

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

DOCUMENTOS 177

**Termos Técnicos Úteis na Catalogação de *Datasets*
Ômicos no Repositório de Dados de
Pesquisa da Embrapa: uma
Compilação de Definições**

*Márcia Izabel Fugisawa Souza
Marcos Cezar Visoli
Carla Cristiane Osawa
Poliana Fernanda Giachetto
Paula Regina Kuser Falcão
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

Mariana Pilatti sob supervisão de Magda Cruciol

Foto da capa

Mariana Pilatti

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

Termos técnicos úteis na catalogação de *datasets* ômicos no Repositório de Dados de Pesquisa da Embrapa: uma compilação de definições / Márcia Izabel Fugisawa Souza ... [et al.]. – Campinas : Embrapa Agricultura Digital, 2021.
PDF (80 p.) : il. color. - (Documentos / Embrapa Agricultura Digital, ISSN 2764-2488; 177).

1. Representação descritiva. 2. Representação temática. 3. Catalogação. 4. *Datasets*. 5. Repositório de Dados de Pesquisa da Embrapa. 6. Termos técnicos. I. Souza, Márcia Izabel Fugisawa. II. Embrapa Agricultura Digital. III. Série.

CDD (21. ed.) 025.32

Carla Cristiane Osawa (CRB-8/10421)

© Embrapa, 2021

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, SP.

Poliana Fernanda Giachetto

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

Paula Regina Kuser Falcão

Física, doutora em Cristalografia de Proteínas, pesquisadora 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.

Luiz Manoel Silva Cunha

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

Leandro Cintra Carrijo

Cientista da Computação, doutor em Bioinformática, 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.

Apresentação

Este trabalho de compilação de termos é fruto de esforços conjuntos realizados no escopo da Ação Gerencial Local (AGL), intitulada “Implantar a gestão de dados de pesquisa na Embrapa Informática Agropecuária”, alinhada ao Programa de Governança de Dados de Pesquisa. Desses esforços participaram ativamente membros da equipe do Laboratório Multiusuário de Bioinformática da Embrapa (LMB) e do Grupo de Pesquisa em Computação Científica, Engenharia da Informação e Automação (GCIA), da Embrapa Agricultura Digital.

Trata-se de um documento de natureza terminológica indispensável à ampliação e à popularização do vocabulário técnico associado à Biblioteconomia aplicada a dados, em especial, na catalogação de *datasets* ômicos no Repositório de Dados de Pesquisa da Embrapa (Redape). Neste documento estão reunidos os principais termos catalográficos utilizados na linguagem corrente e empregados por catalogadores.

A publicação tem o propósito de apoiar os catalogadores – autores de *datasets* e bibliotecários, entre outros atores -, na compreensão e apropriação dos principais conceitos e definições, sobretudo no processo de representação descritiva. Além de contribuir para o mapeamento da linguagem corrente praticada pelos catalogadores e demais interessados na catalogação de *datasets* ômicos, a obra beneficiará diretamente a implementação de boas práticas de catalogação.

Ademais, esta compilação de termos é uma contribuição efetiva para a implementação do processo de gestão de dados biológicos na Embrapa, pois além do seu valor intrínseco, é também um documento complementar às orientações para catalogação de *datasets* no Redape, publicadas na série Documentos, pela Embrapa Agricultura Digital.

Destaca-se, ainda, a contribuição desta compilação de termos técnicos para o processo de gestão de dados de pesquisa na Embrapa, visto que oferece um recurso adicional no apoio a ações de capacitação e divulgação, previstas na Política de Governança de Dados, Informação e Conhecimento da Embrapa (PGDIC).

Sílvia Maria Fonseca Silveira Massruhá

Chefe-geral

Agradecimentos

Os compiladores manifestam reconhecimento e gratidão aos pesquisadores Antonio Nhani Junior e aos analistas Leandro Carrijo Cintra e Luiz Antonio Falaguasta Barbosa, pelas contribuições à elaboração desta compilação de termos para apoiar o processo catalogação de datasets ômicos no Repositório de Dados de Pesquisa da Embrapa (Redape). Essas contribuições foram importantes para que este documento pudesse expressar as especificidades da terminologia inerente à catalogação descritiva de datasets ômicos, e, dessa forma, contribuir para a melhoria do processo de gestão de dados de pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa (LMB).

Sumário

Introdução.....	8
Como utilizar a Compilação de Termos.....	10
Termos Compilados.....	11
Índice.....	62
Referências	69

Introdução

Este documento é uma compilação de termos catalográficos úteis para a atividade de catalogação de *datasets*¹ biológicos no Repositório de Dados de Pesquisa da Embrapa (Redape). Nele estão reunidos os principais vocábulos utilizados na catalogação de dados ômicos, especificamente.

Trata-se de uma obra de natureza compilatória, cuja finalidade principal é auxiliar os catalogadores (pesquisadores e autores de *datasets*) e os bibliotecários, na tarefa de catalogar *datasets* ômicos no Redape. Outra finalidade precípua desta compilação diz respeito à possibilidade de ampliação do vocabulário técnico usual, sobretudo pela reunião de termos relacionados ao universo da Biblioteconomia de Dados.

Cabe salientar que esta obra não tem a pretensão de normalizar, descrever e representar a terminologia especializada; ou seja, não se tratou e nem se pretendeu apresentar definições terminológicas, mas tão somente contribuir para ampliar e socializar novos entendimentos e conhecimentos voltados para catalogação de dados.

O público-alvo a que se destina este documento é formado por biólogos, bioinformatas, bibliotecários, cientistas da informação, cientistas da computação, cientistas de dados, gestores de dados e produtores de *datasets*, dentre outros.

Este documento consolida um trabalho de compilação de termos extraídos de glossários especializados de Biblioteconomia, Ciência da Informação, Ciência da Computação, Ciência de Dados, e de áreas correlatas, com vistas a ampliar o conhecimento dos profissionais envolvidos na catalogação de *datasets* ômicos.

A compilação está estruturada de modo a: 1) apresentar os termos selecionados e extraídos de fontes de informação especializados nas áreas temáticas mencionadas; 2) apresentar definições de termos não existentes nos glossários especializados consultados; 3) fornecer referências bibliográficas às citações empregadas na definição de cada termo compilado, de modo a possibilitar a ampliação da compreensão e o conceito deles; 4) proporcionar a exatidão conceitual, indicando a ocorrência de cada termo em diferentes contextos; 5) eliminar ambiguidades e inadequações lógicas da linguagem natural, facilitando a clareza e a interpretação inequívoca.

Dentre as fontes de informação utilizadas nesta compilação, destacam-se: glossários, tesouros, dicionários, enciclopédias, taxonomias, livros, artigos de periódicos, teses, dissertações, documentos publicados em sites especializados, repositórios e outros tipos de materiais, de origem nacional e internacional. Faz-se oportuno informar que definições e conceitos reunidos nesta compilação foram determinados puramente por razões de conveniência e adequação aos interesses precípuos da catalogação de *datasets*. O critério principal para a inclusão de um termo foi baseado no seu uso corrente pelos profissionais envolvidos com a catalogação de *datasets*.

¹ Optou-se pelo uso do termo '*dataset*' (no idioma inglês) por considerá-lo apropriado para se referir a coleções de 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 + file..., 2016, tradução nossa). 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).

Para termos cuja ocorrência não foi constatada nas fontes consultadas, optou-se pela sua criação e definição, tomando-se por base as prescrições mínimas extraídas da leitura das seguintes fontes: 1) NBR 13789: terminologia - princípios e métodos - elaboração e apresentação de normas de terminologia (Associação Brasileira de Normas Técnicas, 1997a); 2) NBR 13790: terminologia - princípios e métodos - harmonização de conceitos e termos (Associação Brasileira de Normas Técnicas, 1997b); 3) ABNT: diretiva 2 - regras para estrutura e redação de documentos técnicos ABNT (Associação Brasileira de Normas Técnicas, 2017); 4) ISO 704: terminology work - principles and methods (International Organization for Standardization, 2009). As obras consultadas e as obras citadas nesta compilação estão relacionadas no item Referências, ao final deste trabalho.

O trabalho de compilação foi iniciado durante a realização de uma Ação Gerencial Local (AGL), intitulada "Implantar a gestão de dados de pesquisa na Embrapa Agricultura Digital, em alinhamento ao Programa de Governança de Dados de Pesquisa", conduzida em 2020. Esta AGL teve o objetivo de desenhar e propor melhorias, sob a perspectiva da metodologia de governança de processos da Embrapa, para o processo de gestão de dados de pesquisa em curso no Laboratório Multiusuário de Bioinformática da Embrapa (LMB).

Para que se mantenha atualizada, pretende-se que esta compilação seja revisada periodicamente, com a inclusão de novos termos. Acredita-se que o conjunto de termos e definições constantes desta compilação possa servir de subsídios para os trabalhos da Comissão Permanente de Trabalho em Vocabulários Controlados, Agroterminologias e Agrossemântica da Embrapa (GTermos).

Como utilizar a Compilação de Termos

Os verbetes estão organizados em ordem alfabética palavra por palavra, e estruturados da seguinte forma:

Entrada + Definição ou $\Rightarrow \pm$ **Remissiva da Definição \pm Citação**.

A **Entrada** representa a unidade linguística que possui conteúdo semântico da expressão terminológica na linguagem de especialidade. É o termo propriamente dito, o termo principal ou, eventualmente, um termo remissivo. (Brasil, 2013a).

A **Definição** estabelece o sistema de distinções recíprocas que servem para descrever conceitos pertinentes aos termos. (Brasil, 2013a). É o conteúdo que descreve os conceitos pertinentes aos termos (Entrada).

A **Remissiva da Definição** ‘Ver’ esclarece sobre a relação de complementaridade entre termos, por exemplo: uma remissiva para um termo sinônimo. A remissiva ‘Ver’ é feita do termo não definido para o verbete com definição.” (Cunha; Cavalcanti, 2008, p. viii).

A **Citação** fornece a menção, na Definição, de informação extraída de uma fonte, indicando autor(es), ano de publicação e página(s).

Termos Compilados

AACR2

Acrônimo de *Anglo-American Cataloguing Rules*, 2nd ed. Publicação baseada nas regras de catalogação descritiva da *International Standard Bibliographic Description* – ISBD. Fixa normas relativas à descrição das publicações para facilitar o intercâmbio internacional da informação bibliográfica, pela forma escrita convencional, ou legível por máquina. “O principal objetivo da AACR2 é a padronização da catalogação em nível internacional e dar subsídios para o tratamento da informação”. (Martignago et al., 2019, p. 90).

ABNT NBR 6023

Norma que estabelece os elementos a serem incluídos em referências; fixa a ordem dos elementos das referências e estabelece convenções para transcrição e apresentação da informação originada do documento e/ou outras fontes de informação. Destina-se a orientar a preparação e compilação de referências de material utilizado para a produção de documentos e para inclusão em bibliografias, resumos, resenhas, resenhas, resenhas, resenhas e outros. (Associação Brasileira de Normas Técnicas, 2020, p. 1).

Abreviatura

“Representação de uma palavra por meio de algumas de suas sílabas ou letras.” (Brasil, 2013b, p. 15).

Abreviatura do nome do distribuidor

Elemento de descrição do esquema Metadados de Citação. Abreviatura do nome do distribuidor do *dataset*. (Souza et al., 2020).

Abreviatura do nome do produtor

Elemento de descrição do esquema Metadados de Citação. Abreviatura do nome do produtor do *dataset*. (Souza et al., 2020).

Acessibilidade

Acessibilidade (*Accessible*) é um dos princípios FAIR. Princípio que relaciona à capacidade de dados serem acessíveis e recuperáveis. (Go FAIR, 2018). Atributo da qualidade da informação. “A facilidade e a eficiência com que o usuário pode navegar na aplicação para acessar a informação desejada.” (Trindade et al., 2011, p. 782).

Acurácia

“Grau de certeza de um teste diagnóstico representado pelo valor real do efeito medido.” (Brasil, 2013a, p. 15).

Acurácia da informação

Atributo da qualidade da informação. “A percepção de que a informação está correta. A informação do sistema representa a situação do mundo real. Representação do que se supõe e do que se pretende.” (Trindade et al., 2011, p. 782).

“Proximidade entre o valor obtido experimentalmente e o valor verdadeiro na medição de uma grandeza, portanto, a acuracidade da informação é o seu nível de exatidão.” (Cunha; Cavalcanti, 2008, p. 5).

Acurácia/precisão

Precisão. Uma dimensão de medida de qualidade da informação. A acurácia e a precisão avaliam quão correta e factual está a informação fornecida. “Edição de alta qualidade: eliminação de erros tipográficos, conformidade de descrição de nomes pessoais, nomes corporativos e nomes de lugares, uso de abreviaturas padronizadas e assim por diante.” (Bruce; Hillmann, 2004, p. 243, tradução nossa).

Responde ao critério de qualidade: “Têm sido usados métodos aceites para criação ou extração? O que foi feito para garantir valores e estruturas válidos? Os valores-padrão são apropriados e foram usados de maneira adequada?” (Bruce; Hillmann, 2004, p. 251, tradução nossa).

Acurácia/veracidade

Uma dimensão de medida de qualidade da informação. “A acurácia e a veracidade avaliam quão próximo ou distante está a informação da realidade. Ex. A temperatura de todos os pacientes está em 39 graus, será que todos podem estar com febre ao mesmo tempo? Será que é verdade?” (Hoppen; Neto, 2020).

Afiliação do autor

Elemento de descrição do esquema Metadados de Citação. Afiliação do autor do *dataset*. Atributo de uma pessoa. (Souza et al., 2020).

“Entidade à qual um autor se encontra ligado profissionalmente.” (Cunha; Cavalcanti, 2008, p. 168).

“Uma instituição à [...] qual a pessoa é afiliada ou foi afiliada por meio de emprego, associação, identidade cultural etc.” (International Federation of Library Associations and Institutions, 2013, p. 17, tradução nossa).

Afiliação do distribuidor

Elemento de descrição do esquema Metadados de Citação. Afiliação do distribuidor do *dataset*. (Souza et al., 2020).

Entidade à qual o distribuidor se encontra ligado profissionalmente. (Cunha; Cavalcanti, 2008).

Agência

Elemento de descrição do esquema Metadados de Citação. Agência responsável pela geração do outro identificador digital do *dataset*. Exemplo: NCBI. (Souza et al., 2020).

AGROVOC Thesaurus

Vocabulário controlado desenvolvido e mantido pela *Food and Agriculture Organization of the United Nations* (FAO) com cobertura nas áreas de conhecimento da alimentação, nutrição, agricultura, florestas, pescas, nomes científicos e comuns de animais e plantas, meio ambiente, aspectos biológicos e técnicas de cultivo de plantas e de criação de animais. (FAO, 2020).

Amplitude

Atributo da qualidade da informação. “A informação pode ter um alcance amplo ou estreito, ou um foco interno ou externo.” (Trindade et al., 2011, p. 782).

Análise de assunto

“Etapa da análise documental; processo de identificação do conteúdo intelectual de uma obra. Operação base que permeia o processo de representação/recuperação da informação. Consiste na compreensão e interpretação do conteúdo informativo, com a finalidade de extração de conceitos.” (Bonotto et al., 2005, p. 4).

Anonimização dos dados

“Utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta a um indivíduo.” (Serpro, 2020).

“O processo de adaptação de dados para que os indivíduos não possam ser identificados a partir deles. O processo visa transformar os dados em uma forma que não identifica os indivíduos e onde a identificação provavelmente não ocorrerá.” (Anonymisation, 2020, tradução nossa).

“A anonimização compreende uma série de procedimentos a que devem ser submetidos os dados que contém informações pessoais; isso evita que eles revelem a identidade de indivíduos e impede também que esses dados estejam vinculados a outros e diferentes dados que possam revelar indiretamente a identidade dos indivíduos envolvidos.” (Sayão; Sales, 2015, p. 78).

Apresentação

Atributo da qualidade da informação. “A informação pode ser apresentada em forma narrativa, numérica, gráfica ou outras.” (Trindade et al., 2011, p. 782).

Assunto

Elemento de descrição do esquema Metadados de Citação. Assunto do *dataset*. (Souza et al., 2020).

“Tema ou conteúdo tratado por um ou mais autores numa obra, podendo estar explicitado ou não no título.” (Bonotto et al., 2005, p. 6).

“Qualquer conceito ou combinação de conceitos, considerados como um todo.” (Cunha; Cavalcanti, 2008, p. 35).

Atribuição

“O processo humano de atribuir valores a objetos com conteúdo. Ela é baseada na avaliação humana dos conteúdos e na percepção dos valores de atributos apropriados.” (Cunha; Cavalcanti, 2008, p. 194).

“Atribuição de termos de indexação - quer eles apareçam, ou não, no texto - para representar documentos ou dados, de acordo com certas regras.” (Cunha; Cavalcanti, 2008, p. 37).

Atributo

Aspectos de informação de um objeto com conteúdo. Exemplo: Na descrição dos elementos (campos e subcampos) do esquema Metadados de Citação são os seguintes os atributos: Nome (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 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. (Souza et al., 2020, p. 25, 51).

Atualidade

Uma dimensão de medida de qualidade da informação. Responde aos critérios de qualidade: “Os metadados são atualizados regularmente conforme os recursos mudam? Os vocabulários controlados são atualizados?” (Bruce Hillmann, 2004, p. 252, tradução nossa).

Atributo da qualidade da informação. “O quanto que a informação está atualizada. Passa a percepção de idade da informação, ou seja, o tempo decorrido desde a última atualização. Diz respeito a quão atualizada está a informação, e tem relação com confiabilidade.” (Trindade et al., 2011, p. 782).

Autenticação

“Ato de provar uma afirmação, como a identidade de um usuário do sistema de computador.” (FAIR Data Maturity Model Working Group, 2020, p. 39, tradução nossa).

Autodepositante

Autor do *dataset* no exercício do papel de catalogador no repositório de dados. (Souza et al., 2020).

Autodepósito

Ato que possibilita ao próprio autor do *dataset* fazer catalogação e a publicação de seus dados diretamente no repositório de dados sem necessidade de intermediários. (Souza et al., 2020).

Autor

Elemento de descrição do esquema Metadados de Citação. Autor do *dataset*. (Souza et al., 2020).

“[...] pessoa física (individual ou coletiva) ou a pessoa jurídica (Estado, governo, entidades coletivas e similares) que responsabiliza pelo conteúdo de uma obra.” (Cunha; Cavalcanti, 2008, p. 39).

“Autor: pessoa(s) física(s) responsável(eis) pela criação do conteúdo artístico ou intelectual de um documento.” (Associação Brasileira de Normas Técnicas, 2018, p. 2).

Autor corporativo

Elemento de descrição do esquema Metadados de Citação. Autor corporativo do *dataset*. Nomes de entidades coletivas, governamentais, públicas, privadas etc. (Souza et al., 2020).

Autor corporativo se aplica “[...] somente para obras que transmitam ideia do conjunto que representa uma organização como pessoa jurídica, assim como, documentos sobre sua política interna, procedimentos e/ou operações, suas finanças, seu pessoal, seus recursos, relatórios de comissões ou comitês, declarações oficiais sobre questões externas, planos etc.” (Embrapa, 2013, p. 26).

