa Gado de Corte   |
| E-mail do contato          | Endereço de e-mail do contato, seja pessoal ou corporativo.  | Requerido              | jose.silva@empresa.com.br   |

|                                       |   |                        |   |
|---------------------------------------|---|------------------------|---|
| <b>Descrição</b>                      | Texto sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> .   | Não se aplica          | Não se aplica   |
| Texto                                 | Texto sucinto contendo o propósito, a natureza e o escopo do <i>dataset</i> .   | Requerido              | <i>Dataset</i> réplica 1 da glândula salivar Holandês, dados brutos formato <i>fastq</i> .<br>Sequence name; Sequence description; Sequence length; Blast hits count; Blast min e-value; Blast similarity mean; Blast hit description (HSP); Blast hit taxonomy name; Enzyme code; Enzyme name; InterPro accession.   |
| Data                                  | Data em que a descrição do <i>dataset</i> foi realizada.  | Opcional               | 2020-06-08  |
| <b>Assunto</b>                        | Áreas de domínio específicas e relevantes para o <i>dataset</i> .   | Requerido              | Agricultural Sciences<br>Medicine, Health and Life Sciences   |
| <b>Palavras-chave</b>                 | Palavras que expressam o conteúdo-chave do <i>dataset</i> , sendo capazes de representar aspectos relevantes para a compreensão do assunto.   | Não se aplica          | Não se aplica   |
| Termo-chave                           | Termo-chave   | Fortemente recomendado | Carrapato<br>Host-parasite relationship<br>Nucleotide sequences<br>Ticks<br>Transcriptome   |
| Vocabulário-controlado do termo-chave | Vocabulário controlado utilizado para a atribuição de termo-chave   | Fortemente recomendado | AGROVOC <i>Thesaurus</i><br>NAL <i>Thesaurus</i><br>NCBI <i>Taxonomy</i>  |
| URL do vocabulário controlado         | URL do termo-chave onde o mesmo possa ser acessado; esta URL deve ser inserida neste campo 'URL do vocabulário controlado'.   | Fortemente recomendado | URL da palavra-chave 'nucleotide sequence' (AGROVOC <i>Thesaurus</i> ):<br><a href="http://aims.fao.org/aos/agrovoc/c_27583">http://aims.fao.org/aos/agrovoc/c_27583</a><br>URL da palavra-chave 'nucleotide sequences' (NAL <i>Thesaurus</i> ):<br><a href="https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2">https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;w=15979&amp;s=5&amp;t=2</a> |
| <b>Classificação</b>                  | Refere-se à classificação do assunto do <i>dataset</i> , por meio da atribuição de uma ou mais categorias, sob as quais podem ser agrupados <i>datasets</i> de natureza semelhante. | Não se aplica          | Não se aplica   |
| Termo de                              | Indicação da classificação de   | Fortemente             | L Animal Science and Animal Products (NAL <i>Thesaurus</i> )  |

|   |   |                        |  |
|---|---|------------------------|--|
| classificação                           | assunto (categoria).  | recomendado            | Phenomena (AGROVOC <i>Thesaurus</i> )<br>Ixodida (NCBI <i>Taxonomy</i> )   |
| Vocabulário controlado da classificação | Vocabulário controlado utilizado para a classificação de assunto.   | Fortemente recomendado | AGROVOC <i>Thesaurus</i><br>NAL <i>Thesaurus</i><br>NCBI <i>Taxonomy</i>   |
| URL do vocabulário controlado           | URL do termo de classificação; esta URL deve ser inserida neste campo 'URL do vocabulário controlado da classificação'. | Fortemente recomendado | URL do termo de classificação 'phenomena' (AGROVOC <i>Thesaurus</i> ):<br><a href="http://aims.fao.org/aos/agrovoc/c_330704">http://aims.fao.org/aos/agrovoc/c_330704</a><br>URL do termo de classificação 'L Animal Science and Animal Products' (NAL <i>Thesaurus</i> ):<br><a href="https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100">https://agclass.nal.usda.gov/mtwdk.exe?k=default&amp;l=60&amp;s=cid&amp;w=100</a><br>URL do termo de classificação 'Ixodida' (NCBI <i>Taxonomy</i> ):<br><a href="https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&amp;id=6935&amp;lvl=3&amp;p=nucore&amp;keep=1&amp;srchmode=1&amp;unlock">https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Tree&amp;id=6935&amp;lvl=3&amp;p=nucore&amp;keep=1&amp;srchmode=1&amp;unlock</a> |
| <b>Publicação relacionada</b>           | Publicações que foram produzidas a partir de dados do <i>dataset</i> .  | Não se aplica          | Não se aplica  |
| Citação                                 | Referência completa da publicação relacionada ao <i>dataset</i> .   | Recomendado            | GIACHETTO, P. F. Transcriptoma do carrapato dos bovinos. In: GARCIA, M. V.; KOLLER, W. W. (Ed.). <b>Carrapatos na cadeia produtiva de bovinos</b> . Brasília, DF: Embrapa, 2019. cap. 16. p. 205-222.  |
| Tipo de identificador                   | Tipo de identificador digital usado pela publicação relacionada.  | Recomendado            | DOI  |
| Número do identificador                 | Número do identificador digital único da publicação relacionada.  | Recomendado            | <a href="https://doi.org/10.3389/fcimb.2019.00477">https://doi.org/10.3389/fcimb.2019.00477</a>  |
| URL da publicação                       | URL da página web onde se encontra a publicação.  | Opcional               | <a href="https://www.frontiersin.org/journals/cellular-and-infection-microbiology">https://www.frontiersin.org/journals/cellular-and-infection-microbiology</a>  |
| <b>Notas</b>                            | Informações adicionais sobre o <i>dataset</i> .   | Opcional               | Réplica de dados de glândula salivar do carrapato bovino infestando gado Holandês. Ensaio envolve dados de gado Nelore e mestiço Nelore x Holandês.  |
| <b>Idioma</b>                           | Idioma do <i>dataset</i> .  | Opcional               | Inglês<br>Inglês Português<br>Inglês Português Espanhol  |

|                                 |  |                        |   |
|---------------------------------|--|------------------------|---|
| <b>Produtor</b>                 | Pessoa ou instituição com responsabilidade financeira ou administrativa sobre o <i>dataset</i> .   | Não se aplica          | Não se aplica   |
| Nome do produtor                | Nome pessoal ou corporativo do produtor do <i>dataset</i> .  | Recomendado            | José Silva<br>Empresa Brasileira de Pesquisa Agropecuária   |
| Afiliação do produtor           | Nome da instituição de vínculo do produtor.  | Recomendado            | Embrapa LMB   |
| Abreviatura do nome do produtor | Abreviatura do nome do produtor caso seja um nome corporativo.   | Opcional               | Embrapa Gado de Corte<br>Empresa Brasileira de Pesquisa Agropecuária  |
| URL do produtor                 | URL válida da página na web do produtor do <i>dataset</i> .  | Opcional               | URL Empresa Brasileira de Pesquisa Agropecuária:<br><a href="https://www.embrapa.br/">https://www.embrapa.br/</a><br>URL do Laboratório Multiusuário de Bioinformática da Embrapa:<br><a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a> |
| URL do logotipo do produtor     | URL válida da página web onde o logotipo do produtor do <i>dataset</i> pode ser visualizado.   | Opcional               | <a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a>   |
| <b>Data de produção</b>         | Data em que o <i>dataset</i> foi produzido.  | Fortemente recomendado | 2015-05-01  |
| <b>Local de produção</b>        | Local onde o <i>dataset</i> e quaisquer outros materiais relacionados foram produzidos.  | Recomendado            | Laboratório de Biotecnologia Animal Centro de Genômica Funcional  |
| <b>Colaborador</b>              | Pessoa ou instituição responsável pela coleta, gestão ou qualquer outra contribuição relevante para o desenvolvimento do <i>dataset</i> .              | Não se aplica          | Não se aplica   |
| Tipo de colaborador             | Tipo de colaborador do <i>dataset</i> .  | Opcional               | Coletor de dados<br>Curador de dados<br>Líder do projeto<br>Pesquisador   |
| Nome do colaborador             | Nome pessoal ou corporativo do colaborador do <i>dataset</i> .   | Fortemente recomendado | Nome pessoal de colaborador:<br>Paula Regina Kuser Falcão<br>Nome de corporativo (instituição) de colaborador:<br>Laboratório de Biotecnologia Animal Centro de Genômica Funcional  |
| <b>Financiamento</b>            | Informações sobre instituição, agência de fomento, organização e projetos responsável pela concessão de suporte e/ou apoio financeiro para a produção. | Não se aplica          | Não se aplica   |

