

# Avaliação de metadados em repositórios de dados de pesquisa sobre biodiversidade

**Carla Beatriz Marques Felipe**

Doutoranda; Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil;  
felipecarla12@gmail.com; <https://orcid.org/0000-0001-5277-9165>

**Raimunda Fernanda dos Santos**

Doutora; Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil;  
fernanda@face.ufrj.br; <https://orcid.org/0000-0002-7750-3269>

**Resumo:** Este artigo disserta sobre repositórios de dados, em particular a respeito de metadados para a descrição de conjuntos de dados nesses sistemas. Aborda a iniciativa FAIR e o FAIR Data Point como tecnologia que auxilia a recuperação da informação em repositórios de dados. Objetiva avaliar, com base nos metadados FAIR Data Point, a qualidade dos metadados em Repositórios da área de Biodiversidade. Analisa como ocorre a descrição de conjuntos de dados no Repositório de Dados de Estudos Ecológicos e compara os metadados propostos pelo FAIR Data Point, sistema que especifica um conjunto de metadados padronizados a serem implementados em um Repositório. A abordagem do estudo é qualitativa e integra as pesquisas bibliográfica, exploratória, descritiva e documental. Conclui que alguns metadados identificados no Repositório de Dados de Estudos Ecológicos são comuns aos do FAIR Data Point, e outros não apresentam similaridade. Ainda, propõe melhorias na descrição dos conjuntos de dados, tornando sua descrição singular e precisa, sobretudo quanto à representação temática dos dados de pesquisa.

**Palavras-chave:** Repositórios de dados; FAIR; FAIR Data Point; Metadados; Dados de pesquisa

## 1 Introdução

O uso intensivo de máquinas para a geração de dados dentro da ciência pode ser caracterizado como *e-science*. Nesse contexto, os *softwares* são operados para o processamento de dados em ampla escala, os quais passam a ser protagonistas no âmbito científico, visto que, muitas vezes, foram esquecidos nos computadores de seus respectivos pesquisadores.

Segundo Curty (2017), os dados de pesquisa se configuram como matéria-prima fundamental para a ciência e são necessários para novos ciclos de criação e atualização do conhecimento científico, viabilizando, assim, inovação científica e tecnológica. Nesse cenário, surgem formas específicas de publicação de dados que facilitam o compartilhamento destes e, conseqüentemente, o reuso. Logo, essas formas são conhecidas como *Data publishing* ou publicação de dados, que devem ser realizadas nos periódicos de dados, artigos de dados e repositórios de dados de pesquisa.

Por conseguinte, nos repositórios de dados de pesquisa estão incluídos os metadados que possibilitam a representação de dados com vistas a promover sua recuperação. Desse modo, Coyle (2005) ressalta a importância do papel substituto dos metadados tencionando tornar os recursos digitais de fácil identificação, navegação e compartilhamento. Os metadados representam os recursos informacionais e nos repositórios de dados de pesquisa também precisam ser bem estruturados, de modo que possam atender à descrição, referência, significação, uso e reuso de dados em pesquisas científicas.

Assim, surgem iniciativas que visam auxiliar a gestão de dados, dentre elas a iniciativa FAIR, que significa *Findable* (encontrável), *Accessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reutilizável). Essa iniciativa apresenta requisitos que tendem a promover melhores práticas de gestão de dados mediante ações e tecnologias voltadas a essa gestão.

O FAIR *Data Point* consiste em um *software* que trabalha os metadados, em especial para os repositórios de dados de pesquisa, com o emprego de dados relativos à proveniência e à iniciativa FAIR. Objetiva tornar os conjuntos de dados FAIR acessíveis através do uso de metadados legíveis para máquinas e humanos. Ainda, denota ser um instrumento valioso para a gestão e a representação de dados.

Isto posto, são levantados os seguintes questionamentos: os metadados dos repositórios são compatíveis com os da proposta do FAIR *Data Point*? Como ocorre a representação dos conjuntos de dados em repositórios de dados de pesquisa como o de Biodiversidade? Quais metadados são usados tanto no FAIR quanto nesses repositórios?

O objetivo geral da pesquisa é avaliar, com base nos metadados FAIR *Data Point*, a qualidade dos metadados da área de Biodiversidade. E como objetivos específicos: identificar quais metadados estão sendo usados nos repositórios de dados de pesquisa; descrever quais são os metadados que se encontram tanto no FAIR *Data Point* como nos repositórios de dados de pesquisa; comparar os metadados do FAIR *Data Point* com os metadados recuperados nos repositórios.