“Autor entidade: instituição(ões), organização(ões), empresa(s), comitê(s), comissão(ões), evento(s), entre outros, responsáveis por publicações em que não se distingue autoria pessoal.” (Associação Brasileira de Normas Técnicas, 2018, p. 2).

“Autor entidade: instituição ou grupo responsável por publicação na qual não se distingue a autoria pessoal.” (Brasil, 2013b, p. 17).

Autorização

“A função de especificar direitos/privilégios de acesso a recursos, que está relacionado à segurança da informação e segurança do computador em geral e ao controle de acesso em particular.” (FAIR Data Maturity Model Working Group, 2020, p. 39, tradução nossa).

Autor pessoal

Elemento de descrição do esquema Metadados de Citação. (Souza et al., 2020).

Autor pessoal do conjunto de dados “[...] pessoa física criadora de obra intelectual (literária, artística ou científica). O autor é definido por sua contribuição intelectual e/ou artística na composição de uma obra, os demais envolvidos são compreendidos como colaboradores.”(Autor, 2016).

Avaliação de qualidade

“Uma avaliação de qualidade é parte de uma garantia de qualidade.” (Quality assessment, 2019, p. 417, tradução nossa).

“Apreciação da qualidade e da validade da informação e dos dados científicos no decorrer dos trabalhos de síntese e de revisão. Prova e verificação da confiabilidade, ou seja, da medida da qualidade da informação e dos dados científicos.” (Cunha; Cavalcanti, 2008, p. 41)

“Uma avaliação da qualidade faz parte da garantia da qualidade.” (Quality assessment, 2019, p. 417, tradução nossa).

Banco de dados

“[...] representa algum aspecto do mundo real, às vezes chamado de **minimundo** ou de **universo de discurso**. As mudanças no minimundo são refletidas no banco de dados.” (Elmasri; Navathe, 2018, p. 4).

“Um banco de dados é uma coleção logicamente coerente de dados com algum significado inerente. Uma variedade aleatória de dados não pode ser corretamente chamada de banco de dados. [...] Um banco de dados é projetado, montado e preenchido com dados para uma finalidade específica. Ele tem um grupo intencionado de usuários e algumas aplicações previamente concebidas nas quais esses usuários estão interessados.” (Elmasri; Navathe, 2018, p. 5).

Best practices

Ver Melhores práticas.

Biblioteconomia de dados

Campo de conhecimento da biblioteconomia voltado “[...] para os estudos sobre a proveniência, as melhores práticas, o planejamento, a curadoria, entre outras atividades científicas relacionadas à teoria e à práxis com dados, principalmente quando esses dados se relacionam com os processos de pesquisa científica em meio digital.” (Semeler; Pinto, 2019, p. 127).

Big data

Para a Oracle (Dijcks, 2012 citado por Ward; Baker, 2013), *big data* normalmente se refere a três tipos de dados: dados empresariais tradicionais, dados gerados por máquinas/sensores e dados sociais. “Dados empresariais tradicionais - incluem informações de clientes de sistemas de CRM, dados transacionais do ERP, transações comerciais virtuais, dados contábeis. Dados gerados por máquina/sensores - incluem registros de detalhes de chamadas (CDR), weblogs, medidores inteligentes, sensores de manufatura, registros de equipamentos (frequentemente chamados de exaustão digital), dados de sistemas de negociação. Dados sociais - incluem fluxos de feedback de clientes, sites de microblog como o Twitter, plataformas de mídia social como o Facebook.” (Ward; Baker, 2013, tradução nossa).

Essas novas fontes incluem blogs, mídias sociais, redes de sensores, dados de imagens e outras formas de dados que variam em tamanho, estrutura, formato e outros fatores. [...] portanto, afirma, uma definição que é a inclusão de fontes de dados adicionais para aumentar as operações existentes.” (Ward; Barker, 2013, tradução nossa).

“O conceito se refere a estratégias e ferramentas para captura, armazenamento, análise e visualização de grandes volumes e formatos variados de dados, que dentre outros tipos abarcam a informação que todos nós produzimos diariamente ao circular pela internet; e seus usos têm a ver com a tradução e aplicação dessas informações em atividades práticas [...]” (Tautz, 2015).

Big data é um termo associado a dados multivariados e de elevada dimensão, geralmente criados em tempo real e que apresentam um crescimento exponencial (na escala temporal), nomeados de megadados. O conceito de *big data* foi definido inicialmente por 3'V mas a literatura mostrou que seu conceito pode ser expandido para 5'V, assim representados:

Volume: relacionado a grande quantidade de dados gerados, que apresenta um crescimento exponencial ao passar dos anos. (Coneglian et al., 2017, p.133).

Variedade: as fontes de dados são muito variadas, o que aumenta a complexidade das análises; refere-se a diferentes fontes informacionais que são utilizadas atualmente (e-mail, mídias sociais, blogs etc.) e em formatos variados (imagens, textos e vídeos), de modo que os processos analíticos não podem ficar presos a dados estruturados, gerados pelos sistemas de informação tradicionais, o que requer o uso da Web nas análises de dados. (Coneglian et al., 2017, p. 133).

Velocidade: devido ao grande volume e variedade de dados, todo o processamento deve ser ágil para gerar as informações necessárias; configura-se pela rapidez na geração, no processamento e na análise dos dados, de modo que, em algumas aplicações, a velocidade torna-se mais importante que o volume dos dados, devido ao dinamismo dos processos realizados pelas organizações, em que a tomada de decisão está vinculada diretamente com o momento e a velocidade em que é tomada. (McAfee; Brynjolfsson, 2012 citado por Coneglian et al., 2017, p. 133).

Outras duas características inseridas por outros autores são: o valor e a veracidade. “O valor é inserido como consequência das análises de *Big Data*, em que se torna possível adquirir informação com grande valor agregado, como, por exemplo, as tendências de mercado. Já a veracidade surge como requisito para se obter análises mais precisas, pois é necessário utilizar dados verdadeiros para se obter conclusões não equivocadas. “ (Coneglian et al., 2017, p. 133).

Características das fontes de dados observadas

Elemento de descrição do esquema Metadados de Citação. Características das fontes de dados do dataset. (Souza et al., 2020).

Catálogo

Processo técnico para registro de itens tendo em vista a organização de catálogos. “Em sentido mais amplo, a catalogação abrange não somente a descrição bibliográfica, mas também a análise temática.” (Cunha; Cavalcanti, 2008, p. 70).

Catálogo de dados

Processo técnico de representação descritiva e representação temática de *dataset*. (Souza et al., 2020).

Catalogador

Depositante do *dataset* no repositório de dados. (Souza et al., 2020).

“Pessoa responsável pelos registros catalográficos dos documentos que farão parte do acervo de bibliotecas, centros de documentação ou de informação, bem como, se for o caso, dos registros para bases e bancos de dados.” (Cunha; Cavalcanti, 2008, p. 71).

Categoria

“Classe fundamental (ou básica) que resulta da divisão do universo de conhecimentos, de acordo com as características intrínsecas, ou fundamentais, de cada conceito. Faceta geral que se aplica às várias áreas do conhecimento ou a vários assuntos.” (Cunha; Cavalcanti, 2008, p. 74-75).

“Conjunto de objetos com características comuns. Cada categoria contém termos baseados em uma única característica de divisão (teoria da facetação).” (Bonotto et al., 2005, p. 7).

Ciclo de vida dos dados

“[...] o conceito de modelo de ciclo de vida de dados foi formalmente definido, propondo diferentes modelos de ciclo de vida de dados como uma estrutura de alto nível, que abrange todos os aspectos de gerenciamento de dados, desde a criação até o consumo de dados. “Os principais objetivos de um modelo de ciclo de vida de dados são eliminar desperdícios, operar com eficiência e preparar produtos de dados prontos para os usuários finais, atendendo às restrições de qualidade esperadas. [...] modelos de ciclo de vida de dados são geralmente adaptados para campos e interesses específicos, transformando-se em objetivos específicos e diferentes estágios de dados, dependendo das necessidades do designer para cada estágio dos dados.” (Sinaeepourfard et al., 2015, p. 3, tradução nossa).

Ciclo de vida dos dados de pesquisa

“[...] termo amplamente utilizado na atualidade para nomear grandes conjuntos de dados complexos que os aplicativos de processamento de dados tradicionais ainda não conseguem lidar. Por sua enorme extensão em termos de volume, mas também de suas características intrínsecas, velocidade e variedade, *big data* exige tecnologias e métodos analíticos específicos que podem levar à extração de conhecimento e valor.” (Ciclo..., 2016).

“Um ciclo de vida de uma pesquisa rastreia a atividade de um pesquisador ao longo de um projeto; similarmente um ciclo de vida de dados de pesquisa rastreia o caminho dos dados ou das ações que são necessárias nos dados para conduzir a pesquisa para o próximo estágio.” (Rice; Southall, 2016, p. 77, tradução nossa).

Ciclo de vida dos metadados

“O ciclo de vida dos metadados é maior do que o ciclo de vida dos dados: metadados podem ser criados antes que os dados sejam criados ou capturados, por exemplo, para informar sobre os dados que estarão disponíveis no futuro; metadados devem ser mantidos após a remoção dos dados, por exemplo, para informar sobre dados que foram desativados ou retirados.” (Open Data Support, 2014, slide 10, tradução nossa).

Ciência Aberta

“Conceito amplo que abarca diversas práticas do fazer científico, desde o acesso aberto à literatura científica e aos dados abertos de pesquisa, passando pela ciência cidadã, recursos educacionais abertos, software e hardware abertos, entre outras. Faz referência a um modelo metodológico alinhado à filosofia da cultura digital, da colaboração e do compartilhamento das práticas e dos conteúdos científicos.” (Ciência..., 2016).

Citação

Elemento de descrição do esquema Metadados de Citação. Referência completa da publicação relacionada ao *dataset*. (Souza et al., 2020).

“A menção que se faz à opinião de uma autoridade sobre um certo assunto, como reforço de argumentação”. (Mello, 1968 citado por Cunha; Cavalcanti, 2008, p. 82).

“Menção, no texto, de informação colhida em outra fonte.” (Associação Brasileira de Normas Técnicas, 2002, p. 2).

“Menção de uma informação extraída de outra fonte.” (Associação Brasileira de Normas Técnicas, 2018, p. 2).

Clareza

Atributo da qualidade da informação. “A informação deve ser fornecida de forma que seja de fácil compreensão.” (Trindade et al., 2011, p. 782).

Classificação

Elemento de descrição do esquema Metadados de Citação. Classificação do assunto do *dataset*. (Souza et al., 2020).

“Processo de reunir documentos segundo semelhanças ou diferenças, utilizando um sistema pré-determinado. Análise e identificação do conteúdo de documentos, podendo lhes atribuir código.” (Norte, 2010, p. 13-14).

Classificação de *dataset*

“Refere-se à classificação do assunto do *dataset*, por meio da atribuição de uma ou mais categorias, sob as quais podem ser agrupados conjuntos de natureza semelhante.” (Souza et al., 2020, p. 33).

Coautor

“Pessoa física ou entidade coletiva que, associada a outra(s) pessoa(s) ou entidade(s) apresenta um trabalho, sem que se especifique a parte que cada um preparou. Se, pelo contrário, a parte que coube a cada um preparar encontra-se especificada, estes são, individualmente, autores e, portanto, colaboradores na elaboração de um trabalho comum.” (Cunha; Cavalcanti, 2008, p. 88).

Coautoria

“Criação em comum, por dois ou mais autores”. (Cunha; Cavalcanti, 2008, p. 88).

Coerência

Uma das dimensões de medida de qualidade da informação. Coerência na indexação refere-se ao “[...] uso, por diferentes indexadores, ou pelo mesmo indexador em épocas diversas, de descritor(es) idêntico(s) na indexação de um documento ou de documentos que tratam de um mesmo assunto; consistência na indexação, uniformidade na indexação.” (Cunha; Cavalcanti, 2008, p. 90).

Colaborador

Elemento de descrição do esquema Metadados de Citação. Colaborador para o desenvolvimento do *dataset*. (Souza et al., 2020).

Pessoa que participa da produção de um documento, obra ou conjunto de dados, juntamente com o autor. Pessoa ou instituição responsável pela coleta, gestão ou qualquer outra contribuição relevante para o desenvolvimento do conjunto de dados. Em sentido amplo, todo aquele que contribui para a elaboração de um trabalho. Pessoa ou entidade responsável por algum aspecto de uma obra ou de contribuição para ela, mas não pelo conteúdo da obra no seu conjunto ou totalidade. (Cunha; Cavalcanti, 2008, p. 91).

“Aquele que auxilia o autor na construção da obra intelectual, mas não participa da sua elaboração integral.” (Brasil, 2013b, p. 20).

Compilador

“Pessoa responsável por reunir textos ou partes de textos, sem alteração dos seus conteúdos.” (Brasil, 2013b, p. 21).

“É a pessoa que produz uma obra com material extraído de obras de outros autores.” (Compilador, 2003, p. 67).

Completeza

“O sistema provê toda a informação necessária para a execução da atividade. Quando todos os valores necessários estão incluídos.” (Trindade et al., 2011, p. 782).

Completude

Uma dimensão de medida de qualidade da informação. “A completude é definida como o percentual de registros ou campos preenchidos.” (Hoppen; Neto, 2020).

Completude da informação refere-se ao “[...] nível de quão completa é uma dada informação, isto é, se ela cobre todos os aspectos requeridos.” (Cunha; Cavalcanti, 2008, p. 95).

Concisão

Qualidade ou característica do que é conciso. “Relativo a conciso: algo reduzido ao essencial; em poucas palavras; preciso, sucinto, resumido.” (Concisão, 2001, p. 787; Conciso, 2001, p. 787).

Atributo da qualidade da informação. “A informação é apresentada de forma compacta. Apenas a informação que for necessária deve ser fornecida.” (Trindade et al., 2011, p. 782).

Confiabilidade

Atributo da qualidade da informação. “Probabilidade de prevenir erros ou falhas. Está relacionada com o quanto a informação está correta.” (Trindade et al., 2011, p. 782).

“Consistência com que uma dada medida, aplicada repetidamente ao mesmo objeto, produz sempre o mesmo resultado.” (Alvorcem; Dantas, 2020, p. 68).

Conhecimento

“Informação processada e transformada em experiência pelo indivíduo.” (Embrapa, 2019, p. 7).

Consistência

Medida da coerência (ou concordância) entre termos de indexação, atribuídos a uma unidade do-cumentária específica por diversos indexadores; concordância na indexação. (Cunha; Cavalcanti, 2008, p. 103).

A consistência de um conjunto de dados está associada com vários aspectos, como: frequência dos preenchimentos; integração dos dados entre os vários bancos de dados; mudanças nas fórmulas de cálculo ao longo da série; duplicidade ou falta de registros; presença de *outliers* (o que são *outliers* e como tratá-los?). (Hoppen; Neto, 2020).

Contato

Elemento de descrição do esquema Metadados de Citação. Contato sobre o *dataset*. Nome da(s) pessoa(s) ou da instituição para contato sobre o *dataset*. (Souza et al., 2020).

Controle de acesso

Um meio de limitar acesso a um sistema ou a recursos físicos ou virtuais. (Techopedia, 2020).

Controle de qualidade

“Um conjunto de procedimentos executados para garantir que um produto atenda a um padrão especificado.” (Quality control, 2019, p. 417, tradução nossa).

“Utilização de métodos de amostragem, inspeção e prova em todos os níveis de produção de um sistema para gerar produtos e serviços sem defeitos.” (Cunha; Cavalcanti, 2008, p. 107).

Procedimento que visa “[...] assegurar que os produtos, serviços ou processos preencham requisitos específicos e que sejam confiáveis e satisfatórios para os clientes. [...] envolve a análise de um produto, serviço ou processo para determinados níveis mínimos de qualidade.” (Torres, 2011 citado por Branco et al., 2014, p. 2438).

Creative Commons

Organização não governamental sem fins lucrativos, localizada em Mountain View, CA, voltada à expansão da quantidade de obras criativas disponíveis, através de suas licenças que permitem a cópia e o compartilhamento com menos restrições que o modelo tradicional de todos os direitos reservados (*copyright*). (Creative..., 2020).

Credibilidade da informação

“A credibilidade de uma informação é parte de um processo de comunicação, enreda uma relação de confiança, está interligada ao caráter imparcial da informação, do autor e da fonte.” (Fachin et al., 2020, p. 3).

Na Ciência da Informação, entende-se que a credibilidade da informação “[...] é uma propriedade objetiva da informação, associada à ideia de ‘qualidade’ dessa informação em termos de relevância, correção ou utilidade relativamente a um fim específico.” (Terra; Sá, 2012 citados por Fachin et al., 2020, p. 3).

CSV

Acrônimo de *Comma-separated values*. “Formato de arquivo de valores separados por vírgula, geralmente usado para trocar dados entre aplicativos semelhantes de forma diferente.” (CSV, 2020, tradução nossa).

Curador de dados

“Pesquisador ou cientista da informação com conhecimento disciplinar que adiciona valor aos dados por meio de documentação, integração, anotações, *mashup* etc. Promove o compartilhamento e reuso, avalia para a preservação e cria serviços.” (Sayão; Sales, 2019, slide 73).

Curadoria digital

“[...] ações voltadas para o gerenciamento de dados de pesquisa durante o seu ciclo de vida. Envolve "manter, preservar e adicionar valor aos dados.” (Curadoria..., 2016).

“Atividade de gestão e de promoção de uso de dados de pesquisa desde a criação/seleção, assegurando a sua adequação para propósitos atuais e futuros e tornando-os disponíveis para descoberta e reúso.” (Sayão; Sales, 2019, slide 56).

“A curadoria digital é a atividade de manter e agregar valor a um corpo confiável de informações digitais para uso atual e futuro.” (Janée, 2009, p. 816, tradução nossa).

Dado anonimizado

“Dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento.” (Serpro, 2020).

Dados

“[...] um conjunto de registros qualitativos ou quantitativos, conhecido, que organizado, agrupado, categorizado e padronizado adequadamente transforma-se em informação.” (Miranda, 1999 citado por Russo, 2010, p. 14). “Dados são elementos brutos, sem significado, desvinculados da realidade; correspondem a “observações sobre o estado do mundo”, são símbolos e imagens que não dissipam nossas incertezas.” (Davenport, 1998 citado por Russo, 2010, p. 14). “Os dados podem ser totalmente descritos por meio de representações formais, estruturais. Por serem quantificados ou quantificáveis, eles podem ser armazenados em um computador e processados por ele, o que é feito por meio de programas. [...] dados são sinais que não foram processados, correlacionados, integrados, avaliados ou interpretados de qualquer forma, e, por sua vez, representam a matéria-prima a ser utilizada na produção de informações.” (Russo, 2010, p. 15).

“Sequência de símbolos ou valores, produzidos como resultado de um processo natural ou artificial e representados em qualquer meio.” (Embrapa, 2019, p. 7).

No contexto digital, “[...] dados podem ser considerados como todo objeto criado em formato digital (*digital born*) ou convertido para formato digital (digitalizados) que possa ser usado para geração de insights de conhecimento.” (Semeler; Pinto, 2019, p. 118).

Dados abertos

“Dados disponibilizados à sociedade pela Embrapa serão completos, acessíveis, legíveis por máquina e mantidos no formato mais primário, sempre que possível, de modo a facilitar a reutilização, a criação de valor e novos modelos de negócio para a prestação de serviço.” (Embrapa, 2019, p. 13).