|                                     |  |                        |  |
|-------------------------------------|--|------------------------|--|
| Nome do órgão financiador           | Nome da instituição, agência de fomento ou organização que concedeu apoio financeiro para a produção do <i>dataset</i> .                                     | Recomendado            | Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)<br>Empresa Brasileira de Pesquisa Agropecuária (Embrapa)<br>Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp)   |
| Número do contrato                  | Número do contrato de concessão de financiamento   | Recomendado            | 01.11.07.002.03.00 (projeto SEG, da Embrapa)   |
| <b>Distribuidor</b>                 | Pessoa ou instituição designada pelo autor ou produtor do <i>dataset</i> para gerar cópias do trabalho, incluindo quaisquer edições ou revisões necessárias. | Não se aplica          | Não se aplica  |
| Nome do distribuidor                | Nome pessoal ou corporativo do distribuidor do <i>dataset</i> .  | Fortemente recomendado | Nome pessoal:<br>José Silva<br>Nome de corporativo:<br>Empresa Brasileira de Pesquisa Agropecuária   |
| Afiliação do distribuidor           | Nome da instituição à qual o distribuidor do <i>dataset</i> está vinculado.  | Recomendado            | Empresa Brasileira de Pesquisa Agropecuária  |
| Abreviatura do nome do distribuidor | Abreviatura do nome do distribuidor caso seja um nome corporativo.   | Opcional               | Embrapa<br>LMB   |
| URL do logotipo do distribuidor     | URL válida da página na web do distribuidor do <i>dataset</i> .  | Opcional               | URL da Empresa Brasileira de Pesquisa Agropecuária:<br><a href="https://www.embrapa.br/">https://www.embrapa.br/</a><br>URL do Laboratório Multiusuário de Bioinformática da Embrapa:<br><a href="https://www.embrapa.br/laboratorio-de-bioinformatica">https://www.embrapa.br/laboratorio-de-bioinformatica</a>   |
| URL do logotipo                     | URL válida da página web onde o logotipo do distribuidor do <i>dataset</i> pode ser visualizado.   | Opcional               | URL do logotipo da Embrapa:<br><a href="https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844">https://www.embrapa.br/image/layout_set_logo?img_id=1343656&amp;t=1591812010844</a><br>URL do logotipo do Laboratório Multiusuário de Bioinformática da Embrapa:<br><a href="https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png">https://www.dataverse-h.cnptia.embrapa.br/logos/8/layout_set_logo.png</a> |
| <b>Data de distribuição</b>         | Data em que o <i>dataset</i> foi disponibilizado para distribuição.  | Recomendado            | 2020-01-15   |
| <b>Depositante</b>                  | Nome pessoal ou corporativo do depositante do <i>dataset</i> .   | Requerido              | Nome pessoal:<br>José Silva<br>Nome corporativo:<br>Empresa Brasileira de Pesquisa Agropecuária  |

Continua...

|                                      |   |               |   |
|--------------------------------------|---|---------------|---|
| <b>Data do depósito</b>              | Data em que o <i>dataset</i> foi depositado no repositório.                             | Recomendado   | 2020-06-11  |
| <b>Período da cobertura de dados</b> | Período de tempo ao qual os dados se referem ou período de tempo abrangido pelos dados. | Não se aplica | Não se aplica   |
| Data de início                       | Data de início do período de cobertura dos dados.                                       | Recomendado   | 2015-05-01  |
| Data de Término                      | Data de término do período da cobertura dos dados.                                      | Recomendado   | 2015-06-22  |
| <b>Período da coleta de dados</b>    | Período em que os dados foram coletados.  | Não se aplica | Não se aplica   |
| Data de início                       | Data de início do período da coleta dos dados.  | Recomendado   | 2015-05-01  |
| Data de Término                      | Data de término do período da coleta de dados.  | Recomendado   | 2015-06-22  |
| <b>Tipos de dados</b>                | Tipo de dados incluídos no <i>dataset</i> .   | Recomendado   | Código-fonte<br>Dados agregados<br>Dados de pesquisa<br>Dados experimentais<br>Dados biológicos<br>Dados genômicos<br>Dados de sequência <i>RNA</i> |
| <b>Série</b>                         | Informação sobre a série do <i>dataset</i> .  | Não se aplica | Não se aplica   |
| Nome da série                        | Nome de série ao qual o <i>dataset</i> está vinculado                                   | Opcional      | Portfólio de Sanidade Animal<br>Manejo de Pragas e Doenças  |
| Informação sobre a série             | Informação sucinta que descreve as características da série.                            | Opcional      | A Série Portfólio Sanidade Animal e Manejo de Pragas e Doenças abarca o <i>dataset</i> produzidos no âmbito do Projeto RmVAC.                       |
| <b>Software</b>                      | Informações sobre o software usado para gerar o <i>dataset</i> .                        | Não se aplica | Não se aplica   |
| Nome do software                     | Nome do software utilizado para gerar o <i>dataset</i> .                                | Recomendado   | SAS<br><i>Scripts</i><br>Software R<br><i>Trinity</i>   |
| Versão do software                   | Versão do software.   | Recomendado   | Versão 2.8.4<br>Versão 0.01   |
| <b>Material relacionado</b>          | Outro material relacionado ao <i>dataset</i> que está sendo descrito.                   | Opcional      | Projetos anteriores e/ou outros materiais relacionados ao <i>dataset</i> .<br><i>Datasets</i> anteriores  |

Continua...

|  |   |          |   |
|--|---|----------|---|
| <b>Datasets relacionados</b>                   | Quaisquer <i>datasets</i> relacionados ao <i>dataset</i> que está sendo descrito, tais como <i>datasets</i> provenientes de pesquisas anteriores sobre esse assunto.  | Opcional | Não há.   |
| <b>Outras referências</b>                      | Quaisquer outras referências que podem ter servido de material de apoio ao <i>dataset</i> .   | Opcional | Acesso e descoberta de conhecimento produzido por pesquisadores de instituições parceiras do Projeto RmVAC.   |
| <b>Fontes de dados</b>                         | Referem-se a materiais que serviram de fontes para o <i>dataset</i> , como: lista de livros, artigos, publicações em série ou arquivos de dados legíveis por máquina. | Opcional | CSORDAS, B. G.; CUNHA, R. C.; GARCIA, M. V.; SILVA, S. S. da; LEITE, F. L.; ANDREOTTI, R. Molecular characterization of the recombinant protein RmLTIBmCG-LTB: protective immunity against <i>Rhipicephalus (Boophilus) microplus</i> . <b>PLoS One</b> , v. 13, n. 2, p. 1-14, 2018. DOI: 10.1371/journal.pone.0191596 |
| <b>Origem das fontes de dados</b>              | Para materiais históricos, informar a origem das fontes e as regras seguidas no estabelecimento das fontes do <i>dataset</i> .  | Opcional | Literatura que retrata o estado da arte.  |
| <b>Características das fontes observadas</b>   | Características e material de origem do <i>dataset</i> .  | Opcional | Não há  |
| <b>Documentação e acesso a fontes de dados</b> | Nível de documentação do material de origem. Informação sobre quaisquer restrições de acesso aos dados de origem.   | Opcional | Não há  |

Fonte: Souza et al. (2020).