A relevância deste estudo dá-se em compreender como estão sendo efetuadas as práticas de representação de dados nos repositórios e a proposta de representação do FAIR *Data Point* para os repositórios de dados de pesquisa, como, por exemplo, o de Biodiversidade. Os dados coletados foram retirados do Repositório de Dados de Estudos Ecológicos (PPBIO), o qual é voltado para a disseminação de dados de pesquisa da Biodiversidade brasileira. Assim, selecionou-se a área aludida, haja vista ela integrar o *corpus* da pesquisa doutoral de uma das pesquisadoras.

Para um melhor entendimento desta investigação, seguem, nos próximos tópicos, aspectos concernentes aos repositórios de dados de pesquisa e à iniciativa FAIR.

## 2 Repositório de dados de pesquisa

Na esfera científica desponta um novo meio de fazer ciência, cuja produção de dados requer organização para fins de análise, interpretação e inovação científica. Esse novo meio é conhecido como *e-science*, entendida como a grande produção de dados e extenso uso de máquinas na produção destes. Para Gray (2009), a *e-science* é um dos quatro paradigmas que a ciência pode ser classificada, corresponde ao modo como esta foi e é efetuada segundo o tempo e as tecnologias.

O primeiro paradigma concerne à ciência experimental, o segundo relativo à ciência teórica, o terceiro à ciência computacional, com a introdução de computadores no modo de fazer ciência, e o quarto configura-se como o paradigma atual, que se refere ao uso intenso de máquinas e *softwares* na

produção de dados em larga escala, conhecida como *e-science* – ciência guiada pelos dados.

Sendo assim, os dados passam a ser o destaque no fazer científico, sobretudo em ambientes colaborativos, em que o compartilhamento é essencial para o reuso. Dessa maneira, para que esta ação suceda faz-se necessários que tais dados sejam representados e partilhados, possibilitando ao pesquisador que reutilizá-lo-ás poupe tempo e recursos para a execução de uma nova coleta de dados.

À vista disso, a tarefa de implementar estruturas e ferramentas tecnológicas torna-se imprescindível para a divulgação e o compartilhamento de dados, posto que um dado tem estrutura diferente de um objeto informacional, como livro e artigo, por exemplo. Assim, surgem estruturas que facilitam a distribuição e a divulgação. O *Data publishing* ou a publicação de dados, por exemplo, viabilizam a representação de dados mediante uma armação específica com elementos que caracterizam sua descrição (metadados), oportunizando a divisão e a reutilização por outros pesquisadores.

Segundo Curty (2017), as *data publishing* são desdobradas em *data papers* (artigo de dados), *data journal* (periódico de dados) e repositórios de dados – os quais são considerados fundamentais no compartilhamento dos dados, já que são elaborados especificamente para essa função. Os repositórios de dados de pesquisa, por sua vez, têm acentuado ao longo do tempo. Isso fica perceptível ao se pesquisar no Re3Data, ambiente onde é possível localizar mais de 84 (oitenta e quatro) repositórios no contexto da Biodiversidade e de diferentes disciplinas acadêmicas. No Brasil, é possível recuperar 12 (doze) iniciativas no âmbito do R3 Data. Esses dados estão ligados às áreas da saúde, humanidades, biodiversidade, clima e periódicos multidisciplinares.