“Segundo a definição da *Open Knowledge Foundation*, os dados são considerados abertos quando: Qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade (sujeito a, no máximo, a requisitos que preservem a proveniência e a sua abertura). Isso

geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença aberta. [...]. Em 2009, [...] David Eaves propôs as seguintes 'leis' que regem os dados abertos governamentais e aceitas pela comunidade e pelo W3C: 1. Se o dado não pode ser encontrado e indexado na Web, ele não existe; 2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e 3. Se algum dispositivo legal não permitir sua replicação, ele não é útil." (Dados..., 2021).

Dados biológicos

"Os dados biológicos referem-se a um composto ou uma informação derivada de organismos vivos e seus produtos. Um composto medicinal feito de organismos vivos, como um soro ou uma vacina, pode ser caracterizado como dados biológicos. Os dados biológicos são altamente complexos quando comparados com outras formas de dados." (Biological..., 2021, tradução nossa).

Dados brutos

"São os que vêm diretamente dos instrumentos científicos ou coletados diretamente da fonte, sem sofrer qualquer manipulação ou processamento. São também chamados de dados crus ou dados primários." (Dados brutos, 2014).

"Dados brutos são um oxímoro e uma má ideia; pelo contrário, os dados devem ser preparados com cuidado" (Geoffrey Bowker citado por Barrowman, 2018, tradução nossa). "'Cru" carrega uma sensação de natural ou intocado, enquanto "cozido" sugere o resultado de processos cognitivos. Mas os dados são sempre o produto de processos cognitivos, culturais e institucionais que determinam o que coletar e como coletar. Nesse sentido, "dados brutos" são de fato uma contradição em termos. No uso comum do termo "dados brutos", "brutos" significa que nenhum processamento foi realizado após a coleta de dados, mas termo obscurece as várias formas de processamento que necessariamente ocorrem antes da coleta de dados." (Barrowman, 2018, tradução nossa).

Dados de pesquisa

"Registros factuais (pontuações numéricas, registros textuais, imagens e sons) produzidos ou utilizados como fontes primárias para a pesquisa científica e tecnológica e que são necessários para validação dos seus resultados. Variam de acordo com a área do conhecimento e podem estar contidos em documentos textuais, planilhas, estatísticas, cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, arquivos de áudio, vídeo, fotografias, imagens, sequências de proteínas ou genéticas, artefatos, amostras, modelos, algoritmos, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões e protocolos." (Embrapa, 2019, p. 8).

"A definição de dados de pesquisa transforma-se em relação às diferentes disciplinas de conhecimento científico." "[...] a expressão dados de pesquisa significa coisas diferentes e depende do contexto científico em que será utilizada." (Semeler; Pinto, 2019, p. 120).

Dados FAIR

Dados que estão de acordo com os princípios FAIR, e são localizáveis (*Findable*), acessíveis (*Accessible*), interoperáveis (*Interoperable*) e reutilizáveis (*Reusable*), de modo a facilitar a descoberta de conhecimento, auxiliando humanos e máquinas na descoberta e acesso, visando a integração e análise dos dados científicos, e seus algoritmos e fluxos de trabalho associados. (Bezjak et al., 2018, p. 185).

Data

Elemento de descrição do esquema Metadados de Citação. Data de descrição do *dataset*. Data em que o *dataset* foi descrito no repositório. (Souza et al., 2020).

“Indicação do momento, isto é, dia, mês e ano relativos a um documento, a um acontecimento, à publicação de um texto.” (Cunha; Cavalcanti, 2008, p. 113-114).

Dataset

“Uma coleção de conjuntos separados de informações que são tratados como uma única unidade por um computador.” (Dataset, 2021b, tradução nossa).

“*Datasets* ou conjunto de dados são comumente entendidos como uma coleção de dados agrupados de forma estruturada, tabular (linhas de colunas) ou não tabular, por exemplo, CSV ou XML.” (Lira, 2014, p. 16).

“Um *dataset* no Dataverse é um contêiner dos arquivos de dados, da documentação, dos códigos e dos metadados descritivos desse *dataset*.” (Dataverse + file..., 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).

No software Dataverse, “Um dataset é composto por metadados, pelos termos de uso (como licenças) e por arquivos.” (Rocha et al., 2021, p. 9).

“É o conjunto de dados disponibilizados para uma pesquisa, observação, após a extração e tratamento dessas informações.” (Dataset, 2021a)

Datasets relacionados

Elemento de descrição do esquema Metadados de Citação. *Datasets* relacionados ao *dataset* que está sendo descrito. (Souza et al., 2020).

Data citation

“Citação de dados é a prática de referenciar dados de maneira semelhante ao fornecer referências bibliográficas, reconhecendo os dados como uma saída de primeira classe em qualquer processo investigativo. A citação de dados oferece benefícios no suporte à atribuição e crédito adequados

para aqueles que produzem os dados, facilitando a descoberta de dados, apoiando o rastreamento do impacto e reutilização de dados, permitindo a colaboração e reutilização de dados, permitindo a reprodutibilidade dos resultados com base em dados.” (Albertoni et al., 2020, tradução nossa).

“Citação de dados facilita a replicação dos resultados da pesquisa original, enquanto fornece o crédito merecido aos coletores e autores de dados. Citação de dados gerada por um *dataverse* inclui: autor(es), data de distribuição, título, identificador persistente global e URL, universal numerical fingerprint (UNF) e campos opcionais como distribuidor, subconjuntos de dados e versões.” (Crosas, 2011, tradução nossa).

Data de distribuição

Elemento de descrição do esquema Metadados de Citação. Data de distribuição do *dataset*. Data em que o *dataset* foi disponibilizado para distribuição. (Souza et al., 2020).

Data de início

Elemento de descrição do esquema Metadados de Citação. Data de início da cobertura de dados do *dataset*. Data de início de coleta de dados do *dataset*. (Souza et al., 2020).

Data de produção

Elemento de descrição do esquema Metadados de Citação. Data de produção do *dataset*. Data em que o *dataset* foi produzido. (Souza et al., 2020).

Data de término

Elemento de descrição do esquema Metadados de Citação. Data de término da cobertura de dados do *dataset*. Data de término de coleta de dados do *dataset*. (Souza et al., 2020).

Data do depósito

Elemento de descrição do esquema Metadados de Citação. Data de depósito do *dataset*. Data em que o *dataset* foi depositado no repositório. (Souza et al., 2020).

Data steward

“Data Steward, ou Gestor de Dados de Negócio, executa a gestão de dados na área de negócio do empreendimento, visando os interesses dos consumidores e não apenas os interesses da TI. Se a empresa for de larga escala, são necessários gestores específicos para cada um dos segmentos de negócio.” (Impacta, 2021).

“Os gestores de dados são líderes de negócios ou especialistas designados no assunto, que possuem conhecimento detalhado sobre os requisitos de dados e de negócios e que podem traduzir esses requisitos em especificações técnicas.” (Abraham et al., 2019, p. 428, tradução nossa).

Há dois tipos de *data steward* - gestores de dados de negócios atuando como fiduciários de produtos de negócios e gestores de dados técnicos atuando como guardiões especializados e curadores desses ativos (Data Management International, 2010 citado por Peng et al., 2016). Os gestores de dados têm a responsabilidade de garantir e melhorar a procedência e rastreabilidade dos dados e definir e fornecer requisitos de arquivamento de dados aos produtores de dados. Eles são responsáveis por coletar, capturar e transmitir informações de qualidade de dados. (Peng et al., 2016).

Data stewardship

“[...] conjunto de habilidades para garantir que os dados sejam adequadamente gerenciados, compartilhados e preservados, tanto ao longo do ciclo de vida da pesquisa quanto para a preservação a longo prazo.” (European Commission Expert Group on FAIR Data, 2018, p. 13, tradução nossa).

Data Stewardship Wizard

“[...] ferramenta para planejamento de gestão de dados focada em fornecer o maior valor no planejamento de gestão de dados para o projeto em si, ao invés de cumprir obrigações. É baseada na governança de dados FAIR, em que cada decisão relacionada a dados do projeto atua para otimizar a capacidade de encontrar, acessar, interoperar e/ou reutilizar os dados.” (Pergl et al., 2019, p. 1, tradução nossa).

Dataverse

Software Dataverse é uma aplicação web, dedicada para o compartilhamento, preservação, citação, exploração e análise de dados de pesquisa. (Dataverse, 2021).

“Entidade *dataverse* é a estrutura que o software [Dataverse] disponibiliza para representar organizações, grupos ou unidades. [...] Cada entidade *dataverse* contém *datasets*, que são as entidades que representam conjuntos de dados. [...] Uma entidade *dataverse* pode ser definida como se fosse um repositório independente, com todas as autorizações para gerenciamento e operação (como definir papéis, permissões e grupos; criar e gerenciar entidades *dataverses*, *datasets*, arquivos etc.). Cada entidade *dataverse* pode ter marca personalizada (interface, logotipo) e interface configurada para descoberta de dados”. (Rocha et al., 2021, p. 8).

“Cada *dataverse* contém conjuntos de *datasets*, e cada *dataset* contém metadados e arquivos de dados descritivos (incluindo documentação e código que acompanham os dados. Como método de organização, *dataverses* também podem conter outros *dataverses*.” (Fundação Oswaldo Cruz, 2020, p. 6).

Default

“Opção automática utilizada pelo sistema de computação.” (Norte, 2010, p. 17).

Definição

“[...] estabelece o sistema de distinções recíprocas que servem para descrever conceitos pertinentes aos termos.” (Brasil, 2013b, p. 11).

Depositante

Elemento de descrição do esquema Metadados de Citação. Nome do depositante do *dataset* no repositório; preenchido automaticamente com o nome do autor do *dataset*. (Souza et al., 2020).

Depósito de dados

“[...] uma coleção de dados orientadas a assunto, integrada, não-volátil e temporal, de suporte a decisões gerenciais.” (Inmon, 1996 citado por Valente, 2001, p. 16). “Tradução do termo inglês *data warehouse*, [...] consiste em um único repositório de dados, extraídos de bases de dados transacionais e/ou dados externos, acumulados ao longo do tempo, integrados, possibilitando a análise massiva de informações, de forma a permitir melhores tomadas de decisões e a descoberta de conhecimento, sem impactar no desempenho dos bancos de dados do mundo transacional.” (Barquim; Edelstein, 1997 citados por Valente, 2001, p. 16).

Descoberta

“Ato de encontrar novos dados.” (Strasser et al., 2011 p 10, tradução nossa).

“Estratégias e processos usados pelo usuário final para localizar e acessar produtos (metadados, dados e outras informações relacionadas) do ciclo de vida dos dados.” (Discovery, 2021, tradução nossa).

Descrição

Elemento de descrição do esquema Metadados de Citação. Descrição sucinta do propósito, natureza e escopo do *dataset*. (Souza et al., 2020).

“Catalogação descritiva, descrição documentária.” (Cunha; Cavalcanti, 2008, p. 119).

Descrição de metadados

Consiste na descrição de um recurso de informação, por exemplo de um *dataset*, por meio de um conjunto de metadados. A catalogação é a atividade que compreende a tarefa de descrever os metadados de um *dataset*, incluindo-se a representação de assunto. (Souza et al., 2020).

Desempenho

Atributo da qualidade da informação. “A informação pode revelar desempenho pela mensuração das atividades concluídas, progresso realizado ou recursos acumulados.” (Trindade et al., 2011, p. 782).

Dicionário de dados

“Um dicionário de dados cataloga as definições de elementos de dados, tipos de dados, fluxos de dados e outras convenções que são usadas em um sistema de informação. Dicionários de dados têm sido amplamente adotados pela: i) comunidade de banco de dados, onde um dicionário geralmente descreve entidades de banco de dados, esquemas, permissões etc.; e ii) comunidade de desenvolvimento de software, onde um dicionário normalmente descreve os fluxos de informações através do sistema. Em essência, um dicionário de dados é um banco de dados virtual de metadados sobre o próprio sistema de informações. Um dicionário de dados também pode ser referido como um ‘catálogo do sistema.’” (Caverlee, 2009, p. 570, tradução nossa).

Dimensões da qualidade da informação

Categorização de medidas de qualidade usadas para medir a qualidade da informação. Relevância, acurácia, precisão, atualidade, acessibilidade, interpretabilidade e coerência são dimensões da qualidade da informação consideradas pelo *Quality Assurance Framework* (QAF), para dados estatísticos, desenvolvidos pelo *Statistics Canada* (STC). (Bruce; Hillmann, 2004, p 242).

Direito autoral

“Proteção legal que o autor ou responsável (pessoa física ou jurídica) tem sobre a obra.” (Brasil, 2013b, p. 22).

Direito moral do autor

“Direito intransferível sobre a natureza intelectual da obra.” (Brasil, 2013b, p. 22).

Direito patrimonial do autor

“Refere-se às prerrogativas de utilização da obra intelectual e consiste no direito de autor autorizar ou não qualquer forma de exploração de sua obra.” (Brasil, 2013b, p. 23).

Disponibilidade

“Qualidade da informação que pode ser conhecida e utilizada por indivíduos, equipamentos ou sistemas autorizados.” (Embrapa, 2019, p. 8).

Atributo da qualidade da informação. “Refere-se à flexibilidade e compreensão das características da aplicação disponibilizadas para os usuários para especificar e controlar a relação temporal entre os vários componentes hiperímida para que seja disponibilizada uma informação de hiperímida integrada.” (Trindade et al., 2011, p. 782).

Distribuidor

Elemento de descrição do esquema Metadados de Citação. Distribuidor do *dataset*. Pessoa ou instituição designada pelo autor ou produtor do *dataset* para gerar cópias do trabalho, incluindo quaisquer edições ou revisões necessárias. (Souza et al., 2020).

Documentação e acesso a fontes de dados

Elemento de descrição do esquema Metadados de Citação. Documentação e acesso a fontes de dados do *dataset*. (Souza et al., 2020).

DOI

Acrônimo de *Digital Object Identifier*. “[...] um identificador persistente, que é usualmente assinalado a itens digitais como um artigo de periódico ou uma coleção de dados, com o objetivo de identificá-los univocamente e dessa forma serem descobertos e citados.” (DOI, 2016).

DSpace

“[...] plataforma de software de código aberto que permite às organizações: capturar e descrever material digital usando um módulo de fluxo de trabalho de envio ou uma variedade de opções de ingestão programática; distribuir os ativos digitais de uma organização pela web por meio de um sistema de busca e recuperação; preservar os ativos digitais a longo prazo.” (Donohue, 2018, tradução nossa).

Dublin Core

“*Dublin Core* é um conjunto de metadados padronizado para descrever recursos de informação como documentos, vídeos, imagens, serviços e outros artefatos digitais. *Dublin Core* se destina a fornecer um modelo de metadados simples que pode ser adotado em uma ampla gama de comunidades em um esforço para aprimorar a interoperabilidade semântica.” (Caverlee et al., 2009, p. 947, tradução nossa).

Elemento metadado

“Uma propriedade que é usada para descrever uma característica, por exemplo, *Dublin Core* dc: title, Schema.org schema: name ou DataCite Title.” (FAIR Data Maturity Model Working Group, 2020, p. 41, tradução nossa).

E-mail do contato

Elemento de descrição do esquema Metadados de Citação. Endereço de e-mail do contato sobre o *dataset*. (Souza et al., 2020).

Encontrabilidade

Encontrabilidade (*Findable*) é um dos princípios FAIR. Princípio que se relaciona à capacidade de dados e metadados serem encontrados para que possam ser acessados, recuperados e reusados, seja por ação humana ou por sistemas automatizados. (Go FAIR, 2018).

Entrada

“É o termo propriamente dito, o termo principal ou, eventualmente, um termo remissivo.” (Alvorcem; Dantas, 2020, p. 9)

e-Science

“Conceito relacionado às novas práticas de compartilhamento da produção do conhecimento científico e uso de plataformas tecnológicas para a pesquisa colaborativa. Representa uma ciência atuante no uso de dados por meio de computação avançada.” (e-Science, 2016).

Financiamento

Elemento de descrição do esquema Metadados de Citação. Nome da instituição, agência de fomento ou organização que concedeu suporte e/ou apoio financeiro para a produção do *dataset*. (Souza et al., 2020).

Fontes de dados

Elemento de descrição do esquema Metadados de Citação. Fontes de dados para o *dataset*. Materiais que serviram de fontes para o *dataset*, como: lista de livros, artigos, publicações em série ou arquivos de dados legíveis por máquina. (Souza et al., 2020).

Formato de metadados

“Estrutura padronizada e conteúdo consistente para metadados, geralmente em linguagem de marcação extensível legível por máquina (XML) que pode ser representada em outros formatos legíveis por humanos (por exemplo, HTML, pdf etc.). Dentre os padrões para metadados incluem-se EML, FGDC e ISO 19115.” (Strasser et al., 2011, p. 10, tradução nossa).

FRAD

Acrônimo de *Functional Requirements for Authority Data* (em português, Requisitos Funcionais para Dados de Autoridade). Modelo conceitual contemplado pelo RDA, que apresenta os requisitos funcionais para o controle de autoridades e o intercâmbio internacional de dados de autoridade. (Segundo et al., 2013, p. 3).

FRBR

Acrônimo de *Functional Requirements for Bibliographic Records* (em português, Requisitos Funcionais para Registros Bibliográficos). Modelo conceitual de entidade-relacionamento desenvolvido pela *International Federation of Library Associations and Institutions* (IFLA), que relaciona as tarefas do usuário de recuperação e acesso em catálogos on-line de bibliotecas e bases de dados bibliográficos, na perspectiva do usuário. (Functional..., 2021).

Garantia de qualidade

“Processo padronizado projetado e executado para evitar equívocos e erros em produto lançado.” (Quality assurance, 2019, p. 417, tradução nossa).

Garantia de qualidade visa assegurar que determinado produto, serviço ou processo atinja níveis mínimos de qualidade. Garantia da qualidade está empenhada em fazer com que produto, processo, atividade ou serviço atinja de forma eficiente os objetivos para eles traçados. (Branco et al., 2014, p. 2438-2439).

Garantia de qualidade de dados

“Um conjunto de atividades para garantir que os dados sejam gerados e compilados de maneira a atender ao objetivo do projeto.” (Strasser et al., 2011, p. 11, tradução nossa).

Gestão

“Organização, planejamento, implementação e monitoramento de ações para o alcance dos objetivos organizacionais e o uso adequado e seguro dos recursos disponíveis.” (Embrapa, 2019, p. 8).

Gestão de dados

“Processo que contempla as atividades de planejamento, aquisição, organização, estruturação, curadoria e análise de dados, utilizando para isso ferramenta computacional apropriada para o armazenamento de dados, levando em consideração as questões relativas à preservação, à organização, ao compartilhamento, à proteção e à confidencialidade dos mesmos, bem como o acesso e disponibilização para a sociedade quando cabível.” (Embrapa, 2019, p. 8).

Gestão de dados de pesquisa

“Conjunto de práticas de gestão voltadas para o tratamento de dados de pesquisa durante o seu ciclo de vida; inclui todos os aspectos de manutenção, compartilhamento, segurança e preservação.” (Gestão..., 2016).

Gestão do conhecimento

“Processo sistemático, articulado e intencional, apoiado na geração, codificação, disseminação e apropriação de conhecimentos, com o propósito de atingir a excelência organizacional.” (Embrapa, 2019, p. 8).