## Descrição de Elementos do Esquema Metadados Ciência da Vida

**Tabela 3.** Descrição dos elementos do esquema Metadados Ciência da Vida (campo e subcampos) no Redape.

| Campos                | Definição                                | Exigência de preenchimento | Exemplos   |
|-----------------------|--|----------------------------|--|
| <b>Tipo de design</b> | Definição geral do desenho experimental. | Opcional                   | Case control; Cross sectional; Cohort study; Nested case control design; Not specified; Parallel group design; Perturbation design; Randomized controlled trial; Technological design. |

Continua...

|                               |  |                |  |
|-------------------------------|--|----------------|--|
| <b>Tipo de fator</b>          | Fatores utilizados no <i>dataset</i> . Fator experimental; estresse hídrico; estágio de desenvolvimento; tipo de célula. | Opcional       | Age; Biomarkers; Cell surface markers; Cell type/cell line; Developmental stage; Genetic characteristics; Treatment type; etc. |
| <b>Organismo</b>              | Nome taxonômico do organismo usado no <i>dataset</i> ou de onde o material biológico se originou                         | Opcional       | <i>Bos taurus</i><br><i>Escherichia coli</i><br><i>Rhipicephalus microplus</i><br>Carrapato do boi                             |
| <b>Outro organismo</b>        | Não se aplica.   | Não se aplica. | Não se aplica.   |
| <b>Tipo de medida</b>         | Um termo para qualificar o que está sendo medido.  | Opcional       | Cell counting; DNA methylation profiling; Genome sequencing; Protein-RNA binding (RIP-Seq); etc.                               |
| <b>Outro tipo de medida</b>   | Não se aplica.   | Não se aplica. | Não se aplica.   |
| <b>Tipo de tecnologia</b>     | Um termo usado para identificar a tecnologia usada para coletar os dados.  | Opcional       | DNA microarray; Mass spectrometry; Nucleotide sequencing; Protein microarray; etc.   |
| <b>Plataforma tecnológica</b> | O fabricante e o nome da plataforma usada no experimento.  | Opcional       | Illumina<br>HiSeq2500.   |
| <b>Tipo de célula</b>         | Nome da linhagem de células das quais as amostras derivam.   | Opcional       | Larvas; Ninfas; Ovários; Intestinos; Glândula salivar etc.   |

Fonte: Souza et al. (2020).

## Referências

CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004. 1 v. Paginação irregular.

GO FAIR. **FAIR principles**. [2018]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 19 out. 2021.

RIBEIRO, A. M. de C. M. **AACR2**: Anglo-American cataloguing rules, 2nd edition: descrição e pontos de acesso. Brasília, DF: Ed. do Autor, 1995. 577 p.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pela AACR2 2002**: Anglo-American Cataloguing Rules, 2<sup>nd</sup> edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

SOUZA, M. I. F.; VISOLI, M. C.; TORRES, T. Z. **Catálogo de dataset no Repositório de Dados da Embrapa**: a experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa. Campinas: Embrapa Informática Agropecuária, 2020. 116 p. (Embrapa Informática Agropecuária. Documentos, 172). Disponível em:

<<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1127941/1/Doc172-2020.pdf>>.  
Acesso em 23 ago. 2021.

# APÊNDICE C – Revisão na Catalogação de *Datasets* Ômicos no Redape

## Introdução

Este Apêndice C traz orientações para a execução da atividade de revisão na catalogação de *datasets* ômicos, no Repositório de Dados de Pesquisa da Embrapa (Redape). Tais orientações foram elaboradas no âmbito do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB), no escopo do projeto de pesquisa “Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus microplus* e espécies relacionadas” (Projeto RmVAC), com o objetivo de apoiar o trabalho de revisores (bibliotecários e/ou cientistas da informação) na análise de qualidade na catalogação de *datasets*.

Neste Apêndice serão abordados os seguintes tópicos:

- Catalogação de *datasets* no Redape.
- Por que revisar a catalogação de *dataset* no Redape?
- Controle de qualidade e garantia de qualidade.
- Revisão na catalogação de *datasets* no Redape.
- Atividades de revisão na catalogação de *datasets*.

## Catalogação de *datasets* ômicos no Redape

A catalogação de *datasets* é uma atividade essencialmente descritiva, inerente à etapa ‘Descrição’, constantes nos principais modelos de ciclo de vida de dados, como pode ser observado em estudos realizados por Ball (2012) e Sinaeepourfar et al. (2015). A catalogação é uma atividade que, além de descrever, visa conferir qualidade tanto aos metadados quanto aos dados (ou *datasets*).

Esta atividade capaz de gerar impactos em outros processos dela decorrentes, por exemplo, na indexação, na busca e na recuperação de informações, para os quais os vocabulários controlados para tratamento de assuntos tornam-se imprescindíveis e de uso fortemente recomendado para a atribuição de palavra-chave e classificação (ver Apêndice A).

No âmbito das ações de gestão de dados de pesquisa no LMB, a equipe do projeto escolheu o processo de catalogação para ser o objeto de análise e melhorias, visando sua eficiência e eficácia. Esse processo foi desenvolvido no escopo do projeto RmVAC, um projeto-piloto por meio do qual a catalogação de *datasets* foi exercitada no Redape, utilizando a ferramenta Dataverse.

O processo de catalogação de *datasets* ômicos no ambiente do software Dataverse utiliza um formulário para descrição dos metadados e upload de arquivos de dados. O esquema Metadados de Citação, predefinido pelo software, é utilizado para a catalogação de *datasets*. Por conseguinte, este é o esquema genérico de metadados adotado no LMB para a catalogação de *datasets*. Também foi escolhido o esquema Metadados Ciência da Vida, um conjunto de metadados específicos voltados para área biológica. Desse modo, o esquema Metadados de Citação e o esquema Metadados Ciência da Vida podem ser utilizados concomitantemente, desde que o *dataverse*<sup>20</sup> ao qual o *dataset* estiver vinculado, tenha sido criado com essa possibilidade.

Para a representação descritiva, como referência, são adotadas as regras mínimas da AACR2 (Ribeiro, 1995; 2003; Código..., 2004); e para a representação temática são adotados os vocabulários controlados AGROVOC *Thesaurus* (FAO, 2021), NAL *Thesaurus* (Estados Unidos, 2021) e NCBI *Taxonomy* (National Center for Biotechnology Information, 2021).

## **Por que revisar a catalogação de *datasets* ômicos no Redape?**

A busca de qualidade em catalogação deve ser assegurada por meio da adoção de padrões mínimos e universais, sobretudo, de descrição e indexação, os quais contribuem para a coerência e a consistência dos registros em repositórios e para a efetiva recuperação. (Souza; Mostafa, 1999 citado por Branco et al., 2014).

De acordo com Graham (1990), completeza e precisão do registro são os elementos que caracterizam a qualidade da catalogação. A completeza expressa o quanto os metadados representam o documento descrito. A precisão diz respeito ao corretismo das informações fornecidas na descrição do recurso digital (Morris; Wool, 1999).

---

<sup>20</sup> *Dataverse* é aqui entendido como um projeto que reúne conjuntos de dados (*datasets*).

A revisão na catalogação de *datasets* é uma atividade que contribui para que os metadados se tornem aderentes aos princípios FAIR, e, conseqüentemente, possam usufruir dos benefícios de serem encontráveis, acessíveis, interoperáveis e reutilizáveis tanto para humanos como para máquinas.

## **Controle de qualidade e garantia de qualidade**

Controle de qualidade diz respeito à avaliação de um produto, processo, atividade ou serviço. Por sua vez, a garantia de qualidade visa assegurar que determinado produto, serviço ou processo atinja níveis mínimos de qualidade. Ou seja, a garantia da qualidade está empenhada em fazer com que produto, processo, atividade ou serviço atinja de forma eficiente os objetivos traçados. (Branco et al., 2014).

No tocante a determinado processo, por exemplo, o controle de qualidade preocupa-se em assegurar o atendimento aos requisitos técnicos específicos, como também ser confiável e satisfatório para os usuários. Desse modo, o processo de controle de qualidade envolve, necessariamente, a atividade de análise e identificação de atividades e de processos que possam estar em descumprimento de normas específicas estabelecidas para a execução. (Branco et al., 2014).

Após a identificação do processo ou atividade para o qual se pretende exercer o controle sobre sua qualidade, Branco et al. (2014) recomendam os seguintes passos:

- Definir requisitos e critérios para avaliar a efetividade do controle de qualidade.
- Estabelecer diretrizes para o exercício do controle de qualidade.
- Monitorar a manutenção e o desenvolvimento da atividade de controle de qualidade.
- Qualificar os sujeitos envolvidos no processo de controle de qualidade.
- Capacitar profissionais para a melhoria de produto, serviço, atividade e processo.