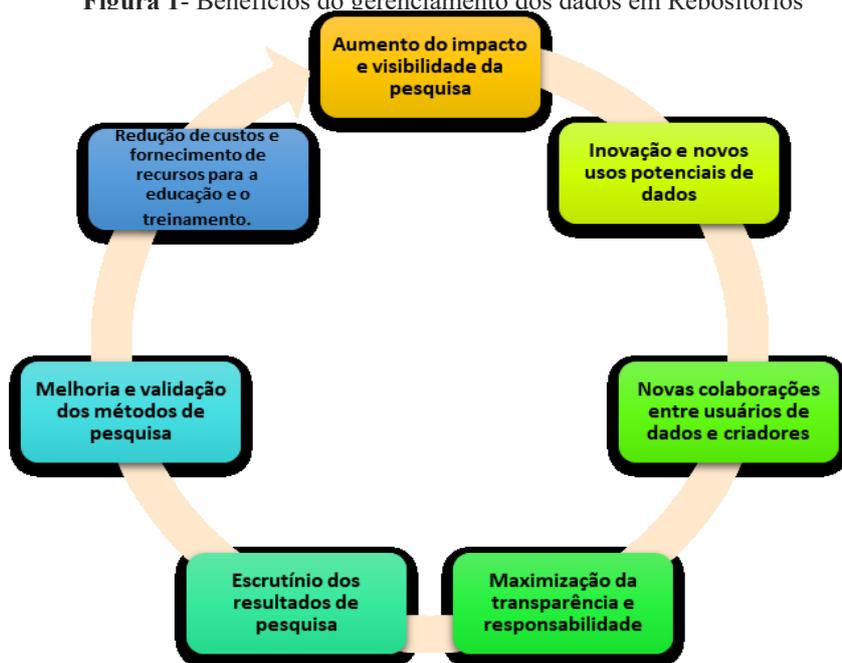
Esses repositórios são armações que permitem acesso ininterrupto aos dados de forma organizada através de metadados e identificadores que facilitam o compartilhamento, a preservação e o reuso de dados. Tais sistemas podem estar vinculados às instituições, aos temas específicos, às comunidades disciplinares ou apenas fruto de projetos de pesquisa. Ainda, conforme Fenner *et al.* (2019, tradução nossa), desempenham importante papel na citação de dados

de pesquisa, já que fornecem serviços de gestão e recuperação com identificadores e metadados exclusivos e vitais para a sua reutilização e citação.

Nesse contexto, é possível afirmar que o repositório de dados é fundamental para o compartilhamento de dados, porquanto sua estrutura é ancorada nessa perspectiva. Logo, a descrição e a recuperação de dados armazenados devem ser operações essenciais a serem efetuadas nesses ambientes, porque não basta apenas fornecer um sistema, é preciso que ocorra a gestão de dados para fins de uso e reuso. Ademais, a gestão desses ambientes garante a preservação de dados a longo prazo e o acesso aos que queiram reutilizar de maneira eficiente.

De acordo com Córdula e Araújo (2019, p. 197), há alguns benefícios ligados ao gerenciamento de dados, conforme pode-se observar na ilustração a seguir:

**Figura 1-** Benefícios do gerenciamento dos dados em Repositórios



Fonte: Adaptado de Córdula e Araújo (2019, p. 197).

A partir da ilustração nota-se a quantidade significativa de benefícios tocantes à implementação de um plano de gerenciamento de dados. Esses benefícios impactam no desenvolvimento científico e tecnológico, de modo a promover a redução de custos (financeiros, de coleta, entre outros) e a transparência com os métodos de pesquisa e os resultados alcançados.

No âmbito de dados de pesquisa, além das práticas de representação de dados há iniciativas que objetivam orientar as atividades a serem efetivadas em prol da gestão de dados. Uma delas é a FAIR, inicialmente desenvolvida no exterior e com divulgação de princípios que auxiliam pesquisadores. A seguir discorrer-se-á acerca dos aspectos relativos a essa iniciativa.

### 3 Iniciativa FAIR

A iniciativa FAIR surge com o propósito de contribuir para a gestão de dados. Inicialmente, em um evento sobre ciências da vida, em 2014, foram lançados princípios que levam o mesmo nome da iniciativa (DIAS *et al.*, 2019). FAIR significa *Findable* (encontrável), *Acessible* (acessível), *Interoperable* (interoperável) e *Reusable* (reutilizável), logo os dados devem estar acessíveis para aqueles que quiserem (re) usar, ser legíveis por máquina, fáceis de encontrar e reutilizáveis. Esses quatro princípios são estendidos em 16 (dezesseis) subprincípios, que auxiliam para o seu entendimento. A figura, a seguir, apresenta de maneira detalhada esses princípios:

**Figura 2 - Princípios FAIR**



Fonte: Fiocruz (2020).

Os princípios elencados anteriormente não se configuram como regras, pois correspondem às orientações que podem ser seguidas ou não pelos

repositórios. Têm como objetivo principal tornar os dados de pesquisa mais acessíveis, facilitando, assim, o compartilhamento e o reuso. O foco não está apenas nos dados, mas também nos metadados. A maioria deles contempla orientações sobre o metadado, sendo entendido como elemento informativo necessário para o entendimento dos dados e do contexto.

Para Arakaki e Arakaki (2020), o conceito de metadados sinaliza uma informação estruturada para fins de identificação, recuperação, seleção, uso, acesso e gerenciamento de uma ampla variedade de recursos. Posto isto, é possível inferir que eles fornecem diferentes pontos de acesso para a descoberta, o uso e o reuso de dados em múltiplas ocasiões.