Glossário

“Lista que define ou explica termos específicos de um texto, uma obra ou uma área do conhecimento.” (Dias; Silva, 2016, p. 56).

“Lista de palavras pouco conhecidas ou estrangeiras ou termos e expressões técnicas acompanha-das de definições ou traduções. Lista de palavras e termos usados em um campo de conhecimento, apresentando definições resumidas de uso especial.” (Glossário, 2003, p. 111).

“Obra que explica o significado de vocábulos pouco comuns, bem como termos técnicos e científicos.” “[...] Lista alfabética de termos específicos de uma área do conhecimento, com definições e explicações.” (Cunha; Cavalcanti, 2008, p. 180).

Governança

“Organização e implementação de políticas, procedimentos, estruturas, culturas, papéis e responsabilidades que delinham, viabilizam e transformam os processos de gestão para atender às necessidades atuais e futuras dos públicos interno e externo da organização.” (Embrapa, 2019, p. 8-9).

Governança de dados

“Envolve políticas, processos e pessoas, além de uma estrutura organizacional para apoiar o gerenciamento de dados institucionais.” (Governança..., 2016).

“[...] uma ação multidisciplinar que tem como finalidade tratar os dados como insumos ativos e tangíveis nas organizações. Para isso, a governança de dados determina políticas, padronizações, processos, papéis, responsabilidades e tecnologias para melhor acompanhar e monitorar os dados gerados, armazenados, utilizados e eliminados na organização.” (Espíndola et al., 2018, p. 276).

“[...] uma estrutura que orienta e estabelece estratégias, políticas e objetivos com a finalidade de gerenciar os dados, como se fossem qualquer outro recurso de uma organização. Assim, para a governança de dados, os dados são importantes recursos para as organizações, os quais devem ser gerenciados, monitorados e acompanhados como qualquer outro insumo que auxilia no cumprimento da missão organizacional e no atingimento de seus objetivos.” (Loftis, 2014 citado por Espíndola et al., 2018, p. 280).

Guestbook

“Na internet, um livro de visitas é um sistema que permite aos visitantes de um site deixar um co-mentário público.” (Livro..., 2019).

“[...] livros de visitas permitem que você colete dados sobre quem está baixando os arquivos de

seus conjuntos de dados. Você pode decidir coletar informações da conta (nome de usuário, nome e sobrenome, afiliação, etc.), bem como criar perguntas personalizadas (por exemplo: Para que você planeja usar esses dados? Você também pode baixar os dados coletados dos livros de visitas habilitados como arquivos do Excel para armazenar e usar fora do *Dataverse*.” (Dataverse guestbooks, 2016, tradução nossa).

Identificador

Atributo orientador para a descrição dos elementos (campos e subcampos) do esquema Metadados de Citação. (Souza et al., 2020).

Identificador digital persistente

Elemento de descrição Metadados de Citação. Identificador digital persistente que identifica o conjunto de dados. Um identificador persistente é uma referência duradoura para um recurso digital. Normalmente, ele possui dois componentes: um identificador exclusivo; e um serviço que localiza o recurso ao longo do tempo, mesmo quando a localização é alterada. O primeiro componente ajuda a garantir a proveniência de um recurso digital (o que ele pretende ser); enquanto o segundo irá garantir que o identificador resolva a localização atual correta. Os identificadores persistentes visam, assim, resolver o problema da persistência do acesso ao recurso citado, particularmente na literatura acadêmica. (Ibict, 2016).

Identificador global único e persistente

“Identificadores globalmente únicos e persistentes removem ambiguidade no significado de seus dados publicados, atribuindo um identificador único a cada elemento de metadados e a cada conceito / medição em seu conjunto de dados.” (Go FAIR, 2018, tradução nossa).

GUI e DOI são dois exemplos de identificadores globais únicos. “GUID é um acrônimo para *Globally Unique Identifier*. [...] às vezes chamado de UUID ou *Universally Unique Identifier*. Um GUID / UUID é um ID globalmente exclusivo (*string*) que identifica uma entidade digital (pessoa, lugar ou coisa) que precisa ser gerenciada e mantida disponível por um tempo de duração. GUIDs estão sendo atribuídos a dados de pesquisa, publicações de pesquisa, pessoas, organizações, bolsas, laboratórios e um número crescente de lugares e coisas. Os GUIDs são exclusivos, registrados com metadados descritivos e administrativos, resolvem em uma página de destino de metadados e seriam potencialmente persistentes (PIDs), oferecendo um redirecionamento contínuo ou uma pedra de exclusão para quaisquer alterações na resolução. Os GUIDs ‘resolvem’ para uma página de destino que deve incluir uma citação completa ou o GUID em um formulário legível e acionável (um link ativo). As citações de recursos devem incluir o GUID como parte do texto da citação.” (Globally..., 2021, tradução nossa).

Idioma

Elemento de descrição do esquema Metadados de Citação. Idioma do *dataset*. (Souza et al., 2020).

Refere-se à língua do texto original. (Cunha; Cavalcanti, 2008, p. 190).

Indexação

“Processo pelo qual se relacionam, de forma sistemática, descritores ou palavras-chave, que permitem a recuperação posterior do conteúdo (e/ou imagem) dos documentos, utilizando vocabulário controlado ou *thesaurus*.” (Alvorcem; Dantas, 2020, p. 134).

Indicadores FAIR

“São derivados dos princípios FAIR e visam formular aspectos mensuráveis de cada princípio e podem ser usados por abordagens de avaliação. Foram desenvolvidos em um processo de decomposição do texto dos princípios e explicações do FAIR, fornecidos pelo artigo original, um artigo posterior e pelo GO-FAIR.” (FAIR Data Maturity Model Working Group, 2020, p. 6, tradução nossa)

Indicadores de maturidade FAIR

Indicadores desenvolvidos para uso no modelo de maturidade FAIR, visando formular aspectos mensuráveis na avaliação de cada princípio FAIR. De acordo com o FAIR Data Maturity Model Working Group (2020, p. 11-12), são os seguintes os indicadores do modelo de maturidade de dados FAIR:

- F1 Metadados são identificados por um identificador persistente
- F1 Dados são identificados por um identificador persistente
- F1 Metadados são identificados por um identificador único global
- F1 Dados são identificados por um identificador único global
- F2 Metadados ricos são fornecidos para permitir a descoberta
- F3 Metadados incluem o identificador para os dados
- F4 Metadados são oferecidos de forma que possam ser coletados e indexados
- A1 Metadados contêm informações que permitem ao usuário obter acesso aos dados
- A1 Metadados podem ser acessados manualmente (ou seja, com intervenção humana)
- A1 Dados podem ser acessados manualmente (ou seja, com intervenção humana)
- A1 Identificador de metadados é resolvido para um registro de metadados
- A1 Identificador de dados é resolvido para um objeto digital
- A1 Metadados são acessados por meio de protocolo padronizado
- A1 Dados são acessíveis por meio de protocolo padronizado

A1 Dados podem ser acessados automaticamente (ou seja, por um programa de computador)

A1.1 Metadados são acessíveis por meio de um protocolo de acesso gratuito

A1.1 Dados são acessíveis por meio de um protocolo de acesso gratuito

A1.2 Dados são acessíveis por meio de um protocolo de acesso que suporta autenticação e autorização

A2 Metadados são garantidos para permanecer disponíveis depois que os dados não estiverem mais disponíveis

I1 Metadados usam representação de conhecimento expressa em formato padronizado

I1 Dados usam representação de conhecimento expressa em formato padronizado

I1 Metadados usam representação de conhecimento compreensível por máquina

I1 Dados usam representação de conhecimento compreensível por máquina

I2 Metadados usam vocabulários compatíveis com FAIR

I2 Dados usam vocabulários compatíveis com FAIR

I3 Metadados incluem referências para outros metadados

I3 Dados incluem referências para outros dados

I3 Metadados incluem referências para outros dados

I3 Dados incluem referências qualificadas para outros dados

I3 Metadados incluem referências qualificadas para outros metadados

I3 Metadados incluem referências qualificadas para outros dados

R1 Pluralidade de atributos precisos e relevantes é fornecida para permitir o reúso

R1.1 Metadados incluem informações sobre a licença sob a qual os dados podem ser reusados

R1.1 Metadados referem-se a uma licença de uso padrão

R1.1 Metadados referem-se a uma licença de uso compreensível por máquina

R1.2 Metadados incluem informação de proveniência de acordo com padrões específicos de comunidade

R1.2 Metadados incluem informação de proveniência de acordo com uma linguagem entre comunidade

R1.3 Metadados estão em conformidade com um padrão de comunidade

R1.3 Dados estão em conformidade com um padrão de comunidade

R1.3 Metadados são expressos em conformidade com um padrão de comunidade compreensível por máquina.

Índice

“Relação de palavras ou frases, ordenadas segundo determinado critério, que localiza os assuntos e remete o leitor para as informações contidas num texto.” (Brasil, 2013b, p. 29).

Informação

“Dados, processados ou não, contidos em qualquer meio, suporte ou formato, que podem ser utilizados para produção e transmissão de conhecimento.” (Embrapa, 2019, p. 9).

Informação sobre a série

Elemento de descrição do esquema Metadados de Citação. Informação sobre a série do *dataset*. (Souza et al., 2020).

Inteligibilidade

Atributo da qualidade da informação. “A informação é facilmente entendida.” (Trindade et al., 2011, p. 782).

Interoperabilidade

“Interoperabilidade é a capacidade de um sistema ou um objeto digital trocar informações entre si. Para tal utilizam-se diversos mecanismos, tais como *WebService* (canal de comunicação direta entre dois sistemas ou objetos) e troca de arquivos (os dados são depositados em um arquivo e esse arquivo é transferido). Para que a comunicação seja estabelecida é necessário adotar ou construir um padrão de comunicação entre esses sistemas ou objetos digitais, de tal forma que seja possível a interpretação dos dados que estão sendo transferidos.” (Furniel et al., 2017?, p. 14).

Habilidade de sistemas trocarem informação e usarem a informação trocada. (Interoperability, 2020).

Interoperabilidade (*Interoperable*) é um dos princípios FAIR. Princípio que se relaciona à capacidade de dados serem interoperáveis com outras aplicações e sistemas, para que possam ser analisados, armazenados, processados, descobertos e reusados. (Go FAIR, 2018).

Interpretabilidade

Atributo da qualidade da informação. “A informação está disponível em linguagem e símbolos apropriados, e as definições são claras.” (Trindade et al., 2011, p. 782).

Licenças *Creative Commons*

Licenças *Creative Commons* são várias licenças públicas que permitem a distribuição gratuita de uma obra protegida por direitos autorais. Uma licença *Creative Commons* é usada quando um autor quer dar às pessoas o direito de compartilhar, usar e construir sobre um trabalho que ele criou. Existem seis tipos de licenças *Creative Commons*, as quais diferem entre si por várias combinações que condicionam os termos de distribuição. São elas (Licenças..., 2020; Creative Commons, 2021):

Atribuição (CC-BY);

Atribuição-Compartilhual (CC BY-SA);

Atribuição-SemDerivações (CC BY-ND);

Atribuição-NãoComercial (CC BY-NC);

Atribuição-NãoComercial-Compartilhual (CC-BY-NC-SA);

Atribuição-NãoComercial-SemDerivações (CC BY-NC-ND).

Licenciamento

“Dados e informações disponibilizados em repositórios institucionais [...] acompanhados das apropriadas licenças de uso, com prioridade para a adoção de modelos de licenças públicas reconhecidos por máquinas.” (Embrapa, 2019, p. 12).

Linguagem artificial

Linguagem de indexação onde os sinônimos, homógrafos, plural/singular são controlados, facilitando o desenvolvimento de buscas exaustivas através do agrupamento de conceitos relacionados. Os termos extraídos a partir de uma lista padronizada garantem a consistência na indexação. O vocabulário controlado é o componente básico da linguagem artificial. (Linguagem artificial, 2003, p. 146-147).

Linguagem documentária

“[...] um instrumento de representação dos assuntos dos documentos que se pode apresentar sob uma notação ou sob forma verbal. (Gomes, 2001, p. 11).

Linguagem livre

Linguagem de indexação que emprega quaisquer termos. (Linguagem livre, 2003, p. 147).

Linguagem natural

Linguagem de indexação onde os termos são extraídos do documento que está sendo indexado. Não existe o risco de ocorrer distorções das palavras do autor, no entanto, não há controle de sinônimos e ambiguidades, exigindo grande esforço no estágio de busca. São as linguagens faladas, apresentam características que dificultam sua utilização para o tratamento da informação. (Linguagem natural, 2003, p. 147).

Livro de visita

Ver *Guestbook*.

Local de produção

Elemento de descrição do esquema Metadados de Citação. Local de produção do *dataset*. (Souza et al., 2020).

Logotipo

“Representação gráfica de empresa, instituição ou produto, que utiliza apenas caracteres para a sua identificação.” (Brasil, 2013b, p. 31).

Mashup

“A combinação de vários conjuntos de dados de várias fontes para criar um novo serviço, visualização ou informação.” (Mashup, 2020, tradução nossa).

Material relacionado

Elemento de descrição do esquema Metadados de Citação. Material relacionado ao *dataset* que está sendo descrito. (Souza et al., 2020).

Melhores práticas

“Métodos ou abordagens geralmente reconhecidos por uma comunidade como corretos ou mais apropriados para adquirir, gerenciar, analisar e compartilhar dados.” (Strasser et al., 2012, p. 9, tradução nossa).

Metadados

“Conjunto de dados estruturados e codificados, a partir de protocolos comuns, que descrevem e permitem acessar, gerenciar, distribuir, compreender, recuperar e/ou preservar outros dados.” (Embrapa, 2019, p. 9).

Metadados fornecem uma descrição básica dos dados, incluindo informações sobre autoria, título, palavras-chave, descrição, contato, produtor, distribuidor, datas, licenças de uso e de acesso, dentre outras. Servem principalmente para assegurar a localização dos dados. (Bezjak et al., 2018, p. 186). Em consonância com os princípios FAIR, em especial, o princípio *Findable* (encontrabilidade), os metadados são essenciais para que a informação seja encontrada.

“Podem ser classificados como descritivos (que servem para identificação e localização), administrativos (para criação, identificação de direitos, controle de acesso etc.) e estruturais (para relacionar objetos digitais)”. (Metadado, 2016).

Metadados enriquecidos

São metadados “[...] apresentados ao usuário com descrições e características associadas aos vocabulários e às informações de proveniência, com semântica mais clara quando comparados ao metadado original (básico). [...] Facilita a atividade dos publicadores ou analistas de dados abertos, com automação parcial do processo de anotação semântica e reuso na geração de metadados enriquecidos; proporciona maior credibilidade aos metadados, uma vez que permite a publicação de metadados com informações de proveniência; promove maior usabilidade dos dados e metadados, uma vez que os metadados são enriquecidos com anotações que fazem uso de padrões de vocabulários já existentes e ontologias específicas. Um conjunto de metadados enriquecidos deve apresentar algumas características como: (i) maior quantidade de atributos semânticos, que contêm um significado claro, capaz de descrever melhor os dados; (ii) facilidade de interpretação e processamento do conteúdo dos conjuntos de dados, uma vez que o metadado estará em linguagem estruturada, ou seja, definida de forma que pode ser processada por máquina; (iii) termos de vocabulários padrões associados, que permitirão ao metadado se integrar a outros dados e recursos na Web.” (Lira, 2014, p. 42-43).

Metadados FAIR

Metadados cujo preenchimento satisfaça os princípios FAIR, contribuindo para que os dados possam ser encontráveis, acessíveis, interoperáveis e reutilizáveis. (Souza et al., 2020).

Métricas de qualidade

“Uma métrica de qualidade é uma definição operacional que descreve, em termos específicos, um atributo do projeto ou do projeto, e como o processo de controle da qualidade irá medi-lo.” (Montes, 2020).

Métricas FAIR

Métricas que medem o nível de aderência e o grau de compatibilidade dos objetos digitais (conjuntos de dados) aos princípios FAIR. (Henning et al., 2019).

NAL

Acrônimo de *National Agricultural Library*, uma das cinco bibliotecas nacionais dos Estados Unidos. Abriga uma das maiores coleções do mundo dedicada à agricultura e ciências correlatas. (Estados Unidos, 2021).

NAL Thesaurus

Vocabulário controlado desenvolvido e mantido pela *National Agricultural Library*, com abrangência ampla em agricultura, biologia e disciplinas correlatas. (Estados Unidos, 2020).

NCBI

Acrônimo de *National Center for Biotechnology Information*, estabelecido em 1988 como uma divisão da *U. S. National Library of Medicine* (NLM) no *U. S. National Institute of Health* (NIH). O NCBI é encarregado de criar sistemas automatizados para armazenar e analisar conhecimentos sobre biologia molecular, bioquímica e genética; facilitar o uso de tais bancos de dados e software pela comunidade médica e de pesquisa; coordenar esforços para reunir informações sobre biotecnologia tanto nacional quanto internacionalmente; e realizar pesquisas em métodos avançados de processamento de informação baseado em computador para analisar a estrutura e função de moléculas biologicamente importantes. (National Center for Biotechnology Information, 2020a).

NCBI Taxonomy

Vocabulário controlado baseado em 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). (National Center for Biotechnology Information, 2020b).

Nome

Elemento de descrição do esquema Metadados de Citação. Nome de autor do *dataset*. Nome de autor pessoal ou de autor corporativo. (Souza et al., 2020).

Palavra ou locução com que se designa uma classe de coisas, pessoas, animais, um lugar, um acidente geográfico, um astro etc. (Nome, 2004, p. 2024).

Nome da série

Elemento de descrição do esquema Metadados de Citação. Nome da série do *dataset*. (Souza et al., 2020).

Nome do distribuidor

Elemento de descrição do esquema Metadados de Citação. Nome do distribuidor do *dataset*. Nome pessoal ou nome corporativo do distribuidor do *dataset*. (Souza et al., 2020).

Nome do órgão inanciador

Elemento de descrição do esquema Metadados de Citação. Nome do órgão financiador da produção do *dataset*. Nome corporativo do financiador do *dataset*. (Souza et al., 2020).

Nome do software

Elemento de descrição do esquema Metadados de Citação. Nome do software utilizado para gerar o *dataset*. (Souza et al., 2020).

Notas

Elemento de descrição do esquema Metadados de Citação. Informações adicionais sobre o *dataset*. (Souza et al., 2020).

Nota: “observação, geralmente em corpo menor, acrescentada para informar, esclarecer ou aprofundar o conhecimento sobre determinado assunto do texto.” (Brasil, 2013b, p. 33).

Número do identificador

Elemento de descrição do esquema Metadados de Citação. Número do identificador digital único que remete para a publicação relacionada ao *dataset*. (Souza et al., 2020).

Número do contrato

Elemento de descrição do esquema Metadados de Citação. Número do contrato de concessão de financiamento dos recursos para a produção do *dataset*. (Souza et al., 2020).

Objetividade

Atributo da qualidade da informação. “A informação não é tendenciosa e é imparcial.” (Trindade et al., 2011, p. 782).

Objetos digitais

“[...] podem representar, códigos, software, protocolos ou outros recursos de pesquisa.” (FAIR Data Maturity Model Working Group, 2020, p. 42, tradução nossa).