A garantia de qualidade visa assegurar que determinado produto, serviço ou processo atinja níveis mínimos de qualidade. No tocante ao processo de catalogação de *datasets*, a descrição padronizada de metadados é fator crítico de garantia de qualidade.

## Revisão de catalogação de *datasets* ômicos no Redape

A catalogação de *datasets* ômicos no Repositório de Dados de Pesquisa da Embrapa é realizada no ambiente do software Dataverse. Para tanto, são adotados padrões internacionais para a representação descritiva e vocabulários controlados para a representação temática, mencionados anteriormente.

A busca de qualidade na catalogação de *datasets* deve ser assegurada por medidas de controle voltadas ao atendimento a padrões e regras mínimas de catalogação, no tocante ao preenchimento de elementos, campos e subcampos do formulário de catalogação no software Dataverse. Todavia, é necessário seguir as orientações para a catalogação de *datasets* apresentadas neste documento<sup>21</sup>, bem como se atentar aos elementos de controle de qualidade para a revisão da catalogação realizada.

Baseados em Branco et al. (2014) e Souza et al. (2020), foram escolhidos os principais elementos necessários ao controle de qualidade na catalogação de *datasets* ômicos no Redape. Os elementos de controle de qualidade para a revisão da catalogação foram:

- Completeza das informações registradas sobre o *dataset*.
- Precisão da descrição quanto à ortografia, pontuação, descrição textual, transcrições de URI, URL, DOI, ORCID etc.
- Clareza e integridade das informações registradas sobre o *dataset*.
- Fidelidade aos padrões e normas adotados para a catalogação de *dataset*.
- Confiabilidade das informações registradas sobre o *dataset*.
- Coerência e consistência dos dados e informações registradas sobre o *dataset*.

É importante ressaltar a necessidade de adoção de padrões e orientações técnicas prescritas para a catalogação no Redape, para que sejam obtidos níveis aceitáveis de qualidade das informações descritas, contidas nos metadados catalogados, o que otimizaria a descoberta e, conseqüentemente, o reuso dos dados catalogados.

---

<sup>21</sup> O documento referido corresponde ao conteúdo na íntegra dos tópicos 1, 2, 3 e 4, bem como aos Apêndices A e B, desta publicação

## Atividades de revisão na catalogação de *datasets* ônicos no Redape

Dada a natureza da atividade, o controle de qualidade em catalogação de *datasets* ônicos deve ser executado por profissionais de Biblioteconomia e/ou Ciência da Informação. A revisão envolve a análise acurada do preenchimento de cada elemento, campo e subcampo do formulário de catalogação no software Dataverse.

A tarefa de revisão e análise da catalogação será guiada pelos elementos de controle de qualidade apresentados anteriormente, devendo o revisor observar as seguintes etapas de controle de qualidade na catalogação de *datasets* ônicos (Souza et al., 2020).

- Examinar todas as informações registradas sobre o *dataset*.
- Descrever todos os erros, omissões e/ou inconsistências encontrados.
- Notificar o catalogador (e/ou autor de dados).
- Solicitar correção de erros e/ou inconsistências identificados.
- Manter o registro de correções efetuadas.
- Encaminhar relatório para o catalogador.

A atividade de revisão da catalogação de *datasets* ônicos deverá ser realizada em conformidade com as matrizes de revisão apresentadas nas Tabelas 1 e 2, a seguir. Os apontamentos para a correção da catalogação poderão ser registrados em formulário próprio, conforme Modelo de Relatório de Revisão na Catalogação de *Datasets* Ônicos (Apêndice D), indicando-se para cada elemento (campo/subcampo), a descrição do erro e/ou inconsistência, a orientação para correção e os impactos na qualidade.

**Tabela 1.** Matriz de atividades de revisão na catalogação de *datasets* ônicos no Redape – Esquema Metadados de Citação.

| Elemento (campo/subcampo) | Descrição do erro e/ou inconsistência                      | Orientação para correção  | Elemento(s) de impacto na qualidade | Documento orientador |
|---------------------------|--|---|-------------------------------------|----------------------|
| Título                    | Ausência de clareza e integridade na atribuição do título. | Atribuir título observando: Tipo de estudo; Amostra estudada; Espécie estudada. | Completeza. Clareza. Integridade.   | Esta publicação      |
| Subtítulo                 | Ausência de informação do subtítulo.                       | Informar subtítulo.   | Completeza.                         | Esta publicação      |

|                            |   |  |   |                 |
|----------------------------|---|--|---|-----------------|
| Título alternativo         | Ausência de informação do título alternativo.   | Informar título alternativo.   | Completeza.   | Esta publicação |
| <b>URL alternativa</b>     | Ausência da URL alternativa.  | Informar URL alternativa válida.   | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Outro identificador</b> | Não se aplica   | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Agência                    | Ausência do nome da agência do outro identificador.   | Informar nome da agência do outro identificador.   | Completeza.   | Esta publicação |
| Identificador              | Ausência de identificador digital. Identificador digital não localizado.  | Informar identificador digital válido.   | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Autor</b>               | Não se aplica   | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Nome                       | Ausência de padronização na descrição do nome.  | Seguir regras de catalogação.  | Fidelidade a padrões.                               | Esta publicação |
| Afiliação                  | Ausência de padrão na descrição de nome corporativo.  | Seguir regras de catalogação.  | Fidelidade a padrões.                               | Esta publicação |
| Sistema do identificador   | Ausência de indicação de sistema identificador.   | Informar nome do sistema identificador.  | Completeza.   | Esta publicação |
| Identificador              | Ausência de identificador persistente válido.   | Informar nome identificador persistente.   | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Contato</b>             | Não se aplica   | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Nome                       | Ausência de padronização na descrição do nome.  | Seguir regras de catalogação.  | Fidelidade a padrões.                               | Esta publicação |
| Afiliação                  | Ausência de padrão na descrição de nome corporativo.  | Seguir regras de catalogação.  | Fidelidade a padrões.                               | Esta publicação |
| E-mail do contato          | E-mail incompleto.  | Informar e-mail válido.  | Precisão.   | Esta publicação |
| <b>Descrição</b>           | Não se aplica   | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Texto                      | Ausência de clareza e coerência textual no campo Texto.<br>Ausência de precisão na descrição quanto à ortografia, pontuação transcrição textual na descrição. | Revisar a redação do campo Texto, observando clareza, coerência e precisão quanto à ortografia, pontuação e transcrição textual. | Clareza e integridade.<br>Coerência e consistência. | Esta publicação |

|   |  |  |   |                 |
|---|--|--|---|-----------------|
| Data                                    | Ausência de informação da Data de descrição.   | Informar Data de descrição.  | Completeza.   | Esta publicação |
| <b>Assunto</b>                          | Ausência de área(s) de domínio representativa do assunto.  | Adicionar outra(s) área(s) de domínio.   | Completeza  | Esta publicação |
| <b>Palavras-chave</b>                   | Não se aplica  | Não se aplica  | Não se aplica   | Não se aplica   |
| Termo-chave                             | Ausência (falta) de termos-chave. Termo-chave selecionado não pertence ao escolhido.   | Inserir mais termos-chave. Corrigir transcrição de termos-chave.   | Completeza. Confiabilidade da informação. Coerência e consistência da informação. | Esta publicação |
| Vocabulário controlado do termo-chave   | Ausência (falta) de indicação de nome(s) do(s) vocabulário(s) controlado(s) usado(s) para o(s) termo(s)-chave.                                       | Inserir o nome do vocabulário(s) controlado(s) para usado para indicar o(s) termo(s)-chave.                            | Completeza da informação.   | Esta publicação |
| URL do vocabulário controlado           | Ausência da URL do termo-chave, que deve inserida neste campo URL do vocabulário controlado. URL do termo-chave não localizado.                      | Informar URL válida do termo-chave.  | Completeza. Precisão.   | Esta publicação |
| <b>Classificação</b>                    | Não se aplica  | Não se aplica  | Não se aplica   | Não se aplica   |
| Termo de classificação                  | Ausência de termos de classificação. Termo de classificação escolhido não corresponde ao assunto do <i>dataset</i> .                                 | Adicionar mais termos de classificação. Eliminar termo de classificação não correspondente ao tema do <i>dataset</i> . | Completeza. Confiabilidade da informação. Coerência e consistência da informação. | Esta publicação |
| Vocabulário controlado da classificação | Ausência de nome(s) de vocabulário(s) controlado(s).   | Informar nome(s) de vocabulário(s) controlado(s).  | Completeza. Precisão. Fidelidade a padrões.                                       | Esta publicação |
| URL do vocabulário controlado           | Ausência da URL de termo(s) de classificação, que deve inserida neste campo URL do vocabulário controlado. URL do termo de classificação não válida. | Adicionar URL de termo(s) de classificação válida. Informar URL do termo válida.                                       | Completeza. Precisão.   | Esta publicação |