Assim, intentando divulgar os princípios FAIR e os serviços FAIR, em 2017 foi criada a iniciativa GO FAIR, que “oferece um ecossistema aberto e inclusivo para indivíduos, instituições e organizações que trabalham juntas por meio de Redes de Implementação” (GO FAIR, 2021). Essa iniciativa é fundamentada em três pilares: *Go Change*, *Go Train* e *Go Build*. O *Go Change* trata das políticas e incentivos para efetuar o FAIR. O *Go Train* versa sobre os treinamentos e o desenvolvimento das práticas FAIR. E o *Go Build* é voltado para as infraestruturas de implementação dos princípios FAIR. Essas ações visam garantir que estes princípios sejam aplicados na gestão de dados e possam, de fato, contribuir para a interoperabilidade e reuso efetivo.

Quanto às tecnologias que facilitam a elaboração dos princípios FAIR, tem-se o FAIR *Data Point*, que surge dentro da iniciativa FAIR intentando facilitar a recuperação da informação mediante um *software*. Este abrange o subsistema de entrada de metadados para os proprietários que querem expor seus dados, bem como metadados que contribuem para a recuperação da informação por parte dos usuários.

O Fair *Data Point* apresenta três objetivos: possibilitar que proprietários/criadores/editores registrem os metadados de seus objetos digitais em consonância com os princípios de dados FAIR; permitir que consumidores/usuários descubram informações sobre objetos digitais nos quais estão interessados, e; promover a interação entre humanos e agentes de *software* (METADATA, 2020).

Ainda segundo a fonte aludida, os componentes do FAIR *Data Point* são: Provedor de Metadados, Controle de Acesso, Esquemas de Metadados e Armazenamento de Metadados RDF. Nesse contexto, o sistema auxilia na instalação e atualização dos metadados para os repositórios. Assim, pode tornar-se uma fonte rica para auxiliar na recuperação, no acesso e no (re) uso de dados, em especial quando sua proposta é tornar os metadados FAIR, assim como a maioria dos princípios FAIR.

#### **4 Procedimentos metodológicos**

A pesquisa é caracterizada como bibliográfica, exploratória, descritiva e documental com abordagem qualitativa. Tem como objetivo geral avaliar, com base nos metadados FAIR *Data Point*, a qualidade dos metadados da área de Biodiversidade.

Segundo Gil (2002), a pesquisa bibliográfica é desenvolvida mediante fontes bibliográficas, sendo elas classificadas em livros, publicações periódicas e materiais informacionais diversos. A principal vantagem desse método é permitir que o pesquisador tenha maior cobertura de uma quantia de fenômenos que poderia pesquisar diretamente. Assim, no presente estudo, foram realizadas pesquisas em bases de dados, como BRAPCI, PERI, Portal de Periódicos CAPES e Biblioteca Digital Brasileira de Teses e Dissertações tentando recuperar e analisar produções científicas relativas aos temas: Repositórios de Dados, metadados, iniciativa e princípios FAIR.

No tocante à pesquisa exploratória, Prodanov e Freitas (2013, p. 51-52) ressaltam que esse tipo de pesquisa “tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar [...]”. Nesse contexto, a investigação permitiu um melhor aprofundamento sobre como ocorre a representação da informação em Biodiversidade através de metadados.

Quanto à pesquisa descritiva, segundo Gil (2002), objetiva descrever as características específicas de determinada população, possibilitando estabelecer relações entre variáveis. Neste estudo, a pesquisa referenciada possibilitou descrever como decorre a representação da informação em um Repositório de

Dados por meio de metadados e comparar com as sugestões do FAIR *Data Point*.

Ainda, sucedeu-se a pesquisa documental. Conforme Gil (2002), nesse tipo de pesquisa as fontes são mais diversificadas, tendo em vista que ainda não receberam um tratamento analítico ou podem ser reinterpretadas em consonância com os objetivos da investigação. Ademais, ela foi empregada para fins de coleta dos dados. Assim, preliminarmente efetuou-se uma busca no Re3Data para a recuperação do Repositório de Dados de Estudos Ecológicos (PPBIO), além de serem consultadas e analisadas fontes, como o *DataOne* e o FAIR *Data Point*. O PPBIO foi selecionado para a prática das análises, haja vista que ser o único repositório disponível no Re3Data que trata acerca da Biodiversidade no Brasil.