“Dados, código e outros resultados de pesquisa: dados, código e outros resultados de pesquisa. Em nível mais básico, os dados ou código são um fluxo de bits ou sequência binária. Para que tenham significado e sejam FAIR, precisam ser representados em formatos padrão e estar acompanhados de identificadores persistentes (PIDs), metadados e documentação”. (European Commission Expert Group on FAIR Data, 2018, p. 35, tradução nossa).

Ontologia

“Modelo formal que permite que o conhecimento seja representado para um domínio específico. Uma ontologia descreve os tipos de coisas que existem (classes), os relacionamentos entre eles (propriedades) e as maneiras lógicas como essas classes e propriedades podem ser usadas juntas (axiomas).” (Ontology, 2020a, tradução nossa).

“Um vocabulário controlado com um número maior e tipos mais específicos de relacionamentos entre os termos do que no caso de tesouros ou taxonomias.” (Ontology, 2020b, tradução nossa).

“Uma estrutura para conceitos inter-relacionados dentro de um domínio; uma ontologia vincularia os termos ‘vapor de água’, ‘umidade relativa’ e ‘pressão de vapor de H₂O’, de modo que um usuário procurando por um desses termos também veria os outros termos relacionados e seus relacionamentos.” (Strasser et al., 2011, p. 10, tradução nossa).

ORCID

Acrônimo de *Open Researcher and Contributor ID*, um identificador único persistente gratuito para pesquisadores; composto de 16 caracteres. ORCID está para o autor assim como o DOI está para um documento digital. (ORCID, 2021).

Organismo

Elemento de descrição do esquema Metadados Ciência da Vida. Nome taxonômico do organismo usado nos *dataset* ou de onde o material biológico se originou. (Souza et al., 2020).

Origem das fontes de dados

Elemento de descrição do esquema Metadados de Citação. Origem das fontes de dados do *dataset*. (Souza et al., 2020).

Outras referências

Elemento de descrição do esquema Metadados de Citação. Outras referências de apoio à produção do *dataset*. (Souza et al., 2020).

Outro identificador

Elemento de descrição do esquema Metadados de Citação. Outro identificador único digital que identifica o *dataset*. (Souza et al., 2020).

Padrão da comunidade

“Um padrão para metadados ou dados que é reconhecido e amplamente usado em uma comunidade.” (FAIR Data Maturity Model Working Group, 2020, p. 39, tradução nossa).

Padrão de disciplina

Ver Padrão da comunidade.

Padrão de domínio

Ver Padrão da comunidade.

Padrão específico da comunidade

Ver Padrão da comunidade.

Padrão de metadados

“Padrão de metadados é um documento de alto nível que estabelece uma maneira comum de estruturar e compreender os dados, e inclui princípios e questões de implementação para a utilização do padrão”. (University of Pittsburgh, 2020, tradução nossa).

“Um padrão de metadados é um requisito que se destina a estabelecer um entendimento comum do significado ou semântica dos dados, para garantir o uso correto e adequado e a interpretação dos dados por seus proprietários e usuários. Para alcançar esse entendimento comum, uma série de características ou atributos dos dados devem ser definidos, também conhecidos como metadados.” (Metadata, 2020, tradução nossa).

Principais padrões de metadados disponíveis: *Dublin Core Metadata Initiative*, *Data Documentation Initiative* (DDI), *Darwin Core*, *European Broadcasting Union* (EBU Core), *Encoded Archival Virtual* (EAD), *Resources Association Description* (VRA Core), *Ecological Metadata Language* (EML), *Content Standard for Digital Geospatial Metadata* (CSDGM), ISO 19115:2003 - *Geographic Information Metadata*, *MARC Standards*, *Metadata Object Description Schema* (MODS), *Public Broadcasting Metadata Dictionary* (PB Core). (Metadata..., 2020).

Padronização

Estabelecimento de termos padronizados para representar, de forma inequívoca, os conceitos de determinado campo do conhecimento visando eliminar a ambiguidade e o uso indiscriminado de sinônimos na comunicação científica. (Cunha; Cavalcanti, 2008, p. 272).

Palavras-chave

Elemento de descrição do esquema Metadados de Citação. Palavras-chave que expressam o conteúdo do *dataset*. (Souza et al., 2020).

Palavra significativa encontrada no título de um documento, no resumo ou no texto. Essa palavra (ou grupo de palavras) caracteriza o conteúdo temático do recurso e é usada em catálogos e índices de assuntos. (Cunha; Cavalcanti, 2008, p. 274).

“Termo que representa a síntese de um texto ou dos aspectos que nele são abordados.” (Bonotto et al., 2005, p. 14).

Paradigma

“Segundo T.S. Kuhn, o termo representa as formulações teóricas que servem implicitamente, por um período de tempo, para legitimar problemas e métodos em um campo de conhecimento.” (Cunha; Cavalcanti, 2008, p. 275).

Parâmetro

“Qualquer elemento cuja modificação de valor altera a solução de um problema, sem lhe alterar a sua natureza.” (Cunha; Cavalcanti, 2008, p. 276).

Período da coleta de dados

Elemento de descrição do esquema Metadados de Citação. Período da coleta de dados do *dataset*. (Souza et al., 2020).

Período da cobertura de dados

Elemento de descrição do esquema Metadados de Citação. Período da cobertura de dados do *dataset*. (Souza et al., 2020).

Plano de gestão de dados

“[...] documento que descreve que dados serão coletados ou gerados; quais as metodologias e padrões que serão utilizados nesses processos; se, como e sob que condições esses dados serão compartilhados e/ou tornados abertos para a comunidade de pesquisa; e como eles serão curados e preservados.” (Plano..., 2016).

“O plano de gestão de dados é o ponto de partida no ciclo de vida dos dados; no entanto é um documento que deve ser revisado frequentemente ao longo do projeto de pesquisa, para garantir que a documentação e o gerenciamento sejam adequados aos dados. [...] Um documento formal que descreve o que o pesquisador irá fazer com seus dados, durante e depois de concluir sua pesquisa. Esse documento garantirá que seus dados estejam seguros no presente e no futuro. A partir do objetivo do projeto, resultados e impactos esperados, plano de gestão descreve como os dados serão gerenciados e acessíveis ao longo de sua vida. [...] Os principais componentes de um plano de gestão de dados incluem informações sobre os dados e seu formato; conteúdo e formato de metadados; políticas de acesso, compartilhamento e reutilização, orçamento; e, planos para armazenamento e gerenciamento de longo prazo.” (DataONE, 2016, tradução nossa).

“Documentos que fornecem aos pesquisadores um mecanismo para declarar como eles gerenciarão os dados associados a pelo menos parte do ciclo de vida dos dados de um projeto de pesquisa.” (Smale et al., 2018, p. 2, tradução nossa).

Plano de gestão de dados acionável por máquina

“Os maDMP (PGDam) são um veículo para relatar as intenções e resultados de um projeto de pesquisa, que permite a troca de informações entre as partes e sistemas relevantes. Eles contêm um inventário das principais informações sobre um projeto e suas saídas (não apenas dados), com um histórico de alterações que as partes interessadas podem consultar para obter informações atualizadas sobre o projeto durante sua vida útil. A estrutura básica requer modelos de dados comuns para troca de informações, atualmente em desenvolvimento no RDA DMP Common Standards WG, bem como um ecossistema compartilhado de serviços que enviam notificações e agem em nome de humanos. Outros componentes da visão incluem políticas acionáveis por máquina, identificadores persistentes (PIDs) (por exemplo, IDs de ORCID, IDs de financiadores, próximos Org IDs, RRDs para recursos biomédicos, protocolos.io, IGSNs para amostras geográficas, etc.) e a remoção de barreiras para compartilhamento de informações.” (Scoping..., 2018, tradução nossa).

Plano de gestão de dados FAIR

Plano que descreve o ciclo de vida da gestão de dados visando à coleta, processamento e/ou gestão de dados, voltado à execução de um projeto de pesquisa. “Em geral, dados de pesquisa devem ser FAIR, ou seja, encontráveis, acessíveis, interoperáveis e reutilizáveis.” (European Commission, 2016, p. 4, 6, tradução nossa).

Plataforma tecnológica

Elemento de descrição do esquema Metadados Ciência da Vida. Fabricante ou nome da plataforma tecnológica usada no experimento para coleta de dados. (Souza et al., 2020).

Precisão

Atributo da qualidade da informação. “A informação deve estar isenta de erros. Nível de detalhe suficiente para o uso que se destina.” (Trindade et al., 2011, p. 783).

Qualidade informativa dos documentos recuperados por um sistema, para atendimento de pedidos relativos a temas específicos. Acurácia. Refere-se “[...] à especificidade e à qualidade dos documentos localizados.” (Cunha; Cavalcanti, 2008, p. 325).

Princípio

“Prescrição, valor fundamental ou diretiva de caráter geral que fornece sentido à norma e serve de critério para a compreensão de seus elementos, relações e funções.” (Embrapa, 2019, p. 10).

Princípios FAIR

Quatro princípios: Encontrabilidade (*Findable*), Acessibilidade (*Accessible*), Interoperabilidade (*Interoperable*) e Reusabilidade (*Reusable*), os quais fornecem diretrizes para os produtores, catalogadores, editores, curadores e publicadores de dados de pesquisa para melhorar a reutilização de dados científicos. Princípios FAIR enfatizam a acionabilidade por máquina; ou seja, a capacidade dos sistemas computacionais de encontrar, acessar, interoperar e reusar dados, sem a mínima intervenção humana. Por quê? Porque os humanos dependem cada vez mais do suporte computacional para lidar com ativos digitais, em razão do aumento no volume, complexidade e velocidade de criação de dados. (Go FAIR, 2018).

Privacidade

Privacidade examina até que ponto o uso de informações pessoais que o sistema detém sobre um usuário está em conformidade com as suposições explícitas ou implícitas em relação a esse uso. (Campos, 2020, slide 10).

Produtor

Elemento de descrição do esquema Metadados de Citação. Produtor do *dataset*. Nome pessoal ou nome corporativo do responsável financeiro ou administrativo do *dataset*. Pessoa física ou jurídica que toma a iniciativa e tem a responsabilidade econômica sobre o recurso produzido. (Souza et al., 2020).

“Organização responsável pela criação do conteúdo de um arquivo legível por máquina, podendo também realizar ou não a comercialização deste produto.” (Cunha; Cavalcanti, 2008, p. 294-295).

Protocolo

“Um método pelo qual um usuário ou máquina pode obter acesso a um recurso, por exemplo HTTP (S) ou FTP para acesso a recursos na Internet, ou um número de telefone e instruções para ligar para a pessoa ou organização que possui ou gerencia o recurso.” (FAIR Data Maturity Model Working Group, 2020, p. 42, tradução nossa).

Proveniência

“Indica a origem e procedência de documentos que, no decorrer das atividades de uma instituição, administração ou pessoa foram criados ou recebidos, acumulados e conservados.” (Cunha; Cavalcanti, 2008, p. 301).

Comumente significa ‘um registro de propriedade de uma obra de arte ou antiguidade, usado como um guia para autenticidade ou qualidade’. [...] A proveniência pode ajudar a estabelecer se uma obra de arte é (ou não) genuína.” O termo proveniência dos dados é usado “para significar ‘as informações necessárias para documentar com precisão o histórico dos dados, incluindo como foram criados e como foram transformados’. Sem a proveniência dos dados, podemos não ter total confiança em uma figura ou tabela resumida.” (Crosas et al., 2017, tradução nossa).

Proveniência refere-se a informações sobre entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, e que podem ser usadas para avaliações sobre sua qualidade, confiabilidade e fidedignidade.” (Groth; Moreau, 2013).

“Histórico de um arquivo de dados ou conjunto de dados, incluindo coleta, transformações, controle de qualidade, análises ou edição.” (Strasser et al., 2012, p. 11, tradução nossa).

“Dados relacionados aonde, quando e como as informações foram adquiridas.” (Provenance, 2013, tradução nossa).

“O termo “proveniência dos dados” se refere a uma trilha de registro que explica a origem de um dado (em um banco de dados, documento ou repositório) junto com uma explicação de como e por que ele chegou ao local atual. Exemplo: em um aplicativo como Biologia Molecular, muitos dados são derivados de bancos de dados públicos, que por sua vez podem ser derivados de papers, mas após algumas transformações (apenas os dados mais significativos foram colocados no banco de dados público), os quais são derivados de experimentos observacionais.” (Gupta, 2009, p. 608, tradução nossa).

Publicação relacionada

Elemento de descrição do esquema Metadados de Citação. Publicação relacionada ao *dataset*, a qual utiliza dados do *dataset* que está sendo descrito. (Souza et al., 2020).

Qualidade de dados

“Qualidade pode ser encarada como um conceito subjetivo sobre a percepção de um indivíduo em relação a um serviço, produto, dado, informação etc. De modo geral, pode ser definida como medidas para que o produto oferecido esteja de acordo com o que se espera dele, podendo este ser uma informação, um dado, um serviço ou um processo. [...] Quando se fala de qualidade aplicada a dados, significa que se espera que eles atendam à medida de perfeição, precisão e conformidade no domínio no qual estão inseridos. Dados que não descrevem fielmente componentes do mundo real diminuem a efetividade dos sistemas, contribuem de forma negativa para as atividades envolvidas em sua utilização e seus impactos podem ser tanto sociais quanto econômicos.” (Melo, 2017, p. 47).

Quantidade

Atributo da qualidade da informação. “O quanto o volume de informação é apropriado para a atividade que está sendo executada.” (Trindade et al., 2011, p. 783).

Query

“Refere-se ao termo usado na busca.” (Query, 2019, p. 417, tradução nossa).

RDA

Acronônimo de *Resource Description and Access*. Padrão para catalogação descritiva sucessor do AACR2. Conjunto de instruções de catalogação baseadas em FRBR e FRAD, para produzir a descrição e pontos de acesso (nome e título) que representam um recurso. RDA é um padrão para descrição de recursos e acesso projetado para o mundo digital. Fornece: 1) uma estrutura flexível para descrever todos os recursos (analógicos e digitais) que são extensíveis para novos tipos de material; 2) dados que são prontamente adaptáveis a estruturas de banco de dados novas e emergentes; 3) dados que são compatíveis com os registros existentes em catálogos on-line. RDA vai além dos códigos de catalogação anteriores, pois fornece diretrizes sobre catalogação de recursos digitais e dá maior ênfase em ajudar os usuários a encontrarem, identificarem, selecionarem e obterem as informações que desejam. (Resource..., 2021).

Recurso

Representação física de um conjunto de dados. Cada recurso pode ser um arquivo de qualquer tipo, um link para um arquivo em qualquer lugar da web ou um link para uma API. Por exemplo, se os dados estão sendo fornecidos em vários formatos ou divididos em diferentes áreas ou períodos de tempo, cada arquivo é um ‘recurso’ diferente que deve ser descrito individualmente. (Resource, 2020).

Referência

“Indicação bibliográfica elaborada de acordo com normas específicas.” (Cunha; Cavalcanti, 2008, p. 312).

Conjunto padronizado de elementos descritivos retirados de um documento, permitindo sua identificação no todo ou em parte. Esses elementos identificam documentos impressos ou registrados em qualquer suporte físico, como: livros, periódicos, materiais audiovisuais, materiais em formatos eletrônicos, acessados, capturados ou consultados sob diferentes formatos e protocolos como *http* (*hypertext transfer protocol*), usado pelo *www* (*World Wide Web*), *ftp* (*file transfer protocol*) etc. (Embrapa, 2013, p. 20).

Relevância

Atributo da qualidade da informação. “A informação deve estar relacionada com o interesse e as necessidades de informação do usuário.” (Trindade et al., 2011, p. 783).

Repositório de dados de pesquisa

Local “[...] mais apropriado para o arquivamento voltado para o compartilhamento e para a preservação. Os repositórios de dados de pesquisa são banco de dados que recebem, gerenciam e disseminam dados e conjuntos de dados de pesquisa. “Os repositórios de dados de pesquisa são bancos de dados que recebem, gerenciam e disseminam dados e conjuntos de dados de pesquisa. Eles constituem o lugar mais apropriado para que seus dados sejam preservados e possam ser re-cuperados, acessados e citados por outros pesquisadores, ou seja, tenham visibilidade em escala mundial.” (Sayão; Sales, 2015, p. 50). Há diferentes tipos de repositórios e centros de dados para hospedagem de dados assim denominados: repositórios institucionais, repositórios temáticos, repositórios de dados de artigos de periódicos e repositórios governamentais.

Repositórios institucionais são aqueles apoiados por instituições, sobretudo, de pesquisa e de ensino, por exemplo: Alice - Repositório Acesso Livre à Informação Científica da Embrapa, que se destina a reunir, organizar, armazenar, preservar e disseminar, na íntegra, informações científicas produzidas por pesquisadores da Embrapa; Infoteca-e - Repositório de Informação Tecnológica da Embrapa, que reúne e permite acesso a informações sobre tecnologias produzidas pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa), as quais se relacionam às áreas de atuação de suas demais unidades de pesquisa; ARCA - Repositório Institucional da Fiocruz (<https://www.arca.fiocruz.br/>), que reúne, hospeda, preserva, disponibiliza e dá visibilidade à produção intelectual da Fundação Oswaldo Cruz. (Nascimento et al., 2019).

Repositórios temáticos são aqueles dedicados à cobertura de determinada área de conhecimento, por exemplo: Repositório Institucional de Produção Científica da Escola Nacional de Saúde Pública (<http://www6.ensp.fiocruz.br/repositorio/>), que reúne a produção científica gerada pela Instituição. (Escola Nacional de Saúde Pública Sergio Arouca, 2020).

Repositórios de dados de artigos de periódicos são “[...] repositórios vinculados a um ou mais periódicos científicos. Neste tipo de repositório os autores depositam os dados relativos aos artigos publicados pelos periódicos. Por exemplo: DRYAD (<https://datadryad.org>) que armazena dados de artigos revisados por pares de mais de 150 periódicos da área de biociências. (Sayão; Sales, 2015, p. 51).

Repositórios governamentais são aqueles apoiados por agências governamentais, por exemplo: DATA.GOV (www.data.gov), que reúne dados e informações de todas as esferas governamentais norte-americanas. (Data.gov, 2021).

Repositório institucional

“Repositório público acessível na internet que visa armazenar, preservar e disseminar dados ou informações produzidos pela instituição.” (Embrapa, 2019, p. 10).

Sistemas de informação que servem para armazenar, preservar, organizar e disseminar amplamente os resultados de pesquisa de instituições de ensino e pesquisa, utilizando um software. No Brasil, software mais conhecido é o DSpace. (Repositório..., 2020).

Um local on-line para coletar, preservar e divulgar materiais como artigos de revistas e versões digitais de teses e dissertações. Pode também incluir outros recursos digitais gerados por uma instituição de pesquisa, como documentos administrativos, notas de curso ou objetos de aprendizagem. (Ibict, 2016).

Exemplos: Repositório Acesso Livre à Informação Científica da Embrapa (Alice), que reúne a produção científica resultante das pesquisas geradas pela Empresa. (Embrapa, 2020); Repositório de Informação Tecnológica da Embrapa (Infoteca-e), que reúne e permite acesso a informações sobre tecnologias produzidas pela Embrapa. (Embrapa, 2021).

Representação descritiva

Representação descritiva, ou catalogação, é conhecida como “[...] processo de decisão multidimensional que estrutura e padroniza os diferentes aspectos de um item informacional, tornando-o único e passível de recuperação e uso. (Aganette et al., 2017, p. 177).