|                                 |   |   |  |   |
|---------------------------------|---|---|--|---|
| <b>Publicação relacionada</b>   | Não se aplica   | Não se aplica   | Não se aplica                                    | Não se aplica                                   |
| Citação                         | Ausência de padronização na referência da citação.  | Seguir regras de referência bibliográfica.  | Fidelidade a padrões.                            | Associação Brasileira de Normas Técnicas (2020) |
| Tipo de identificador           | Ausência de indicação de tipo de identificador.   | Informar tipo de identificador.   | Completeza.                                      | Esta publicação                                 |
| Número do identificador         | Ausência de número de identificador válido.   | Informar nome identificador válido.   | Completeza. Precisão.                            | Esta publicação                                 |
| URL da publicação               | Ausência da URL da publicação.  | Informar URL válida da publicação.  | Completeza. Precisão.                            | Esta publicação                                 |
| <b>Notas</b>                    | Ausência de clareza e coerência textual da nota. Ausência de precisão, quanto à ortografia, pontuação no texto da nota. | Revisar a redação de nota, observando clareza, coerência e precisão textual quanto à ortografia, pontuação. | Clareza e integridade. Coerência e consistência. | Esta publicação                                 |
| <b>Idioma</b>                   | Ausência de informação sobre o idioma <i>dataset</i> .  | Informar o idioma do <i>dataset</i> .   | Completeza.                                      | Esta publicação                                 |
| <b>Produtor</b>                 | Não se aplica   | Não se aplica   | Não se aplica                                    | Não se aplica                                   |
| Nome do produtor                | Ausência de padronização na indicação do nome do produtor.  | Seguir regras de catalogação.   | Fidelidade a padrões.                            | Esta publicação                                 |
| Afiliação do produtor           | Ausência de padrão na descrição de nome corporativo.  | Seguir regras de catalogação.   | Fidelidade a padrões.                            | Esta publicação                                 |
| Abreviatura do nome do produtor | Ausência de abreviatura do nome do produtor.  | Informar a abreviatura do nome do produtor.   | Fidelidade a padrões.                            | Esta publicação                                 |
| URL do produtor                 | Ausência da URL do produtor.  | Informar URL válida do produtor.  | Completeza. Precisão.                            | Esta publicação                                 |
| URL do logotipo do produtor     | Ausência da URL do logotipo do produtor.  | Informar URL válida do logotipo do produtor.  | Completeza. Precisão.                            | Esta publicação                                 |
| <b>Data de produção</b>         | Ausência de data de produção.   | Informar data de produção.  | Completeza. Precisão.                            | Esta publicação                                 |
| <b>Local de produção</b>        | Ausência de local de produção.  | Informar local de produção.   | Completeza. Precisão.                            | Esta publicação                                 |

|                                      |   |  |                       |                 |
|--------------------------------------|---|--|-----------------------|-----------------|
| <b>Colaborador</b>                   | Não se aplica   | Não se aplica  | Não se aplica         | Esta publicação |
| Tipo de colaborador                  | Ausência de informação sobre o tipo de colaborador.   | Selecionar tipo de colaborador.  | Completeza. Precisão. | Esta publicação |
| Nome do colaborador                  | Ausência de padronização na indicação do nome do colaborador.   | Seguir regras de catalogação.  | Fidelidade a padrões. | Esta publicação |
| <b>Financiamento</b>                 | Não se aplica   | Não se aplica  | Não se aplica         | Não se aplica   |
| Nome do órgão financiador            | Ausência do nome do órgão financiador.  | Indicar nome do órgão financiador.   | Completeza. Precisão. | Esta publicação |
| Número do contrato                   | Ausência do número do contrato de financiamento.  | Informar número de contrato de financiamento.  | Completeza. Precisão. | Esta publicação |
| <b>Distribuidor</b>                  | Não se aplica   | Não se aplica  | Não se aplica         | Não se aplica   |
| Nome do distribuidor                 | Ausência do nome do distribuidor do <i>dataset</i> . Ausência de padronização na indicação do nome do distribuidor. | Indicar nome do distribuidor. Seguir regras de catalogação na indicação do nome do distribuidor. | Completeza. Precisão. | Esta publicação |
| Afiliação do distribuidor            | Ausência de padrão na descrição do nome da instituição ao qual o distribuidor está vinculado                        | Seguir regras de catalogação.  | Fidelidade a padrões. | Esta publicação |
| Abreviatura do nome do distribuidor  | Ausência de abreviatura do nome do distribuidor.  | Informar a abreviatura do nome do distribuidor.  | Fidelidade a padrões. | Esta publicação |
| URL do distribuidor                  | Ausência da URL do distribuidor.  | Informar URL válida do distribuidor.   | Completeza. Precisão. | Esta publicação |
| URL do logotipo do distribuidor      | Ausência da URL do logotipo do distribuidor.  | Informar URL válida do logotipo do distribuidor.   | Completeza. Precisão. | Esta publicação |
| <b>Data de distribuição</b>          | Ausência de data de distribuição.   | Informar data de distribuição.   | Completeza. Precisão. | Esta publicação |
| <b>Depositante</b>                   | Ausência de padronização na indicação do nome do depositante.   | Seguir regras de catalogação na indicação do nome do depositante.                                | Completeza. Precisão. | Esta publicação |
| <b>Data do depósito</b>              | Ausência de data de depósito.   | Informar data de depósito.   | Completeza. Precisão. | Esta publicação |
| <b>Período da cobertura de dados</b> | Não se aplica   | Não se aplica  | Não se aplica         | Não se aplica   |

|                                   |  |  |   |                 |
|-----------------------------------|--|--|---|-----------------|
| Data de início                    | Ausência de data de início da cobertura dos dados.   | Informar data de início da cobertura dos dados.  | Completeza.<br>Precisão.                            | Esta publicação |
| Data de término                   | Ausência de data de término da cobertura dos dados.  | Informar data de término da cobertura dos dados.   | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Período de coleta de dados</b> | Não se aplica  | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Data de início                    | Ausência de data de início da coleta dos dados.  | Informar data de início da coleta dos dados.   | Completeza.<br>Precisão.                            | Esta publicação |
| Data de término                   | Ausência de data de término da coleta dos dados.   | Informar data de término da coleta dos dados.  | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Tipo de dados</b>              | Ausência de informação do(s) tipo(s) de dados do <i>dataset</i> .  | Informar tipo(s) de dados do <i>dataset</i> .  | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Série</b>                      | Não se aplica  | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Nome da série                     | Ausência do nome da série do <i>dataset</i> .  | Informar nome da série, caso exista.   | Completeza.<br>Precisão.                            | Esta publicação |
| Informação sobre série            | Ausência de informações sobre a série do <i>dataset</i> . Ausência de clareza e coerência textual na informação sobre a série. | Fornecer informações sucintas sobre a série, caso exista. Revisar redação da informação sobre a série. | Clareza e integridade.<br>Coerência e consistência. | Esta publicação |
| <b>Software</b>                   | Não se aplica  | Não se aplica  | Não se aplica                                       | Não se aplica   |
| Nome de software                  | Ausência do nome do software utilizado.  | Indicar o nome do software   | Completeza.<br>Confiabilidade das informações.      | Esta publicação |
| Versão do software                | Ausência da informação sobre a versão do software utilizado.   | Indicar o nome do software.  | Completeza.<br>Confiabilidade das informações.      | Esta publicação |
| <b>Material relacionado</b>       | Ausência de informação sobre material relacionado.   | Informar material relacionado ao <i>dataset</i> , caso exista.   | Completeza.   | Esta publicação |
| <b>Datasets relacionados</b>      | Ausência de informação de <i>datasets</i> relacionados.  | Informar <i>datasets</i> relacionados, caso existam.   | Completeza.   | Esta publicação |