À vista disso, o objetivo do PPBIO consiste em:

intensificar estudos sobre biodiversidade no Brasil, descentralizar a produção científica dos centros desenvolvidos academicamente, integrar atividades de pesquisa e divulgar os resultados para diferentes finalidades, como gestão ambiental e educação (PPBIO, 2012).

Nesse contexto, o PPBIO desenvolve várias iniciativas que facilitam a disseminação da informação em biodiversidade, incluindo sítios de coleta, repositório de dados, banco de imagens e divulgação de produtos, como teses e dissertações. A iniciativa escolhida para esta pesquisa foi o repositório de dados.

Assim, algumas áreas do conhecimento apresentam programas para fornecer acesso aos dados de repositórios similares, e no contexto da Biodiversidade há o *DataOne* (cujo PPBIO está inserido) – programa fundado em 2009 que visa prestar suporte à pesquisa e à recuperação de dados ambientais e sobre a terra, além de fornecer recursos e materiais de apoio aos pesquisadores dessa área do conhecimento (DATAONE, 2021a). Como o PPBIO integra o *DataOne*, seus conjuntos de dados são recuperados por meio dessa última fonte, ou seja, quando clicamos em sua página para acessar os dados, o usuário é direcionado para o sítio do *DataOne*. Entretanto, o primeiro passo para a coleta dos dados foi observar os metadados sugeridos pelo PPBIO.

A avaliação efetuada neste estudo partiu do segundo objetivo do FAIR *Data Point*, que diz respeito à disponibilização da informação sobre dados contidos nos repositórios por meio de metadados. A avaliação foi elaborada a partir da camada de metadados que alude acerca da descrição do conjunto de dados disponíveis nos repositórios. É importante frisar que foram analisados apenas os metadados descritivos, sem considerar os voltados para a estrutura, a preservação, o acesso e o uso dos dados.

Os metadados descritivos são aqueles que apresentam as características e os contextos intelectuais dos recursos para fins de identificação, recuperação, seleção, aquisição, contextualização e compreensão (ARAKAKI; ARAKAKI, 2020). Sendo assim, os metadados sugeridos pelo Fair *Data Point* para fins de análise e discussão dos dados desta pesquisa seguem descritos no quadro abaixo:

**Quadro 1-** Metadados de conjunto de dados

METADADOS	CARÁTER	DESCRIÇÃO
<i>Type</i>	Obrigatório	Necessário ser do tipo deat:Dataset
<i>Title</i>	Obrigatório	Nome do conjunto de dados com a <i>tag</i> de idioma
<i>Publisher</i>	Obrigatório	Organização (ões) ou Pessoa(s) responsável(is) pelo conjunto de dados
<i>hasVersion</i>	Obrigatório	Versão do conjunto de dados
<i>description</i>	Opcional	Descrição do conjunto de dados com a <i>tag</i> de idioma
<i>conformsTo</i>	Opcional	A especificação do esquema de metadados do conjunto de dados (ShEx)
<i>metadataIssued</i>	Opcional	Data de criação da entrada do conjunto de dados
<i>metadataModified</i>	Opcional	Data da última modificação da entrada do conjunto de dados
<i>language</i>	Opcional	Idioma do conjunto de dados
<i>license</i>	Opcional	Licença associada ao conjunto de dados
<i>rights</i>	Opcional	Direitos sobre o conjunto de dados
<i>accessRights</i>	Opcional	Descrição dos direitos de acesso
<i>isPartOf</i>	Obrigatório	Relação com os metadados pai.
<i>label</i>	Opcional	Nome do conjunto de dados com a <i>tag</i> de idioma
<i>keyword</i>	Opcional	Palavras-chave relacionadas ao conjunto de dados com a <i>tag</i> de idioma
<i>metadataIdentifier</i>	Obrigatório	Identificador de metadado
<i>MetadataIssued</i>		Data de criação da entrada de metadados
<i>MetadataModified</i>		Data da última modificação da entrada de metadados
<i>distribution</i>	Obrigatório	Lista de URLs de distribuição
<i>theme</i>	Obrigatório	Lista de termos que descrevem o assunto do conjunto de dados
<i>ContactPoint</i>	Opcional	Organização responsável e mantenedora do metadado
<i>landingPage</i>	Opcional	Página inicial do conjunto de dados

Fonte: Adaptado de FAIR *Data Point* (2020).

Foram selecionados três conjuntos de dados que se configuram como parte do PPBIO dentro do *DataOne* para a observação da descrição e da

comparação com os metadados propostos pelo FAIR *Data Point*. O primeiro passo foi observar os metadados e perceber como estavam organizados o DataOne – PPBIO. Nessa etapa, apreendeu-se que a descrição ocorre em categorias e os metadados estão incorporados a estas.