Representação do conhecimento

“Tanto (a) um conjunto de conceitos em um vocabulário controlado, uma ontologia ou tesauro, ou (b) um modelo de dados, ou seja, uma estrutura bem definida para descrever e estruturar metadados ou dados.” (FAIR Data Maturity Model Working Group, 2020, p. 40, tradução nossa).

Reputação

Atributo da qualidade da informação. “A informação é considerada verdadeira com relação à sua fonte ou conteúdo.” (Trindade et al., 2011, p. 783).

Reusabilidade

Reusabilidade (*Reusable*) é um dos princípios FAIR. Princípio que se relaciona à capacidade dos dados e metadados serem reusáveis, para replicarem e/ou se combinarem para que possam ser reusados. (Go FAIR, 2018, tradução nossa).

Reúso

“Uso dos dados para propósitos diferentes daquele pelo qual foram originalmente coletados. Geralmente são usados por outros pesquisadores que não os autores dos dados.” (Reúso, 2016).

“Uso dos dados para uma finalidade diferente daquela para a qual foram coletados.” (Strasser et al., 2011, p. 11, tradução nossa).

Revisão na catalogação

Atividade que revisa a catalogação e contribui para que os metadados se tornem aderentes aos princípios FAIR, e que possam usufruir os benefícios de serem encontráveis, acessáveis, interoperáveis e reutilizáveis. (Souza et al., 2020).

A revisão na catalogação mira dois alvos: a completeza, que expressa o quanto os metadados representam o recurso digital descrito; a precisão, que diz respeito ao corretismo das informações fornecidas na descrição do recurso. (Graham, 1990 citado por Branco et al., 2014, p. 2439).

Revisor da catalogação

Profissional com formação em Biblioteconomia ou em Ciência da Informação, designado para a tarefa de revisão na catalogação. (Souza et al., 2020, p. 19-20).

Segurança

Atributo da qualidade da informação. “O acesso à informação é mantido restrito apropriadamente para garantir a sua segurança.” (Trindade et al., 2011, p. 783).

Série

Elemento de descrição do esquema Metadados de Citação. Série do *dataset*. (Souza et al., 2020).

“Unidade de arquivamento, ou seja, documentos ordenados de acordo com o quadro de arranjo, ou conservados como uma unidade porque se relacionam às mesmas atividades e funções ou ao mesmo tipo documental. Em administração de arquivos, corresponde a certas peças que apresentam uma ou mais características comuns.” (Cunha; Cavalcanti, 2008, p. 332).

“Conjunto ilimitado de publicações não periódicas sobre um tem específico, com autores e títulos próprios, reunidos sob título comum.” (Brasil, 2013b, p. 40).

Sigla

“Caso especial de abreviatura formada pelas letras ou sílabas iniciais de um nome geralmente composto.” (Brasil, 2013b, p. 41).

Sistema de banco de dados

“[...] é um sistema computadorizado que permite que os usuários criem e mantenham um banco de dados. O SGBD é um sistema de software de uso geral que facilita o processo de definição, cons-

trução, manipulação e compartilhamento de bancos de dados entre diversos usuários e aplicações.” (Elmasri; Navathe, 2018, p. 5).

“Um Sistema de Gerenciamento de Banco de Dados (SGBD) - do inglês Data Base Management System (DBMS) — é o sistema de software responsável pelo gerenciamento de um ou mais bancos de dados. Seu principal objetivo é retirar da aplicação cliente a responsabilidade de gerenciar o acesso, a persistência, a manipulação e a organização dos dados. O SGBD disponibiliza uma interface para que seus clientes possam incluir, alterar ou consultar dados previamente armazenados.” (Sistema..., 2021).

Sistema identificador

Elemento de descrição do esquema Metadados de Citação. Sistema identificador digital do autor do *dataset*. (Souza et al., 2020).

SKOS

Acrônimo de *Simple Knowledge Organization System* (em português, Sistema Simples de Organização do Conhecimento). “Uma recomendação do Consórcio W3C para representação de tesouros, esquemas de classificação, taxonomias, sistemas de cabeçalho de assunto ou qualquer outro tipo de vocabulário controlado estruturado.” (Simple..., 2020, tradução nossa).

Software

Elemento de descrição do esquema Metadados de Citação. Software utilizado para gerar o *dataset*. (Souza et al., 2020).

String

“Sequência de caracteres (letras, números ou símbolos) de entrada ou saída de dados.” (Alves et al., 2008, p. 56).

Subtítulo

Elemento de descrição do esquema Metadados de Citação. Subtítulo do *dataset*. Nome secundário atribuído ao *dataset* para generalizar ou especificar a abrangência do título. (Souza et al., 2020).

“Título secundário apresentado após o título, que visa esclarecê-lo ou complementá-lo de acordo com o conteúdo da obra.” (Brasil, 2013b, p. 41).

“Informações apresentadas em seguida ao título, visando esclarecê-lo ou complementá-lo de acordo com o conteúdo da publicação.” (Associação Brasileira de Normas Técnicas, 2018, p. 3).

Taxonomia

“Uma coleção de termos de vocabulário controlado organizados em uma estrutura hierárquica. Cada termo em uma taxonomia está em um ou mais relacionamentos pai/filho (mais amplo/mais restrito) com outros termos na taxonomia.” (Taxonomy, 2020, tradução nossa).

“Taxonomia é uma representação formal dos relacionamentos entre os itens em uma estrutura hierárquica.” (Taxonomy, 2013, tradução nossa).

Taxonomia multifacetada de dados

Ferramenta conceitual e operacional para organização e representação dos tipos de dados que permite reunir, classificar e relacionar a multidimensionalidade que as propriedades e naturezas de um dado podem manifestar em diferentes contextos e domínios do conhecimento nos quais ele pode ser percebido, registrado e usado. A ferramenta é construída com base nos referenciais teórico, conceitual, metodológico e tecnológico das aplicações dos Sistemas de Organização do Conhecimento (SOC) e da Engenharia da Informação. Assim, por exemplo, um dado de temperatura, 27°C, pode ser operacionalizado em sistemas de informação, como um dado alfanumérico, quanti ou qualitativo, primário ou secundário, sigiloso ou aberto etc., de acordo com o propósito de uso ou finalidade e alinhado com as etapas decorrentes de gestão de dados às quais será submetido, entre elas coleção, planejamento, estruturação, curadoria, análise e serviços decorrentes de dados. (Pierozzi Júnior et al., 2020).

Termo

“Palavra, rótulo semântico. Palavra ou expressão (locução) empregada para a inclusão temática de um item em sistema de informação e para a recuperação posterior. Palavra ou expressão empregada para denotar um conceito.” (Cunha; Cavalcanti, 2008, p. 360).

“Uma entrada em um vocabulário controlado, esquema, taxonomia ou ontologia.” (Term, 2013, tradução nossa).

Termo de classificação

Elemento de descrição do esquema Metadados de Citação. Termo de classificação do *dataset*. (Souza et al., 2020).

Termo de indexação

“Representação de um assunto por meio de um termo derivado da linguagem natural, de preferência um nome ou uma locução nominal ou um símbolo de classificação”. (Bonotto et al., 2005, p. 18).

Termo específico

“Termo preferido que representa um conceito que é mais específico que aquele em questão. [...] Termo que, numa relação genérica, expressa um conceito cuja intenção é mais ampla.” (Termo específico, 2015).

Termo genérico

“Termo preferido que representa um conceito mais amplo em relação a outro em questão. O escopo do termo específico fica completamente no escopo do termo genérico. [...] Termo que, numa relação genérica, expressa um conceito cuja extensão é mais ampla.” (Termo genérico, 2015).

Termo livre

“Termo simples ou composto, retirado da linguagem natural, com valor significativo, não incluído na relação de descritores padronizados de um tesauro ou vocabulário controlado.” (Bonotto et al., 2005, p. 17).

Termo preferido

“Termo que deve ser aceito e usado por especialistas no assunto. [...] Um termo recomendado por um comitê técnico deve ser considerado um termo preferido, enquanto um termo admitido deve representar um sinônimo aceitável para um termo preferido. Termos descontinuados devem ser rejeitados.” (International Organization for Standardization, 2009, p. 34-36, tradução nossa).

Termo-chave

Elemento de descrição do esquema Metadados de Citação. Termo-chave das palavras-chave do *dataset*. (Souza et al., 2020).

Temporalidade

Atributo da qualidade da informação. “A informação está suficientemente atualizada. A idade da informação é adequada para a necessidade do usuário.” (Trindade et al., 2011, p. 783).

Tesauro

“Vocabulário estruturado e controlado no qual os conceitos são representados por termos, organizados de modo que as relações entre os conceitos sejam explicitadas e os termos preferenciais sejam acompanhados por entradas iniciais para sinônimos ou quase-sinônimos.” (Thesaurus, 2020, tradução nossa).

Instrumentos de controle terminológico utilizados para representar conteúdos de documentos e solicitações de busca. “Os tesouros são estruturas sistemáticas de conceitos. Estas estruturas conceituais apresentam termos e relacionamentos entre termos. Os relacionamentos expressos

no tesauro podem ser classificados em hierárquicos (gênero e espécie e partitivos), associativos e de equivalência. (Campos, 2001 citado por Campos et al., 2008, p. 3).

“Vocabulário controlado e dinâmico de descritores relacionados semântica e genericamente, que cobrem de forma extensiva um ramo específico de conhecimento.” (Cunha; Cavalcanti, 2008, p. 362).

Texto

Elemento de descrição do esquema Metadados de Citação. Texto do elemento metadado Descrição do *dataset*. Texto sucinto contendo o propósito, a natureza e o escopo do *dataset*. (Souza et al., 2020).

Tipo de célula

Elemento de descrição do esquema Metadados Ciência da Vida. Nome da linhagem de células usadas no *dataset* e das quais as amostras derivam. (Souza et al., 2020).

Tipo de colaborador

Elemento de descrição do esquema Metadados de Citação. Tipo de colaborador do *dataset*. (Souza et al., 2020).

Tipo de dados

Elemento de descrição do esquema Metadados de Citação. Tipo de dados do *dataset*. (Souza et al., 2020).

Tipo de design

Elemento de descrição do esquema Metadados Ciência da Vida. Definição geral do desenho experimental utilizado na produção do *dataset*. (Souza et al., 2020).

Tipo de fator

Elemento de descrição do esquema Metadados Ciência da Vida. Nome taxonômico do organismo usado no *dataset* ou de onde o material biológico se originou. (Souza et al., 2020).

Tipo de identificador

Elemento de descrição do esquema Metadados de Citação. Tipo de identificador digital utilizado pela publicação relacionada ao *dataset*. (Souza et al., 2020).

Tipo de medida

Elemento de descrição do esquema Metadados Ciência da Vida. Termo utilizado para qualificar o que está sendo medido. (Souza et al., 2020).

Tipo de tecnologia

Elemento de descrição do esquema Metadados Ciência da Vida. Termo utilizado para identificar a tecnologia empregada para coletar os dados. (Souza et al., 2020).

Tipologia de dados de pesquisa

Modelo conceitual de representação dos fenômenos de pesquisa que são objetos de estudo dos pesquisadores, e que possibilita um conhecimento mais qualificado sobre o(s) tipo(s) de dado(s) que poderão subsidiar a pesquisa. (Pierozzi Júnior et al., 2020).

Sistema conceitual de organização do conhecimento que considera a ampla variedade de dados gerados pela pesquisa, e que os organiza sob uma estrutura de categorias previamente definidas e capazes de reunir e classificar dados de pesquisa em grupos de características ou propriedades semelhantes. (Pierozzi Júnior, 2020).

Tipos de dados de pesquisa

“Os dados de pesquisa podem diferir em suas tipologias. Dependendo do assunto ou da disciplina científica, as definições utilizadas podem incluir uma vasta tipologia de objetos digitais e não digitais, como documentos (texto, palavras), memorandos, notas, avaliações, registros de casos, protocolos de estudo, planilhas, notas de laboratório, notas de campo, diários, questionários, transcrições, filmes, fitas de áudio e vídeo, arquivos digitais de áudio e vídeo, proteínas, sequências genéticas, DNA, cristalografia, coordenadas geográficas, slides, artefatos, espécimes, organismos, coleções de objetos digitais adquiridos e/ou gerados durante o processo de pesquisa, conteúdo em banco de dados (vídeo, áudio, texto, imagens), modelos, algoritmos, scripts de uma aplicação de computador (entrada, saída, arquivos de log para software de análise, software de simulação, esquemas), metodologias, fluxos de trabalho, informações técnicas, procedimentos e protocolos.” (Dudziak, 2016; Henderson, 2017 citados por Semeler; Pinto, 2019, p. 119).

Título

Elemento de descrição do esquema Metadados de Citação. Título atribuído ao *dataset*, e pelo qual é formalmente conhecido. (Souza et al., 2020).

“Palavra, frase, caractere ou grupo de caracteres que, normalmente, aparecem num item, dando nome a este ou à obra nele contida.” (Cunha; Cavalcanti, 2008, p. 364).

“Frase ou expressão no início da publicação, capítulo, seção, ilustração para identificar, distinguir, individualizar a obra ou para indicar o assunto tratado.” (Brasil, 2013b, p. 42).

Título alternativo

Elemento de descrição do esquema Metadados de Citação. Título alternativo atribuído ao *dataset*, sendo geralmente o nome pelo qual é comumente conhecido. (Souza et al., 2020).

Tratamento da informação

“Conjunto de ações referentes à produção, recepção, categorização, utilização, reprodução, transmissão, distribuição, ao acesso, transporte, arquivamento, armazenamento, avaliação e destinação (eliminação ou guarda permanente) ou ao controle da informação restrita ou sigilosa.” (Embrapa, 2019, p. 10).

URI

Acrônimo de *Uniform Resource Identifier*. “Um conjunto genérico de nomes ou endereços, que consiste em uma string de caracteres que remetem para um recurso. URI compreende as URLs e URNs, que são tipos de URIs.” (Alves et al., 2008, p. 56).

“Uma *string* que identifica com exclusividade praticamente qualquer coisa, incluindo um edifício físico ou conceitos mais abstratos, como cores. Pode ou não ser resolvível na web.” (URI, 2020, tradução nossa).

URL

Acrônimo de *Uniform Resource Locator*. “Localizador que permite identificar e acessar um serviço na rede web.” (Alves et al., 2008, p. 56).

“Um identificador global comumente chamado de ‘endereço da web’. Um URL pode ser resolvível na web. Todos os URLs HTTP são URIs; no entanto, nem todos os URIs são URLs.” (URL, 2020, tradução nossa).

URL alternativa

Elemento de descrição do esquema Metadados de Citação. URL alternativa válida que aponta para onde o *dataset* também pode ser visualizado. (Souza et al., 2020).

URL da publicação

Elemento de descrição do esquema Metadados de Citação. URL da válida da página web onde se encontra a publicação relacionada ao *dataset*. (Souza et al., 2020).

URL do distribuidor

Elemento de descrição do esquema Metadados de Citação. URL válida da página web do distribuidor do *dataset*. (Souza et al., 2020).

URL do logotipo do distribuidor

Elemento de descrição do esquema Metadados de Citação. URL válida do logotipo do distribuidor *dataset*. (Souza et al., 2020).

URL do logotipo do produtor

Elemento de descrição do esquema Metadados de Citação. URL válida do logotipo do produtor do *dataset*. (Souza et al., 2020).

URL do produtor

Elemento de descrição do esquema Metadados de Citação. URL válida da página web do produtor do *dataset*. (Souza et al., 2020).

URL do termo-chave

Elemento de descrição do esquema Metadados de Citação. URL válida do termo-chave do *dataset*. (Souza et al., 2020).

URL do termo de classificação

Elemento de descrição do esquema Metadados de Citação. URL válida do termo de classificação do *dataset*. (Souza et al., 2020).

URL do vocabulário controlado

Elemento de descrição do esquema Metadados de Citação. URL do vocabulário controlado onde o termo poderá ser acessado. (Souza et al., 2020).

URN

Acrônimo de *Uniform Resource Name*. Refere-se ao “[...] nome e endereço de um objeto na internet que possui certa garantia de persistência dos recursos informacionais.” (Cunha; Cavalcanti, 2008, p. 371).

“[...] um tipo de URI que usa URN Scheme e que tem por objetivo a identificação única do recurso, de forma persistente e independente da sua localização. Nome histórico de um URI.” (URN, 2020, tradução nossa).

Valor

Atributo da qualidade da informação. “A informação provê benefícios e vantagens com o seu uso.” (Trindade et al., 2011, p. 783).

Verbetes

“Definições e acepções, explicações relativas a uma expressão, palavra, termo ou vocábulo, conforme se apresentam em dicionários, enciclopédias, vocábulos e obras similares.” (Cunha; Cavalcanti, 2008, p. 736).

“Conjunto articulado de informações referentes a uma unidade de descrição.” (Verbetes, 2003, p. 247).

Versão do software

Elemento de descrição do esquema Metadados de Citação. Versão do software utilizado para gerar o *dataset*. (Souza et al., 2020).

Vocabulário compatível com FAIR

“Um vocabulário que é documentado, resolvido e compreensível por máquina usando identificadores globalmente únicos e persistentes.” (FAIR Data Maturity Model Working Group, 2020, p. 40, tradução nossa).

Vocabulário controlado

“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.” (Cunha; Cavalcanti, 2008, p. 378).

“Lista de termos autorizados predefinidos para serem usados na indexação (humana ou automatizada; uma forma de organizar o conhecimento para posterior recuperação.” (Controlled..., 2020, tradução nossa).

“Conjunto de termos cuidadosamente selecionados que são usados para descrever unidades de informação; usado para criar taxonomias, tesouros e ontologias. Em configurações tradicionais, os termos nos vocabulários controlados são palavras ou frases, em uma configuração de dados vinculados, normalmente são atribuídos identificadores únicos (URIs) que, por sua vez, se vinculam a frases descritivas.” (Controlled..., 2013, tradução nossa).

“Um conjunto de termos ou conceitos que podem ser usados na descrição de um recurso. Isso inclui taxonomias, ontologias e tesouros.” (FAIR Data Maturity Model Working Group, 2020, p. 44, tradução nossa).

Vocabulário controlado da classificação

Elemento de descrição do esquema Metadados de Citação. Vocabulário controlado utilizado para a classificação de assunto do *dataset*. (Souza et al., 2020).

Vocabulário controlado do termo-chave

Elemento de descrição do esquema Metadados de Citação. Vocabulário controlado utilizado para a atribuição de termos-chave do *dataset*. (Souza et al., 2020).

W3C

Acrônimo de *World Wide Web Consortium*, uma comunidade internacional composta por organizações membros e por equipes de trabalho em tempo integral voltadas para o desenvolvimento de padrões web (The World Wide Web Consortium, 2020).