|   |  |   |             |                 |
|---|--|---|-------------|-----------------|
| <b>Outras referências</b>                             | Ausência de informação sobre outras referências ao <i>dataset</i> .                      | Informar outras referências relacionadas ao <i>dataset</i> , caso existam.              | Completeza. | Esta publicação |
| <b>Fontes de dados</b>                                | Ausência de informação sobre fontes de dados do <i>dataset</i> .                         | Informar fontes de dados do <i>dataset</i> , caso existam.                              | Completeza. | Esta publicação |
| <b>Origem das fontes de dados</b>                     | Ausência de informação sobre origem das fontes dos dados do <i>dataset</i> .             | Informar origem das fontes de dados do <i>dataset</i> , caso exista.                    | Completeza. | Esta publicação |
| <b>Características das fontes de dados observadas</b> | Ausência de informação sobre características das fontes de dados observadas.             | Informar sobre características das fontes de dados observadas, caso existam.            | Completeza. | Esta publicação |
| <b>Documentação e acesso a fontes de dados</b>        | Ausência de informação sobre documentação e acesso a fontes de dados do <i>dataset</i> . | Informar sobre documentação e acesso a fontes de dados do <i>dataset</i> , caso exista. | Completeza. | Esta publicação |

Fonte: Adaptado de Souza et al. (2020).

**Tabela 2.** Matriz de atividades de revisão na catalogação de *datasets* ômicos no Redape - Esquema Metadados Ciência da Vida.

| <b>Elemento (campo)</b>     | <b>Descrição do erro e/ou inconsistência</b> | <b>Orientação para correção</b> | <b>Elemento(s) de impacto na qualidade</b> | <b>Documento orientador</b> |
|-----------------------------|--|---------------------------------|--|-----------------------------|
| <b>Tipo de design</b>       | Ausência de informação.                      | Selecionar tipo de design.      | Completeza.                                | Esta publicação             |
| <b>Tipo de fator</b>        | Ausência de informação.                      | Selecionar tipo de fator.       | Completeza.                                | Esta publicação             |
| <b>Organismo</b>            | Ausência de informação.                      | Selecionar organismo.           | Completeza.                                | Esta publicação             |
| <b>Outro organismo</b>      | Não se aplica                                | Não se aplica                   | Não se aplica                              | Não se aplica               |
| <b>Tipo de medida</b>       | Ausência de informação.                      | Selecionar tipo de medida.      | Completeza.                                | Esta publicação             |
| <b>Outro tipo de medida</b> | Não se aplica                                | Não se aplica                   | Não se aplica                              | Não se aplica               |
| <b>Tipo de tecnologia</b>   | Ausência de informação.                      | Selecionar tipo de tecnologia.  | Completeza.                                | Esta publicação             |

Continua...

|                               |                         |                                    |             |                 |
|-------------------------------|-------------------------|------------------------------------|-------------|-----------------|
| <b>Plataforma tecnológica</b> | Ausência de informação. | Selecionar plataforma tecnológica. | Completeza. | Esta publicação |
| <b>Tipo de célula</b>         | Ausência de informação. | Informar tipo de célula.           | Completeza. | Esta publicação |

Fonte: Adaptada de Souza et al. (2020).

O arquivo do relatório gerado pelo revisor, com apontamentos para a correção da catalogação (Apêndice D), deverá ser anexado junto aos demais arquivos do *dataset* em revisão, diretamente no Redape. Ao catalogador (e/ou autor do *dataset*) caberá analisar, discutir e fazer as correções solicitadas pelo revisor, devolvendo ao sistema o arquivo do relatório por ele corrigido.

Concluída a etapa de revisão na catalogação, o catalogador poderá fazer a publicação do *dataset* no Redape.

## Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação – referências: elaboração. 2. ed. Rio de Janeiro, 2020. 68 p. Versão corrigida 2.

BALL, A. **Review of data management lifecycle models**. Bath University of Bath, 2012. 15 p.

BRANCO, Z. de S.; MACHADO, D. R.; CESTARI, B. H. P. de S.; OLIVEIRA, Z. P. Controle de qualidade em catalogação cooperativa. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 18., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014. 21 p.  
CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB: Imprensa Oficial do Estado de São Paulo, 2004. 1 v. Paginação irregular.

ESTADOS UNIDOS. Department of Agriculture. **NAL Agricultural Thesaurus and Glossary**. Disponível em: <<https://agclass.nal.usda.gov/agt.shtml>>. Acesso em: 19 out. 2021.

FAO. **AGROVOC**. Rome, 2021. Disponível em: <<https://agrovoc.fao.org/browse/agrovoc/en/>>. Acesso em: 19 out. 2021.

GRAHAM, P. S. Quality in cataloging: making distinctions. **The Journal of Academic Librarianship**, v. 16, n. 4, p. 213-218, Sept.1990.

MORRIS, D. E.; WOOL, G. Cataloging: librarianship's best bargain. **Library Journal**, v. 124, n. 11, p. 44-46, 1999.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **Taxonomy**. Bethesda. Disponível em: <<https://www.ncbi.nlm.nih.gov/taxonomy>>. Acesso em: 19 out. 2021.

RIBEIRO, A. M. de C. M. **AACR2**: Anglo-American cataloguing rules, 2nd edition: descrição e pontos de acesso. Brasília, DF: Ed. do Autor, 1995. 577 p.

RIBEIRO, A. M. de C. M. **Catálogo de recursos bibliográficos pela AACR2 2002**: Anglo-American Cataloguing Rules, 2<sup>nd</sup> edition, 2002 revision. Brasília, DF: Ed. do Autor, 2003. 1 v. Paginação irregular.

SINAEPOURFAR, A.; MASIP-BRUIN, X.; GARCIA, J.; MARÍN-TORDERA, E. **A survey on data lifecycle models**: discussions toward the 6Vs challenges. [Barcelona: Universitat Politècnica de Catalunya, 2015]. Technical Report, UPC-DAC-RR-2015-18.

SOUZA, M. I. F.; VISOLI, M. C.; TORRES, T. Z. **Catálogo de dataset no Repositório de Dados da Embrapa**: a experiência do Projeto-piloto de Implantação de Gestão de Dados de Pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa. Campinas: Embrapa Informática Agropecuária, 2020. 116 p. (Embrapa Informática Agropecuária. Documentos, 172). Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1127941/1/Doc172-2020.pdf>>. Acesso em 23 ago. 2021.

## APÊNDICE D – Modelo de Relatório de Revisão na Catalogação de *Datasets* Ômicos

### Apontamentos para a correção da catalogação de *datasets* ômicos

**Projeto:** RmVac – Prática para estruturação de *dataverse* com sub *dataverses* [exemplo]

**Dataverse:** Solução de inovação 1 do *dataverse* RmVAC – Prática 2 (S1-RmVAC-Pratica2) [exemplo]

**Dataset:** <https://doi.org/10.80272/NUT6RT> [exemplo]

As orientações contidas nas Tabelas 1 e 2, apresentadas a seguir, devem ser utilizadas para informar ao catalogador as correções necessárias para a revisão na catalogação de *datasets* ômicos, respectivamente, no Esquema Metadados de Citação e no Esquema Metadados Ciência da Vida.