Na segunda etapa, foram observados quais metadados sugeridos pelo *Fair Data Point* estão presentes no PPBIO. Assim, esse sistema sugere em sua estrutura um conjunto de metadados obrigatórios e opcionais. Nessa dinâmica, a análise realizada nesta pesquisa também considerou essas características dos metadados.

Diante disso, em cada conjunto de dados selecionado observou-se se os metadados obrigatórios e opcionais sugeridos pelo *FAIR data point* estavam presentes nos conjuntos de dados, com mesmo nome ou não. Caso o metadado não aparecesse nomeado tal como é apresentado na sugestão do *Fair Data Point* considerou-se a possibilidade de existir metadados com sinônimos, mas que exercesse a mesma função na representação dos dados. A seguir apresentar-se-á os resultados da pesquisa segundo os parâmetros metodológicos definidos e descritos anteriormente.

## 5 Análise e discussão dos resultados

Para alcançar o primeiro objetivo específico – identificar quais metadados estão sendo utilizados nos repositórios de dados de pesquisa – foram observados os metadados aplicados no repositório PPBIO. Assim, descrição dos conjuntos de dados no *DataOne* ocorre em diferentes níveis, cujos metadados principais se encaixam em categorias genéricas (comuns em todos os conjuntos de dados), e os metadados específicos variam segundo as tipologias de dados a serem representados.

Dessa maneira, na parte geral da descrição de dados, dentro do *DataOne*, foram percebidos os seguintes metadados: “*Identifier*” (Identificador), “*Abstract*” (Resumo) e “*Keywors*” (Palavras-chave). Na tabela que contém a descrição individual de cada dado que forma o conjunto, a representação é realizada separadamente.

Os metadados recuperados foram: “*Entity Name*” (Nome da Entidade), “*Object Name*” (Nome do objeto), “*Online Distribution Info*” (Informação de Distribuição online), “*Size*” (Tamanho), “*Text Format*” (Formato de Texto) e “*Number Of Records*” (Número de registro). Ainda, identificamos “*Attribute Information*” (Informação de Atributo), que contém outros metadados inerentes à coleta dos dados e à espécie, tais como: “*Name*” (nome), relativo à pessoa que coletou os dados, “*Definition*” (definição), “*Method*” (Método), entre outros. Essa descrição vai depender do dado a ser representado.

Outra categoria em comum a todos os conjuntos de dados analisados diz respeito à “*People and Associated Parties*” (Pessoas e Partes Associadas), aos produtores, à organização e às pessoas que colaboraram com a pesquisa. Desse modo, foram encontrados os respectivos metadados: “*Data Set Creators*” (Criadores de conjuntos de dados), que se desdobram em “*Individual*” (individual), “*Organization*” (Organização), “*Addres*” (Endereço), “*Position*” (Posição), “*Email Address*” (Endereço de e-mail). Ademais, Contatos do conjunto de dados que se desdobram em Individual e Organização, e por último Parte Associada que se divide em “*Individual*” (individual), “*Organization*” (Organização) e “*Addres*” (Endereço).

Há, também, um campo para “Outra Entidade”. Diante do exposto, entende-se que esse campo não cabe em qualquer outra categoria entre as citadas. Diante disso, os metadados vinculados a esse campo foram: “*Entity Name*” (Nome de Entidade), “*Description*” (Descrição), “*Physical Structure Description*” (Descrição da estrutura física), “*Object Name*” (Nome do Objeto), “*Size*” (Tamanho), “*Externally Defined Format*” (Formato definido externamente) e “*Online Distribution Info*” (Informações de distribuição online).

Foram identificados, do mesmo modo, no *DataOne* fora das categorias gerais aludidas, e não sendo comum a todos os conjuntos, os seguintes metadados: “*Taxonomic Range*” (Alcance Taxonômico) e “*Methods & Sampling*” (Métodos e Amostragem), “*Data Set Usage Rights*” (Direitos de uso do conjunto de dados), “*Geographic Region*” (Região geográfica) e “*Temporal*”

Coverage” (cobertura temporal). Para mais das classes comum a todos os conjuntos, tem-se o metadado “Access Control” (Controle de acesso).