Índice

A

AACR2
ABNT NBR 6023
Abreviatura
Abreviatura do nome do distribuidor
Abreviatura do nome do produtor
Acessibilidade
Acurácia
Acurácia da informação
Acurácia/precisão

Acurácia/veracidade
Afiliação do autor
Afiliação do distribuidor
Afiliação do produtor
Agência
AGROVOC *Thesaurus*
Amplitude
Análise de assunto
Anonimização dos dados
Apresentação
Assunto
Atualidade
Atribuição
Atributo
Autenticação
Autodepositante
Autodepósito
Autor
Autor corporativo
Autorização
Autor pessoal
Avaliação de qualidade

B

Banco de dados
Best practices
Biblioteconomia de dados
Big Data

C

Características das fontes de dados observadas
Catalogação
Catalogação de dados
Catalogador

Categoria
Ciclo de vida dos dados
Ciclo de vida dos dados de pesquisa
Ciclo de vida dos metadados
Ciência aberta
Citação
Clareza
Classificação
Classificação de *datasets*
Coautor
Coautoria
Coerência
Colaborador
Compilador
Completeza
Completude
Concisão
Confiabilidade
Conformidade
Conhecimento
Consistência
Contato
Controle de acesso
Controle de qualidade
Creative Commons
Credibilidade da informação
CSV
Curador de dados
Curadoria digital

D

Dado anonimizado
Dados
Dados abertos
Dados biológicos
Dados brutos
Dados de pesquisa
Dados FAIR
Data
Dataset
Datasets relacionados
Data citation
Data de distribuição
Data de início
Data de produção
Data de término
Data do depósito

Depositante

Data steward
Data stewardship
Data Stewardship Wizard
Dataverse
Default
Definição
Depositante
Depósito de dados
Descoberta
Descrição
Descrição de metadados
Desempenho
Dicionário de dados
Dimensões da qualidade da informação
Direito autoral
Direito moral do autor
Direito patrimonial do autor
Disponibilidade
Distribuidor
Documentação e acesso a fontes de dados
DOI
DSpace
Dublin Core

E

Elemento metadados
E-mail do contato
Encontrabilidade
Entrada
e-Science

F

Financiamento
Fontes de dados
Formato de metadados
FRAD
FRBR

G

Garantia de qualidade
Garantia de qualidade de dados
Gestão
Gestão de dados
Gestão de dados de pesquisa
Gestão do conhecimento
Glossário
Governança
Governança de dados
Guestbook

I

Identificador
Identificador digital persistente
Identificador global único e persistente
Idioma
Indexação
Indicadores FAIR
Indicadores de maturidade FAIR
Índice
Informação
Informação sobre a série
Inteligibilidade
Interoperabilidade
Interpretabilidade

L

Licenças *Creative Commons*
Linguagem artificial
Linguagem documentária
Linguagem livre
Linguagem natural
Livro de visita
Local de produção Logotipo

M

Mashup
Material relacionado
Melhores práticas
Metadados
Metadados enriquecidos
Metadados FAIR
Métricas de qualidade
Métricas FAIR

N

NAL
NAL *Thesaurus*
NCBI
NCBI *Taxonomy*
Nome
Nome da série
Nome do distribuidor
Nome do órgão financiador
Nome do software
Notas
Número do identificador
Número do contrato

O

Objetividade
Objetos digitais
Ontologia
ORCID
Organismo
Origem das fontes de dados
Outras referências
Outro identificador

P

Padrão da comunidade
Padrão de pesquisa
Padrão de domínio
Padrão específico da comunidade
Padrão de metadados
Padronização
Palavras-chave
Paradigma
Parâmetro
Período da coleta de dados
Período da cobertura de dados
Plano de gestão de dados
Plano de gestão de dados acionável por máquina
Plano de gestão de dados FAIR
Plataforma tecnológica
Precisão
Princípio
Princípios FAIR
Privacidade
Produtor
Protocolo
Proveniência
Publicação relacionada

Q

Qualidade de dados
Qualidade de metadados
Quantidade
Query

R

RDA
Recurso
Referência
Relevância
Repositório de dados de pesquisa

Repositório institucional
Representação descritiva
Representação do conhecimento
Reputação
Reusabilidade
Reúso
Revisão na catalogação Revisor
da catalogação

S

Segurança
Série
Sigla
Sistema de banco de dados
Sistema do identificador
SKOS
Software
String
Subtítulo

T

Taxonomia
Taxonomia multifacetada de dados
Termo
Termo de classificação
Termo de indexação
Termo específico
Termo genérico
Termo livre
Termo preferido
Termo-chave
Temporalidade
Tesauro
Texto
Tipo de célula
Tipo de colaborador
Tipo de dados
Tipo de design
Tipo de fator
Tipo de identificador
Tipo de medida
Tipo de tecnologia
Tipologia de dados de pesquisa
Tipos de dados de pesquisa
Título
Título alternativo
Tratamento da informação

U

URI
URL

URL alternativa
URL da publicação
URL do distribuidor
URL do logotidpo do distribuidor
URL do logotidpo do produtor
URL do produtor
URL do termo-chave
URL do termo de classificação
URL do vocabulário controlado
URN

V

Valor
Verbetes
Versão do software
Vocabulário compatível com FAIR
Vocabulário controlado
Vocabulário controlado da classificação
Vocabulário controlado do termo-chave

W

W3C

Referências

- ABRAHAM, R.; SCHNEIDER, J.; VOM BROCKE, J. Data governance: a conceptual framework, structured review, and research agenda. **International Journal of Information Management**, v. 49, p. 424-438, Dec. 2019. DOI: 10.1016/j.ijinfomgt.2019.07.008.
- AGANETTTE, E. C.; TEIXEIRA, L. M. D.; AGANETTE, K. de J. P. A representação descritiva nas perspectivas do século XXI: um estudo evolutivo dos modelos conceituais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 22, n. 50, p. 176-187, set./dez. 2017. DOI: 10.5007/1518-2924.2017v22n50p176.
- ALBERTONI, R.; BROWNING, D.; COX, S.; GONZALEZ BELTRAN, A.; PEREGO, A.; WINSTANLEY, P. (Ed.). **Data Catalog Vocabulary (DCAT) - version 2**. [Cambridge, MA]: World Wide Con-sortium (W3C), 2020. W3C Recommendation. Disponível em: <<https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/>>. Acesso em: 20 set. 2020.
- ALVES, M. das D. R.; SOUZA, M. I. F.; SANTOS, A. D. dos. **Manual de representação descritiva de recursos eletrônicos no Sistema Agência**. Campinas: Embrapa Informática Agropecuária, 2008. 142 p. (Embrapa Informática Agropecuária. Documentos, 81). Disponível em: <<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/48773>>. Acesso em: 27 ago. 2021.
- ALVORCEM, R. M.; DANTAS, J. de O. **Glossário administrativo da Embrapa**. Brasília, DF: Embrapa, 2020. 257 p. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1120598/1/GlossarioadministrativodaEmbrapa.pdf>>. Acesso em: 16 abr. 2021.
- ANONYMISATION. In: FAO. Agricultural Information Management Standards Portal. **Glossary**. Rome: AIMS, 2020. Disponível em: <<http://aims.fao.org/capacity-development/glossary/anonymisation>>. Acesso em: 12 jan. 2021.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT**: diretiva 2 – regras para estrutura e redação de documentos técnicos ABNT. Rio de Janeiro, 2017. 67 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR 6022**: informação e documentação – artigo em publicação periódica técnica e/ou científica impressa – apresentação. Rio de Janeiro, 2018. 8 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação – referências: elaboração. 2. ed. Rio de Janeiro, 2020. 68 p. Versão corrigida 2.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 10520**: informação e documentação - apresentação de citações em documentos - apresentação. Rio de Janeiro, 2002. 7 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 13789**: terminologia – princípios e métodos – elaboração e apresentação de normas de terminologia. Rio de Janeiro, 1997a. 17 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 13790**: terminologia – princípios e métodos – harmonização de conceitos e termos. Rio de Janeiro, 1997b. 6 p.

AUTOR. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 13 jan. 2021.

BALESTERO, M. de S.; ALMEIDA, G. M. de B.; PIEROZZI JUNIOR, I. Quando o especialista de domínio e as novas tecnologias entram em cena: impactos na definição terminológica. **Linguasagem**, São Carlos, v. 30, n. 1, p. 1-27, jan./jun. 2019. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/197125/1/AP-Linguasagem-Ivo.pdf>>. Acesso em: 22 jun. 2021.

BARROWMAN, N. Why data is never raw: on the seductive myth of information free of human judgement. **The New Atlantis**, n. 56, Summer/Fall 2018. Disponível em: <<https://www.thenewatlantis.com/publications/why-data-is-never-raw>>. Acesso em: 26 out. 2021.

BEZJAK, S.; CONZETT, P.; FERNANDES, P.; GÖRÖGH, E.; HELBIG, K.; KRAMER, B.; LABASTIDA, I.; NIEMEYER, K.; PSOMOPOULOS, F.; ROSS-HELLAUER, T.; SCHNEIDER, R.; TENNAT, J.; VERBAKEL, E.; CLYBURNE-SHERIN, A.; BRINKEN, H.; HELLER, L. **Open science training handbook**: version 1.0. 2018. 205 p. DOI: 10.5281/zenodo.1212496.

BIOLOGICAL data. In: WIKIPEDIA: the free encyclopedia. 2021. Disponível em: <https://en.wikipedia.org/wiki/Biological_Data>. Acesso em: 9 ago. 2021.

BONOTTO, M. E. K. K.; FERREIRA, G. I. S.; FLORES, H. F. de; BITTENCOURT, I. C.; DIAS, M. L.; VAN DER LAAN, R. H. **Glossário de ciência da informação**: glossário da área de organização e tratamento da informação. Porto Alegre: Faculdade de Biblioteconomia e Comunicação da UFRGS, [2005].

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. **Bibliotecas universitárias e o acesso público à informação**: articulando leis, tecnologias, práticas e gestão: anais... Belo Horizonte: UFMG, 2014. p. 2434-2452.

BRASIL. Ministério da Saúde. Secretaria Executiva. Subsecretaria de Assuntos Administrativos. **Glossário temático**: ciência e tecnologia em saúde. Brasília, DF, 2013a. 56 p.

BRASIL. Ministério da Saúde. Secretaria Executiva. Subsecretaria de Assuntos Administrativos. Coordenação-Geral de Documentação e Informação. **Glossário temático**: gestão editorial. Brasília, DF, 2013b. 56 p.

BRUCE, T. R.; HILLMANN, D. I. The continuum of metadata quality: defining, expressing, exploiting. In: HILLMANN, D.; WESTBROOKS, E. L. **Metadata in practice**. Chicago, American Library Association, 2004. chap. 15, p. 238-256.

CAMPOS, M. L. M. **Regulamentação e gestão de dados**: leis gerais, licenciamento e certificação. [Rio de Janeiro: UFRJ; Grupo de Engenharia do Conhecimento, 2020]. 56 slides.

CAMPOS, M. L. M.; FREITAS, K. F.; CAMPOS, M. L. de A.; GOMES, H. E.; CAMPOS, L. M. O uso de tesouro como base terminológica para a elaboração de ontologias de domínio: uma experiência com o domínio do folclore e cultura popular. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENCONTRO ENANCIB, 9., 2008, São Paulo. **Diversidade**

cultural e políticas de informação. [São Paulo: ANCIB, 2008]. 15 p. Disponível em: <<http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3015/2141>>. Acesso em: 20 set. 2020.

CAVERLEE, J. Data dictionary. In: LIU, L.; ÖZSU, M. T. (Ed.). **Encyclopedia of database systems.** New York: Springer, 2009. p. 570.

CAVERLEE, J.; MITRA, P.; LAARSGARD, M. Dublin Core. In: LIU, L.; ÖZSU, M. T. (Ed.). **Encyclopedia of database systems.** New York: Springer, 2009. p. 947.

CICLO de vida de dados de pesquisa. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário.** Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 13 jan. 2021.

CIÊNCIA aberta. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário.** Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jan. 2021.

COMPILADOR. In: SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos:** arquivística, biblioteconomia, documentação, informática. Campinas: Editora Átomo, 2003. p. 67.

CONCISÃO. In: HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss da língua portuguesa.** Rio de Janeiro: Objetiva, 2001. p. 787.

CONCISO. In: HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss da língua portuguesa.** Rio de Janeiro: Objetiva, 2001. p. 787.

CONEGLIAN, C. S.; GONÇALVES, P. R. V. A.; SEGUNDO, J. E. S. O profissional da informação na era do Big Data. **Encontros Biblio**, v. 22, n 50, p. 128-143, set./dez. 2017. DOI: 10.5007/1518-2924.2017v22n50p128.

CONTROLLED vocabulary. In: FAO. Agricultural Information Management Standards Portal. **Glossary.** Rome: AIMS, 2020. Disponível em: <<http://aims.fao.org/glossary/controlled-vocabulary>>. Acesso em: 15 jan. 2021.

CONTROLLED vocabulary. In: HYLAND, B.; ATEMEZING, G.; PENDLETON, M.; SRIVASTAVA, B. (Ed.). **Linked data glossary.** [Cambridge, MA]: World Wide Consortium (W3C), 2013. (W3C Working Group Note). Disponível em: <<https://www.w3.org/TR/ld-glossary/#controlled-vocabulary>>. Acesso em: 20 set. 2020.

CREATIVE COMMONS. **Sobre as licenças:** a função das nossas licenças. Disponível em: <<https://br.creativecommons.net/licencas/>>. Acesso em: 5 fev. 2021.

CREATIVE Commons. In: WIKIPÉDIA: a enciclopédia livre. 2020. Disponível em: <https://pt.wikipedia.org/wiki/Creative_Commons>. Acesso em: 23 set. 2020.

CROSAS, M. The Dataverse Network®: an open-source application for sharing, discovering and preserving data. **D-Lib Magazine**, v. 17, n. 1-2, Jan./Feb. 2011. DOI: 10.1045/january2011-crosas.

CROSAS, M.; OKUHN, J.; PASQUIER, T.; SELTZER, M. **Citation ++:** data citation, provenance,

and documentation. 2017. Disponível em: <<https://dataverse.org/presentations/citation-data-citation-provenance-and-documentation>>. Acesso em: 14 jan. 2021.

CSV. In: EUROPEAN UNION OPEN DATA PORTAL. **Glossary**. Disponível em: <<https://data.europa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

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

CURADORIA digital. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jan. 2021.

DADOS abertos. In: BRASIL. Governo do Brasil. **Governo digital**. Disponível em: <<https://www.gov.br/governodigital/pt-br/dados-abertos>>. Acesso em: 12 jul. 2021.

DADOS brutos. In: CENTRO DE INTEGRAÇÃO DE DADOS E CONHECIMENTO PARA SAÚDE. **Glossário de termos**. Salvador, 2021. Disponível em: <<https://cidacs.bahia.fiocruz.br/glossario--de-termos/>>. Acesso em: 26 out. 2021.

DATAONE. **Tutorials on data management**: lesson 3 - data management planning. 2016. 29 slides Disponível em: <https://old.dataone.org/sites/all/documents/education-modules/pptx/L03_Da-taManagementPlanning.pptx>. Acesso em: 2 ago. 2020.

DATASET. In: CAMBRIDGE ADVANCED LEARNER'S DICTIONARY & THESAURUS. [Cambridge, UK]: Cambridge University Press, 2021b. Disponível em: <<https://dictionary.cambridge.org/dictionary/english/dataset>>. Acesso em: 17 ago. 2021.

DATASET. In: CENTRO DE INTEGRAÇÃO DE DADOS E CONHECIMENTO PARA SAÚDE. **Glossário de termos**. [Salvador], 2021a. Disponível em: <<https://cidacs.bahia.fiocruz.br/glossario-de-termos/>>. Acesso em: 26 out. 2021.

DATAVERSE. **About the Project**. Disponível em: <<https://dataverse.org/about>>. Acesso em: 27 ago. 2021.

DATAVERSE guestbooks. In: DATAVERSE PROJECT. **Dataverse management**. [Cambridge: Harvard College], 2016. Disponível em: <<https://guides.dataverse.org/en/4.5/user/dataset-management.html>>. Acesso em: 14 jan. 2021.

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.

DATA.GOV. **Data.gov**. [S. l.], 2021. Disponível em: <<https://www.data.gov/>>. Acesso em: 26 jan. 2021.

DIAS, E.; SILVA, F. M. P. Ensino léxico: a construção da autoria na produção de glossários escolares. **Linguagem em (Re)vista**, v. 11, n. 21, p. 44-69, jan./jun. 2016.

DISCOVERY. In: DATA DOCUMENTATION INITIATIVE. Glossary. [S. l.], DDI Alliance, 2021. Disponível em: <<https://ddialliance.org/resources/ddi-glossary>>. Acesso em: 26 jan. 2021.

DOI. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jan. 2021.

DONOHUE, T. Introduction. In: DONOHUE, T. **DSpace 6.x documentation**. 2018. Disponível em: <<https://wiki.lyrasis.org/display/DSDOC6x/Introduction>>. Acesso em 15 jul. 2021.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 7. ed. São Paulo: Pearson, 2018. 1126 p.

EMBRAPA. **Manual para referência de recursos de informação na Embrapa [recurso eletrônico]**. 3. ed. rev. e ampl. Brasília, DF, 2013. 235 p.

EMBRAPA. Política de Governança de Dados, Informação e Conhecimento da Embrapa. Boletim de **Comunicações Administrativas**, Brasília, DF, v. 45, n. 16, p. 1-19, 5 abr. 2019. Disponível em: <<https://www.embrapa.br/politica-de-governanca-de-dados-informacao-e-conhecimento>>. Acesso em: 13 abr. 2021.

EMBRAPA. **Repositório Acesso Livre à Informação Científica da Embrapa (Alice)**. Brasília, DF, [2020]. Disponível em: <<https://www.alice.cnptia.embrapa.br/alice/>>. Acesso em: 26 set. 2020.

EMBRAPA. **Repositório de Informação Tecnológica da Embrapa (Infoteca-e)**. Brasília, DF, [2021]. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/infoteca.jsp>>. Acesso em: 7 jan. 2021.

e-SCIENCE. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jan. 2021.

ESCOLA NACIONAL DE SAÚDE PÚBLICA SERGIO AROUCA. **Repositório temático - repositório de produção científica**. [Rio de Janeiro, 2020]. Disponível em: <<http://www6.ensp.fiocruz.br/repositorio/>>. Acesso em: 26 set. 2020.

ESPÍNDOLA, P. L.; SALM JUNIOR, J. F.; ROSA, F.; JULIANI, J. P. Governança de dados aplicada à Ciência da Informação: análise de um sistema de dados científicos para a área da saúde. **RDBCI: Revista Digital Biblioteconomia e Ciência da Informação**, v. 16, n. 3, p. 274-298, set./dez. 2018. DOI: 10.20396/rdbci.v16i3.8651080.

ESTADOS UNIDOS. Department of Agriculture. **NAL Agricultural Library**. Disponível em: <<https://agclass.nal.usda.gov/agt.shtml>>. Acesso em: 5 jan. 2021.

ESTADOS UNIDOS. Department of Agriculture. **NAL Agricultural Thesaurus and Glossary home**. Disponível em: <<https://agclass.nal.usda.gov/agt.shtml>>. Acesso em: 18 set. 2020.

EUROPEAN COMMISSION. Directorate-General for Research & Innovation. **H2020 programme: guidelines on FAIR data management in Horizon 2020: version 3.0**. [Brussels, Belgium], 2016. 12 p.

EUROPEAN COMMISSION EXPERT GROUP ON FAIR DATA. **Turning FAIR into reality**: final report and action plan from the European Commission Expert Group on FAIR Data. Brussels, 2018. 76 p. Disponível em: <https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf>. Acesso em: 31 mar. 2021.