**Tabela 1.** Orientações para a revisão na catalogação de *datasets* ômicos no Redape - Esquema Metadados de Citação.

| Elemento (campo/subcampo) | Descrição do erro e/ou inconsistência                      | Orientação para correção  | Elemento(s) de impacto na qualidade     | Documento orientador |
|---------------------------|--|---|---|----------------------|
| <b>Título</b>             | Ausência de clareza e integridade na atribuição do título. | Atribuir título observando: Tipo de estudo; Amostra estudada; Espécie estudada. | Completeza.<br>Clareza.<br>Integridade. | Esta publicação      |
| Subtítulo                 | Ausência de informação do subtítulo.                       | Informar subtítulo.   | Completeza.                             | Esta publicação      |
| Título alternativo        | Ausência de informação do título alternativo.              | Informar título alternativo.  | Completeza.                             | Esta publicação      |
| <b>URL alternativa</b>    | Ausência da URL alternativa.                               | Informar URL alternativa válida.  | Completeza.<br>Precisão                 | Esta publicação      |

Continua...

|                            |   |  |                          |                 |
|----------------------------|---|--|--------------------------|-----------------|
| <b>Outro identificador</b> | Não se aplica   | Não se aplica                                    | Não se aplica            | Não se aplica   |
| Agência                    | Ausência do nome da agência do outro identificador.                         | Informar nome da agência do outro identificador. | Completeza.              | Esta publicação |
| Identificador              | Ausência de identificador digital.<br>Identificador digital não localizado. | Informar identificador digital válido.           | Completeza.<br>Precisão. | Esta publicação |
| <b>Autor</b>               | Não se aplica   | Não se aplica                                    | Não se aplica            | Não se aplica   |
| Nome                       | Ausência de padronização na descrição do nome.                              | Seguir regras de catalogação.                    | Fidelidade a padrões.    | Esta publicação |
| Afiliação                  | Ausência de padrão na descrição de nome corporativo.                        | Seguir regras de catalogação.                    | Fidelidade a padrões.    | Esta publicação |
| Sistema do identificador   | Ausência de indicação de sistema identificador.                             | Informar nome do sistema identificador.          | Completeza.              | Esta publicação |
| Identificador              | Ausência de identificador persistente válido.                               | Informar nome identificador persistente.         | Completeza.<br>Precisão  | Esta publicação |
| <b>Contato</b>             | Não se aplica   | Não se aplica                                    | Não se aplica            | Não se aplica   |
| Nome                       | Ausência de padronização na descrição do nome.                              | Seguir regras de catalogação.                    | Fidelidade a padrões.    | Esta publicação |
| Afiliação                  | Ausência de padrão na descrição de nome corporativo.                        | Seguir regras de catalogação.                    | Fidelidade a padrões.    | Esta publicação |

Continua...

|                                       |   |  |   |                 |
|---------------------------------------|---|--|---|-----------------|
| E-mail do contato                     | E-mail incompleto.  | Informar e-mail válido.  | Precisão.   | Esta publicação |
| <b>Descrição</b>                      | Não se aplica   | Não se aplica  | Não se aplica   | Não se aplica   |
| Texto                                 | Ausência de clareza e coerência textual no campo Texto.<br>Ausência de precisão na descrição quanto à ortografia, pontuação e transcrição textual na descrição. | Revisar a redação do campo Texto, observando clareza, coerência e precisão quanto à ortografia, pontuação e transcrição textual. | Clareza e integridade.<br>Coerência e consistência                                      | Esta publicação |
| Data                                  | Ausência de informação da Data de descrição.  | Informar Data de descrição.  | Completeza.   | Esta publicação |
| <b>Assunto</b>                        | Ausência de área(s) de domínio representativa do assunto.   | Adicionar outra(s) área(s) de domínio.   | Completeza.   | Esta publicação |
| <b>Palavras-chave</b>                 | Não se aplica   | Não se aplica  | Não se aplica   | Não se aplica   |
| Termo-chave                           | Ausência (falta) de termos-chave.<br>Termo-chave escolhido não pertence ao vocabulário escolhido.   | Inserir mais termos-chave.<br>Corrigir transcrição de termo-chave.   | Completeza.<br>Confiabilidade da informação.<br>Coerência e consistência da informação. | Esta publicação |
| Vocabulário controlado do termo-chave | Ausência (falta) de indicação de nome(s) do(s) vocabulário(s) controlado(s) usado(s) para o(s) termo(s)-chave   | Inserir o nome do vocabulário(s) controlado(s) para usado para indicar o(s) termo(s)-chave.                                      | Completeza da informação  | Esta publicação |
| URL do vocabulário controlado         | Ausência da URL do termo-chave, que deverá ser inserido neste campo (e não a URL do vocabulário controlado).<br>URL do termo-chave não localizado.              | Informar URL válida do termo-chave.  | Completeza.<br>Precisão   | Esta publicação |

Continua...

|   |   |   |   |  |
|---|---|---|---|--|
| <b>Classificação</b>                    | Não se aplica   | Não se aplica   | Não se aplica   | Não se aplica  |
| Termo de classificação                  | Ausência de termos de classificação.<br>Termo de classificação escolhido não corresponde ao assunto do <i>dataset</i> .   | Adicionar mais termos de classificação.<br>Eliminar termo de classificação não correspondente ao tema do <i>dataset</i> . | Completeza.<br>Confiabilidade da informação.<br>Coerência e consistência da informação. | Esta publicação  |
| Vocabulário controlado da classificação | Ausência de nome(s) de vocabulário(s) controlado(s).  | Informar nome(s) de vocabulário(s) controlado(s).   | Completeza.<br>Precisão.<br>Fidelidade a padrões.                                       | Esta publicação  |
| URL do vocabulário controlado           | Ausência da URL de termo(s) de classificação, que deverá ser inserido neste campo (e não a URL do vocabulário controlado).<br>URL do termo de classificação não válida. | Adicionar URL de termo(s) de classificação válida.<br>Informar URL do termo válida.                                       | Completeza.<br>Precisão.  | Esta publicação  |
| <b>Publicação relacionada</b>           | Não se aplica   | Não se aplica   | Não se aplica   | Não se aplica  |
| Citação                                 | Ausência de padronização na referência da citação.  | Seguir regras de referência bibliográfica.  | Fidelidade a padrões.   | NBR 6023: informação e documentação – referências: elaboração. |
| Tipo de identificador                   | Ausência de indicação de tipo de identificador.   | Informar tipo de identificador.   | Completeza.   | Esta publicação  |
| Número do identificador                 | Ausência de número de identificador válido.   | Informar nome identificador válido.   | Completeza.   | Esta publicação  |

|                                 |  |   |   |                 |
|---------------------------------|--|---|---|-----------------|
|                                 |  |   | Precisão.   |                 |
| URL da publicação               | Ausência da URL da publicação.   | Informar URL válida da publicação.  | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Notas</b>                    | Ausência de clareza e coerência textual da nota.<br>Ausência de precisão, quanto à ortografia, pontuação no texto da nota. | Revisar a redação de nota, observando clareza, coerência e precisão textual quanto à ortografia, pontuação. | Clareza e integridade.<br>Coerência e consistência. | Esta publicação |
| <b>Idioma</b>                   | Ausência de informação sobre o idioma <i>dataset</i> .   | Informar o idioma do <i>dataset</i> .   | Completeza.   | Esta publicação |
| <b>Produtor</b>                 | Não se aplica  | Não se aplica   | Não se aplica                                       | Não se aplica   |
| Nome do produtor                | Ausência de padronização na indicação do nome do produtor.   | Seguir regras de catalogação.   | Fidelidade a padrões.                               | Esta publicação |
| Afiliação do produtor           | Ausência de padrão na descrição de nome corporativo.   | Seguir regras de catalogação.   | Fidelidade a padrões.                               | Esta publicação |
| Abreviatura do nome do produtor | Ausência de abreviatura do nome do produtor.   | Informar a abreviatura do nome do produtor.   | Fidelidade a padrões.                               | Esta publicação |
| URL do produtor                 | Ausência da URL do produtor.   | Informar URL válida do produtor.  | Completeza.<br>Precisão.                            | Esta publicação |
| URL do logotipo do produtor     | Ausência da URL do logotipo do produtor.   | Informar URL válida do logotipo do produtor.  | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Data de produção</b>         | Ausência de data de produção.  | Informar data de produção.  | Completeza.<br>Precisão.                            | Esta publicação |