Na tabela, abaixo, são apresentados os metadados do repositório que apresentaram similaridade concernentes às orientações do *Fair Data Point*:

**Quadro 2 - Comparação entre metadados**

<b>Metadados Fair Data Point</b>	<b>Caráter dos metadados Fair Data Point</b>	<b>Metadados do Repositório PPBIO</b>
<i>Type</i>	Obrigatório	Não identificado
<i>Title</i>	Obrigatório	Identificado
<i>Publisher</i>	Obrigatório	Identificado
<i>hasVersion</i>	Obrigatório	Não contém
<i>description</i>	Opcional	Identificado
<i>conformsTo</i>	Opcional	Não identificado
<i>issued</i>	Opcional	Não contém
<i>modified</i>	Opcional	Não contém
<i>language</i>	Opcional	Não contém
<i>license</i>	Opcional	Não contém
<i>rights</i>	Opcional	Identificado
<i>Label</i>	Opcional	Identificado
<i>accessRights</i>	Obrigatório	Não identificado
<i>isPartOf</i>	Opcional	Não identificado
<i>keyword</i>	Opcional	Identificado
<i>metadataIdentifier</i>	Obrigatório	Identificado
<i>MetadataIssued</i>		Não identificado
<i>MetadataModified</i>		Não identificado
<i>distribution</i>	Obrigatório	Identificado
<i>Theme</i>	Obrigatório	Não identificado
<i>ContactPoint</i>	Opcional	Identificado
<i>landingPage</i>	Opcional	Não contém

Fonte: Dados da pesquisa.

Na representação acima pode-se perceber os metadados que foram ou não foram encontrados no PPBio. Assim, exprime-se os seguintes entendimentos: identificado significa que o metadado que é proposto pelo *Fair Data Point* é encontrado no PPBIO, mesmo que tenha uma nomenclatura diferente; e não identificado, que não foi possível constatar sua presença, ou seja, que não está sendo fornecido.

Apesar de não terem sido identificados todos os metadados no PPBIO, conforme sugerido pelo *FAIR Data Point*, a descrição dos conjuntos de dados do repositório em questão contempla características relevantes de identificação dos dados como, por exemplo, nome do conjunto de dados, palavras-chave, pessoas responsáveis pelo conjunto de dados, direitos de acesso e distribuição.

Sendo assim, observou-se que 40% dos metadados identificados no PPBIO atendem os metadados sugeridos pelo Fair *Data Point*.

No repositório, alguns metadados foram substituídos por uma referência que apresenta o autor, o título do conjunto de dados e o identificador. O metadado identificador aparece tanto na referência quanto na descrição geral do conjunto de dados.

Alguns metadados recuperados no PPBIO são compatíveis com os do FAIR *Data Point*, apesar de aparecerem nomeados de outra forma, porque possuem o mesmo significado e objetivo. A saber: “*Title*” (título), “*Publisher*” (Editor), “*Description*” (descrição), “*rights*” (direito), “*Label*” (rótulo), “*Keywords*” (palavras-chave), “*MetadataIdentifier*” (identificador de metadado), “*Distribution*” (distribuição) e “*contact point*” (ponto de contato).

No tocante aos metadados sugeridos pelo Fair *Data Point* e não encontrados no PPBIO, como “idioma” e “licença”, sugere-se que sejam contemplados no repositório PPBIO para fins de representação e identificação da licença concedida e do *status* dos dados em relação à cadeia de (re) uso. Acredita-se que o metadado “idioma” é importante para a descrição de dados, considerando que embora o repositório seja nacional, os dados a serem armazenados podem ser provenientes de pesquisas cujo *corpus* ou universo foi internacional.

O metadado “tema”, por sua vez, também não foi identificado na representação dos conjuntos de dados analisados. Infere-se que pelo fato do repositório ser temático os criadores do repositório não acharam pertinente sua inserção, contudo, assim como qualquer área, a Biodiversidade (enquanto domínio de conhecimento) contempla diferentes conceitos e termos que carecem de serem especificados no metadado “assunto”.

Os metadados “versão”, “emitido” e “modificado” foram substituídos no repositório pelo metadado “*Temporal Coverage*” (cobertura temporal), que demonstra todo o período em que os dados foram coletados e eventuais alterações realizadas durante esse período.

À vista disso, é possível afirmar que o FAIR *Data Point* está no caminho certo para apoiar a implementação dos princípios FAIR tocante aos metadados.

Apesar de no contexto pesquisado o repositório ser mais específico que os metadados do FAIR *Data Point*, o PPBIO apresenta muitos metadados importantes em sua composição para a descrição de dados, fazendo com que o usuário não se perca ao realizar a consulta.