FACHIN, J.; ARAUJO, N. C. de; SOUSA, J. C. de. Credibilidade de informações em tempos de COVID-19. **Revista Interamericana de Bibliotecologia**, v. 43, p. eRf3/1-eRf3/11, sept./dic. 2020. DOI: 10.17533/udea.rib.v43n3eRf3.

FAIR DATA MATURITY MODEL WORKING GROUP. **FAIR data maturity model**: specification and guidelines. [S. l.]: RDA, 2020. 45 p. DOI: 10.15497/rda00050.

FAO. **AGROVOC**. Rome: AIMS, 2020. Disponível em: <<http://aims.fao.org/vest-registry/vocabularies/agrovoc>>. Acesso em: 18 set. 2020.

FUNCTIONAL Requirements for Bibliographic Records (FRBR). In: LIBRARIANSHIP STUDIES & INFORMATION TECHNOLOGY. **Glossary of Library & Information Science**. 2021. Disponível em: <<https://www.librarianshipstudies.com/2015/04/glossary-of-library-information-science.html>>. Acesso em: 1 jul. 2021.

FUNDAÇÃO OSWALDO CRUZ. Vice-presidência de Ensino, Informação e Comunicação. Coordenação de Informação e Comunicação. **Relatório final**: piloto de repositório de dados de pesquisa (fase 1). Rio de Janeiro, 2020. 32 p. Disponível em: <https://www.arca.fiocruz.br/bitstream/icit/41414/4/Relat%c3%b3rioFinal_Piloto%20OGP_%20FASE%2001_Rev01.pdf>. Acesso em: 31 ago. 2021.

FURNIEL, A. C. da M.; MENDONÇA, A. P. B.; SILVA, R. M. da. **Recursos educacionais abertos**: conceitos e princípios. Rio de Janeiro: Fiocruz, [2017?]. 30 p.

GESTÃO de dados de pesquisa. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jan. 2021.

GLOBALLY Unique Identifiers (GUIDs) (DOIs, ARKs, etc.). In: SMITHSONIAN LIBRARIES. **Library research tools**. 2021. <<https://library.si.edu/research/guids-help-make-your-data-findable>>. Acesso em: 19 ago. 2021.

GLOSSÁRIO. In: SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação, informática. Campinas: Editora Átomo, 2003. p. 111. GO FAIR. FAIR principles. [Hamburg, 2018]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 16 jul. 2020.

GO FAIR. **FAIR principles**. [Hamburg, 2018]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 16 jul. 2020.

GOMES, H. E. Prefácio. In: CAMPOS, M. L. de A. **Linguagem documentária**: teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001. p. 11-14.

GOVERNANÇA de dados. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016].

Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jul. 2021.

GROTH, P.; MOREAU, L. (Ed.). **PROV-Overview**: an overview of the PROV family of documents. [Cambridge, MA]: W3C, 2013. (W3C Working Group Note). Disponível em: <<https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/>>. Acesso em: 26 jul. 2020.

GUPTA, A. Data provenance. In: LIU, L.; ÖZSU, M. T. (Ed.). **Encyclopedia of database systems**. New York; Springer, 2009. p. 608.

HENNING, P. C.; RIBEIRO, C. J. S.; SANTOS, L. O. B.; SANTOS, P. X. dos. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, v. 25, n. 2, p. 389-412, maio/ago. 2019. DOI: 10.19132/1808-5245252.389-412.

HOPPEN, J.; NETO, E. **4 métricas fundamentais para qualidade de dados em analytics**. Florianópolis: Aquarela Analytics, 2020. Disponível em: <<https://www.aquare.la/4-metricas-fundamen-tais-para-qualidade-de-dados-em-analytics/>>. Acesso em: 22 set. 2020.

IBICT. **Cariniana**: Rede Brasileira de Serviços de Preservação Digital: glossário. Brasília, DF, [2016]. <<http://cariniana.ibict.br/index.php/2016-02-19-18-56-47>>. Acesso em: 17 jul. 2020.

IMPACTA. Data steward. In: IMPACTA. **Gestão de dados**: o que é e porque ter na sua empresa? 1 fev. 2021. Disponível em: <<https://www.impacta.com.br/blog/gestao-dados-porque-ter-na-empresa/>>. Acesso em: 28 ago. 2021.

INTEROPERABILITY. In: EUROPEAN UNION OPEN DATA PORTAL. **Glossary**. Disponível em: <<https://data.europa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional requirements for authority data**: a conceptual model. [München: K. G. Saur], 2013. 54 p. (IFLA working group on Functional Requirements and Numbering of Authority Records (FRANAR): final report).

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 704**: terminology work – principles and methods. Geneva, 2009. 75 p.

JANÉE, G. Digital curation. In: LIU, L.; ÖZSU, M. T. (Ed.). **Encyclopedia of database systems**. New York: Springer, 2009. p. 816.

LICENÇAS Creative Commons. In: WIKIPÉDIA: a enciclopédia livre. 2020. Disponível em: <https://pt.wikipedia.org/wiki/Licen%C3%A7as_Creative_Commons>. Acesso em: 23 set. 2020.

LINGUAGEM artificial. In: SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação, informática. Campinas: Editora Átomo, 2003. p. 146- 147.

LINGUAGEM livre. In: SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação, informática. Campinas: Editora Átomo, 2003. p. 147.

LINGUAGEM natural. In: SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos**: arquivística, biblioteconomia, documentação, informática. Campinas: Editora Átomo, 2003. p. 147.

LIRA, M. A. B. de. **Uma abordagem para enriquecimento semântico de metadados para publicações de dados abertos**. 2014. 96 p. Dissertação (Mestrado em Ciência da Computação) - Centro de Informática, Universidade Federal de Pernambuco, Recife. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/11570>>. Acesso em: 26 jul. 2020.

LIVRO de visitas. In: WIKIPÉDIA: a enciclopédia livre. 2019. Disponível em: <<https://pt.wikipedia.org/wiki/URN>>. Acesso em: 23 set. 2020. Acesso em: 14 jan. 2020.

MARTIGNAGO, D.; MADRID, G.; JULIANI, J. P.; SALM JUNIOR, J. F.; BASTOS, P. R. P. de. Governança de dados aplicada no processo de catalogação. **Revista Brasileira de Biblioteconomia e Documentação**, v. 15, n. 2, p. 81-105, maio/ago. 2019.

MASHUP. In: EUROPEAN UNION OPEN DATA PORTAL. **Glossary**. Disponível em: <<https://data.europa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

MELO, J. O. de S. F. **Metodologia de avaliação de qualidade de dados no contexto do linked data**. 2017. 111 p. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista "Júlio Mesquita Filho", Marília, SP. Disponível em: <<https://repositorio.unesp.br/handle/11449/150870>>. Acesso em: 17 jan. 2021.

METADADO. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 31 ago. 2021.

METADATA standard. In: WIKIPÉDIA: a enciclopédia livre. 2020. Disponível em: <https://en.wikipedia.org/wiki/Metadata_standard>. Acesso em: 23 set. 2020.

MONTES, E. **Métricas da qualidade**. [São Paulo]: Escritório de Projetos, 2020. Disponível em: <<https://escritoriodeprojetos.com.br/metricas-da-qualidade>>. Acesso em: 23 set. 2020.

NASCIMENTO, A. G. do; QUEIROZ, C. F. de; ARAUJO, L. D. de. **A importância dos repositórios e da preservação digital para a pesquisa científica**. [Rio de Janeiro: Fiocruz; Icict, 2019]. 93 slides. Disponível em: <https://www.arca.fiocruz.br/bitstream/icict/38900/4/Reposit%C3%B3rios_Claudete_Luciana_Andrea_UFF_2019.pdf>. Acesso em: 26 set. 2020.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **NCBI home**. Bethesda. Disponível em: <<https://www.ncbi.nlm.nih.gov/>>. Acesso em: 18 set. 2020a.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **Taxonomy**. Bethesda. Disponível em: <<https://www.ncbi.nlm.nih.gov/taxonomy>>. Acesso em: 18 set. 2020b.

NOME. In: HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss da língua portuguesa**. Rio de Janeiro: Objetiva, 2001. p. 2001.

NORTE, M. B. **Glossário de termos técnicos em ciência da informação**: inglês/português. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2010. 50 p. DOI: 10.36311/2011.978-

85-7983-075-4.

ONTOLOGY. In: EUROPEAN UNION OPEN DATA PORTAL. Glossary. [2020a]. Disponível em: <<https://data.europa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

ONTOLOGY. In: FAO. Agricultural Information Management Standards Portal. **Glossary**. Rome: AIMS, 2020b. Disponível em: <<http://aims.fao.org/glossary/ontology>>. Acesso em: 12 jan. 2021.

OPEN DATA SUPPORT. **Introduction to metadata management**: presentation metadata. [Brussels]: European Commission, 2014. 42 slides. Training Module 1.4.

ORCID. **ORCID connecting research and researchers**. Disponível em: <<https://orcid.org/>>. Acesso em: 26 jan. 2021.

PENG, G.; RITCHEY, N. A.; CASEY, K. S.; KEARNS, E. J.; PRIVETTE, J. L.; SAUNDERS, D.; JONES, P.; MAYCOCK, T.; ANSARI, S. Scientific stewardship in the open data and big data era — roles and responsibilities of stewards and other major product stakeholders. **D-Lib Magazine**, v. 22, n. 5/6, May/June 2016. DOI: 10.1045/may2016-peng.

PERGL, R.; HOOFT, R.; SUCHÁNEK, M.; KNAISL, V.; SLIFKA, J. “Data Stewardship Wizard”: a tool bringing together researchers; data stewards, and data experts around data management planning. **Data Science Journal**, v. 18, p. 1-8, 2019. DOI: 10.5334/dsj-2019-059.

PIEROZZI JÚNIOR, I. **Tipologia de dados de pesquisa da Embrapa**. [Campinas: Embrapa Informática Agropecuária, 2020]. 10 p.

PIEROZZI JÚNIOR, I.; TORRES, T. Z.; SOUZA, M. I. F. **Modelo conceitual de tipologia de dados de pesquisa da Embrapa: considerações preliminares**. [Campinas: Embrapa Informática Agropecuária, 2020]. 17 p. (Nota técnica).

PLANO de gestão de dados. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 31 ago. 2021.

PROVENANCE. In: HYLAND, B.; ATEMEZING, G.; PENDLETON, M.; SRIVASTAVA, B. (Ed.). **Linked data glossary**. [Cambridge, MA]: World Wide Consortium (W3C), 2013. (W3C Working Group Note). Disponível em: <<https://www.w3.org/TR/ld-glossary/#provenance>>. Acesso em: 20 set. 2020.

QUALITY assessment. In: NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **The NCBI handbook**. 2. ed. Bethesda, 2019. p. 417. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK143764/pdf/Bookshelf_NBK143764.pdf>. Acesso em: 13 jan. 2021.

QUALITY assurance. In: NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **The NCBI handbook**. 2. ed. Bethesda, 2019. p. 417. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK143764/pdf/Bookshelf_NBK143764.pdf>. Acesso em: 13 jan. 2021.

QUALITY control. In: NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **The NCBI handbook**. 2. ed. Bethesda, 2019. p. 417. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK143764/pdf/Bookshelf_NBK143764.pdf>. Acesso em: 13 jan. 2021.

QUERY. In: NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (Estados Unidos). **The NCBI handbook**. 2. ed. Bethesda, 2019. p. 417. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK143764/pdf/Bookshelf_NBK143764.pdf>. Acesso em: 13 jan. 2021.

REPOSITÓRIO institucional. In: WIKIPÉDIA: a enciclopédia livre. 2020. Disponível em: <https://pt.wikipedia.org/wiki/Reposit%C3%B3rio_institucional>. Acesso em: 26 set. 2020.

RESOURCE. In: EUROPEAN UNION OPEN DATA PORTAL. **Glossary**. Disponível em: <<https://data.europa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

RESOURCE Description and Access. In: LIBRARIANSHIP STUDIES & INFORMATION TECHNOLOGY. **Glossary of Library & Information Science**. 9 May 2021. Disponível em: <<https://www.librarianshipstudies.com/2015/04/glossary-of-library-information-science.html>> . Acesso em: 1 jul. 2021.

REÚSO. In: FUNDAÇÃO OSWALDO CRUZ. **Glossário**. Rio de Janeiro, [2016]. Disponível em: <<https://portal.fiocruz.br/glossario>>. Acesso em: 14 jan. 2021.

RICE, R.; SOUTHALL, J. **The data librarian's handbook**. London: Facet Publishing, 2016. 192 p.

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: 30 ago. 2021.

RUSSO, M. **Fundamentos em Biblioteconomia e Ciência da Informação**. Rio de Janeiro: E-paper Serviços Editoriais, 2010. (Coleção Biblioteconomia e Gestão de Unidades de Informação. Série Didáticos, n. 1).

SAYÃO, L. F.; SALES, L. F. Curadoria de dados de pesquisa em repositórios. In: ENCONTRO DA REDE SUDESTE DE REPOSITÓRIOS INSTITUCIONAIS, 1., 2019, Rio de Janeiro. **Anais...** Rio de Janeiro: Fiocruz/Icict/UFRJ, 2019. 80 slides.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. 90 p.

SCOPING machine-actionable DMPs. 9 July 2018. Disponível em: <<https://blog.dmpool.org/2018/07/09/scoping-machine-actionable-dmps/>>. Acesso em: 15 jul. 2021.

SEGUNDO, W. R.; GUEDES, M. G.; ANDRADE, M.; SHINTAKU, M. A implementação de controle de autoridade em sistemas baseados em Dspace. In: CINFORM: Encontro Nacional de Ensino e

Pesquisa em Informação, 11., 2013, Salvador. **Anais...** Salvador: UFBA. 13 p. Disponível em: <<http://www.cinform2013.ufba.br/modulos/submissao/Upload-164/48903.pdf>>. Acesso em: 1 jul. 2021.

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.

SERPRO. **Glossário LGPD**. Disponível em: <<https://www.serpro.gov.br/lgpd/menu/a-lgpd/glossario-lgpd>>. Acesso em: 22 dez. 2020.

SIMPLE Knowledge Organization System. In: FAO. Agricultural Information Management Standards Portal. **Glossary**. Rome: AIMS, 2020. Disponível em: <<http://aims.fao.org/glossary/simple--knowledge-organization-system>>. Acesso em: 12 jan. 2021.

SINAEPOURFARD, 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).

SMALE, N.; UNSWORTH, K.; DENYER, G.; BARR, D. The history, advocacy and efficacy of data management plans. **BioRxiv**, 443499, 2018. 30 p. DOI: 10.1101/443499.

SISTEMA de gerenciamento de banco de dados. In: WIKIPÉDIA: a enciclopédia livre. 2021. Disponível em: <https://pt.wikipedia.org/wiki/Sistema_de_gerenciamento_de_banco_de_dados>. Acesso em: 12 jul. 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: 13 abr. 2021.

STRASSER, C.; COOK, R.; MICHENER, W.; BUDDEN, A. **Primer on data management: what you always wanted to know**. [Santa Barbara, CA: DataONE, 2011] 11 p. (DataONE best practices primer). Disponível em: <https://old.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf>. Acesso em: 2 ago. 2020.

TAUTZ, C. **Big Data**: uma questão ética em aberto. 2015. Disponível em: <<https://portal.fiocruz.br/noticia/big-data-uma-questao-etica-em-aberto>>. Acesso em: 23 set. 2021.

TAXONOMY. In: FAO. Agricultural Information Management Standards Portal. **Glossary**. Rome: AIMS, 2020. Disponível em: <<http://aims.fao.org/glossary/taxonomy>>. Acesso em: 12 jan. 2021.

TAXONOMY. In: HYLAND, B.; ATEMEZING, G.; PENDLETON, M.; SRIVASTAVA, B. (Ed.). **Linked data glossary**. [Cambridge, MA]: World Wide Consortium (W3C), 2013. (W3C Working Group Note). Disponível em: <<https://www.w3.org/TR/ld-glossary/#taxonomy>>. Acesso em: 20 set. 2020.

TECHOPEDIA. **Access control**. Disponível em: <<https://www.techopedia.com/definition/5831/access-control>>. Acesso em: 20 ago. 2020.

TERM. In: HYLAND, B.; ATEMEZING, G.; PENDLETON, M.; SRIVASTAVA, B. (Ed.). **Linked data glossary**. [Cambridge, MA]: World Wide Consortium (W3C), 2013. (W3C Working Group Note). Disponível em: <<https://www.w3.org/TR/ld-glossary/#term>>. Acesso em: 20 set. 2020.

TERMO específico. In: **TESAURO OC**. [São Paulo: USP/ECA], 2015. Disponível em: <<http://www2.eca.usp.br/mapasconceituais/vocab/index.php?tema=99&/termo-especifico>>. Acesso em: 5 jan. 2021.

TERMO genérico. In: **TESAURO OC**. [São Paulo: USP/ECA], 2015. Disponível em: <<http://www2.eca.usp.br/mapasconceituais/vocab/index.php?tema=79&/termo-genrico>>. Acesso em: 5 jan. 2021.

THESAURUS. In: FAO. Agricultural Information Management Standards Portal. **Glossary**. Rome: AIMS, 2020. Disponível em: <<http://aims.fao.org/glossary/thesaurus>>. Acesso em: 12 jan. 2021.

TRINDADE, A. L. B.; OLIVEIRA, M.; BECKER, G. V. Análise dos atributos para avaliação da qualidade da informação nos ambientes de intranet para apoio à gestão do conhecimento. **READ**, v. 70, n. 3, p. 776-801, set./dez. 2011. Disponível em: <<https://seer.ufrgs.br/read/article/view/38523/24746>>. Acesso em: 22 set. 2020.

UNIVERSITY OF PITTSBURGH. University library System. **Course & subject guides: metadata & discovery @ Pitt metadata standards: what is a metadata standard?** [Pittsburgh], 2020. Disponível em: <<https://pitt.libguides.com/metadatadiscovery/metadata-standards>>. Acesso em: 22 dez. 2020.

URI. In: EUROPEAN UNION OPEN DATA PORTAL. **Glossary**. Disponível em: <<https://data.euro-pa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

URL. In: EUROPEAN UNION OPEN DATA PORTAL. **Glossary**. Disponível em: <<https://data.euro-pa.eu/euodp/en/glossary>>. Acesso em: 20 set. 2020.

URN. In: WIKIPEDIA: a enciclopédia livre. 2020. Disponível em: <<https://pt.wikipedia.org/wiki/URN>>. Acesso em: 23 set. 2020. Acesso em: 14 jan. 2021.

VALENTE, D. L. **Um estudo de metodologia para criação de um depósito de dados**. 2001. 105 p. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Instituto de Informática, Porto Alegre.

VERBETE. In: SANTOS, G. C.; RIBEIRO, C. M. **Acrônimos, siglas e termos técnicos: arquivística, biblioteconomia, documentação, informática**. Campinas: Editora Átomo, 2003. p. 247.

WARD, J. S.; BAKER, A. **Undefined by data: a survey of big data definitions**. 2013. 2 p. Disponível em: <<https://arxiv.org/abs/1309.5821v1.pdf>>. Acesso em: 23 set. 2021.

THE WORLD WIDE WEB CONSORTIUM. **About W3C**. [Cambridge, MA]: World Wide Consortium (W3C), 2020]. Disponível em: <<https://www.w3.org/Consortium/>>. Acesso em: 21 set. 2020.

Embrapa

Agricultura Digital