|                                     |  |   |                          |                 |
|-------------------------------------|--|---|--------------------------|-----------------|
| <b>Local de produção</b>            | Ausência de local de produção  | Informar local de produção.   | Completeza.<br>Precisão. | Esta publicação |
| <b>Colaborador</b>                  | Não se aplica  | Não se aplica   | Não se aplica            | Esta publicação |
| Tipo de colaborador                 | Ausência de informação sobre o tipo de colaborador.  | Selecionar tipo de colaborador.   | Completeza.<br>Precisão. | Esta publicação |
| Nome do colaborador                 | Ausência de padronização na indicação do nome do colaborador.  | Seguir regras de catalogação.   | Fidelidade a padrões.    | Esta publicação |
| <b>Financiamento</b>                | Não se aplica  | Não se aplica   | Não se aplica            | Não se aplica   |
| Nome do órgão financiador           | Ausência do nome do órgão financiador  | Indicar nome do órgão financiador.  | Completeza.<br>Precisão. | Esta publicação |
| Número do contrato                  | Ausência do número do contrato de financiamento.   | Informar número de contrato de financiamento.   | Completeza.<br>Precisão. | Esta publicação |
| <b>Distribuidor</b>                 | Não se aplica  | Não se aplica   | Não se aplica            | Não se aplica   |
| Nome do distribuidor                | Ausência do nome do distribuidor do <i>dataset</i> .<br>Ausência de padronização na indicação do nome do distribuidor. | Indicar nome do distribuidor.<br>Seguir regras de catalogação na indicação do nome do distribuidor. | Completeza.<br>Precisão. | Esta publicação |
| Afiliação do distribuidor           | Ausência de padrão na descrição do nome da instituição à qual o distribuidor está vinculado.                           | Seguir regras de catalogação.   | Fidelidade a padrões.    | Esta publicação |
| Abreviatura do nome do distribuidor | Ausência de abreviatura do nome do distribuidor.   | Informar a abreviatura do nome do distribuidor.   | Fidelidade a padrões.    | Esta publicação |

|                                      |   |  |                          |                 |
|--------------------------------------|---|--|--------------------------|-----------------|
| URL do distribuidor                  | Ausência da URL do distribuidor.                              | Informar URL válida do distribuidor.                           | Completeza.<br>Precisão. | Esta publicação |
| URL do logotipo do distribuidor      | Ausência da URL do logotipo do distribuidor.                  | Informar URL válida do logotipo do distribuidor.               | Completeza.<br>Precisão. | Esta publicação |
| <b>Data de distribuição</b>          | Ausência de data de distribuição.                             | Informar data de distribuição.                                 | Completeza.<br>Precisão. | Esta publicação |
| <b>Depositante</b>                   | Ausência de padronização na indicação do nome do depositante. | Seguir regras de catalogação na indicação do nome depositante. | Completeza.<br>Precisão. | Esta publicação |
| <b>Data do depósito</b>              | Ausência de data de depósito.                                 | Informar data de depósito.                                     | Completeza.<br>Precisão. | Esta publicação |
| <b>Período da cobertura de dados</b> | Não se aplica   | Não se aplica  | Não se aplica            | Não se aplica   |
| Data de início                       | Ausência de data de início da cobertura dos dados.            | Informar data de início da cobertura dos dados.                | Completeza.<br>Precisão. | Esta publicação |
| Data de término                      | Ausência de data de término da cobertura dos dados.           | Informar data de término da cobertura dos dados.               | Completeza.<br>Precisão. | Esta publicação |
| <b>Período de coleta de dados</b>    | Não se aplica   | Não se aplica  | Não se aplica            | Não se aplica   |
| Data de início                       | Ausência de data de início da coleta dos dados.               | Informar data de início da coleta dos dados.                   | Completeza.<br>Precisão. | Esta publicação |

Continua...

|                             |   |   |   |                 |
|-----------------------------|---|---|---|-----------------|
| Data de término             | Ausência de data de término da coleta dos dados.  | Informar data de término da coleta dos dados.   | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Tipo de dados</b>        | Ausência de informação do(s) tipo(s) de dados do <i>dataset</i> .   | Informar tipo(s) de dados do <i>dataset</i> .   | Completeza.<br>Precisão.                            | Esta publicação |
| <b>Série</b>                | Não se aplica   | Não se aplica   | Não se aplica                                       | Não se aplica   |
| Nome da série               | Ausência do nome da série do <i>dataset</i> .   | Informar nome da série, caso exista.  | Completeza.<br>Precisão.                            | Esta publicação |
| Informação sobre série      | Ausência de informações sobre a série do <i>dataset</i> .<br>Ausência de clareza e coerência textual na informação sobre a série. | Fornecer informações sucintas sobre a série, caso exista.<br>Revisar redação da informação sobre a série. | Clareza e integridade.<br>Coerência e consistência. | Esta publicação |
| <b>Software</b>             | Não se aplica   | Não se aplica   | Não se aplica                                       | Não se aplica   |
| Nome de software            | Ausência do nome do software utilizado.   | Indicar o nome do software  | Completeza.<br>Confiabilidade das informações.      | Esta publicação |
| Versão do software          | Ausência da informação sobre a versão do software utilizado.  | Indicar o nome do software.   | Completeza.<br>Confiabilidade das informações.      | Esta publicação |
| <b>Material relacionado</b> | Ausência de informação sobre material relacionado.  | Informar material utilizado relacionado ao <i>dataset</i> , caso exista.                                  | Completeza.   | Esta publicação |

Continua...

|   |  |   |             |                 |
|---|--|---|-------------|-----------------|
| <b>Datasets relacionados</b>                          | Ausência de informação de <i>datasets</i> relacionados.                                  | Informar <i>dataset</i> relacionados, caso existam.                                     | Completeza. | Esta publicação |
| <b>Outras referências</b>                             | Ausência de informação sobre outras referências ao <i>dataset</i> .                      | Informar outras referências relacionadas ao <i>dataset</i> , caso existam.              | Completeza. | Esta publicação |
| <b>Fontes de dados</b>                                | Ausência de informação sobre fontes de dados do <i>dataset</i> .                         | Informar fontes de dados do <i>dataset</i> , caso existam.                              | Completeza. | Esta publicação |
| <b>Origem das fontes de dados</b>                     | Ausência de informação sobre origem das fontes dos dados do <i>datasets</i> .            | Informar origem das fontes de dados do <i>dataset</i> , caso exista.                    | Completeza. | Esta publicação |
| <b>Características das fontes de dados observadas</b> | Ausência de informação sobre características das fontes de dados observadas.             | Informar sobre características das fontes de dados observadas, caso existam.            | Completeza. | Esta publicação |
| <b>Documentação e acesso a fontes de dados</b>        | Ausência de informação sobre documentação e acesso a fontes de dados do <i>dataset</i> . | Informar sobre documentação e acesso a fontes de dados do <i>dataset</i> , caso exista. | Completeza. | Esta publicação |

Fonte: Adaptado de Souza et al. (2020).

**Tabela 2.** Orientações para a revisão na catalogação de *datasets* ômicos no Redape - Esquema Metadados Ciência da Vida.

| <b>Elemento (campo)</b>     | <b>Descrição do erro e/ou inconsistência</b> | <b>Orientação para correção</b> | <b>Elemento(s) de impacto na qualidade</b> | <b>Documento orientador</b> |
|-----------------------------|--|---------------------------------|--|-----------------------------|
| <b>Tipo de design</b>       | Ausência de informação.                      | Selecionar tipo de design.      | Completeza.                                | Esta publicação             |
| <b>Tipo de fator</b>        | Ausência de informação.                      | Selecionar tipo de fator.       | Completeza.                                | Esta publicação             |
| <b>Organismo</b>            | Ausência de informação.                      | Selecionar organismo.           | Completeza.                                | Esta publicação             |
| <b>Outro organismo</b>      | Não se aplica                                | Não se aplica                   | Não se aplica                              | Não se aplica               |
| <b>Tipo de medida</b>       | Ausência de informação.                      | Selecionar tipo de medida.      | Completeza.                                | Esta publicação             |
| <b>Outro tipo de medida</b> | Não se aplica                                | Não se aplica                   | Não se aplica.                             | Não se aplica               |

Fonte: Adaptado de Souza et al. (2020).