Por pertencer ao *DataOne*, o PPBIO passa por uma avaliação FAIR com relação aos seus metadados por parte dos usuários objetivando gerar um “Relatório de Avaliação de Metadados”. Para tanto, é solicitado ao usuário que após cadastrar os metadados seja feita uma avaliação quanto aos quatro princípios FAIR. Até a data da realização desta pesquisa, o resultado dessa avaliação mostrou-se da seguinte forma:

**Figura 3** - Relatório de avaliação de metadados



Fonte: DataONE (2021b).

Conforme a ilustração, esse relatório foi gerado com base na avaliação de todos os usuários que já cadastraram metadados e incluíram dados de suas pesquisas no sistema. É possível observar que nos quesitos “localizável” e “acessível” os metadados estão bem avaliados, porém nos atributos “interoperável” e “reutilizável” depreende-se necessitar de melhorias pautadas nesses princípios.

## 6 Considerações finais

Os dados de pesquisa vêm ganhando, cada vez mais, espaço no âmbito da comunicação científica. Em vista disso, surgem novas formas de publicação de

dados (*Data publishing*) que facilitam a preservação e o compartilhamento, visto apresentarem características específicas para tal. Nesse contexto, sucedem os repositórios de dados, cujas ferramentas contribuem para preservar a longo prazo os dados de pesquisa, fornecendo informações como metadados e identificadores de dados com vistas ao (re) uso.

A avaliação dos metadados configura-se como uma atividade importante, visto serem elementos fundamentais que condicionam a recuperação, o acesso, o uso e o reuso dos dados. Assim, o FAIR *DataPoint* mostrou ser uma valiosa ferramenta de avaliação dos metadados, posto que seu papel é garantir que os metadados sejam compatíveis com a linguagem e com a atividade realizada pelos usuários.

Na conjuntura pesquisada, os metadados do PPBIO favorecem o entendimento, a representação e a recuperação dos conjuntos de dados disponíveis. Alguns são comuns aos do FAIR *Data Point* e outros, não. Acredita-se que o repositório surgiu primeiro que o FAIR *Data Point* e o PPBIO pode implementar melhorias no tocante à descrição dos conjuntos de dados, ascendendo o número de metadados para a descrição detalhada e precisa dos dados no sistema. Após a elaboração da pesquisa, pode-se afirmar que os dados do PPBIO estão sendo descritos em consonância com os princípios FAIR, porém merecem atenção, em especial no quesito relativo ao metadado “tema”.

Como esse metadado não apareceu em qualquer dos conjuntos analisados, considera-se essencial sua implementação, pois mesmo que o repositório seja temático esse sistema apresenta dados dos mais diversos temas ligados à Biodiversidade (tais como flora, fauna, espécies em extinção, etc.), e sua especificação contribuirá para a precisão na busca e na recuperação dos dados.

Para mais, recomenda-se a criação e a implementação de um sistema híbrido de organização do conhecimento no PPBIO a ser elaborado com base na linguagem empregada pelos usuários na descrição dos dados (Folksonomia) e com o suporte dos formalismos e métodos usados na construção de instrumentos de controle terminológicos. Esse instrumento objetiva auxiliar na representação, na busca e na recuperação semântica dos dados com vistas a promover o uso e o

reuso dos dados de pesquisa na comunidade da Biodiversidade e áreas correlatas.

## Referências

ARAKAKI, Ana Carolina Simionato; ARAKAKI, Felipe Augusto. Dados e metadados: conceitos e relações. **Ci. Inf.**, Brasília, DF, v. 49, n. 3, p. 34-45, set./dez. 2020.

CÓRDULA, Flavio Ribeiro; ARAÚJO, Wagner Junqueira. O compartilhamento de dados científicos na era do e-science. *In*: DIAS, Guilherme Ataíde; OLIVEIRA, Bernadina Maria Juvenal de (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019.

COYLE, Karem. Understanding Metadata and Its Purpose. **The Journal of Academic Librarianship**, [s. l.], v. 3, n. 2, p. 160-63, 2005.

CURTY, Renata Gonçalves. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais [...]**. Marília: UNESP, 2017. p. 1-20.

DATAONE. **About DataOne**. 2021a. Disponível em: <https://www.dataone.org/about/>. Acesso em: 20 out. 2021.

DATAONE. **Metadata Assessment Report**. 2021b. Disponível em: <https://www.dataone.org/fair/>. Acesso em: 7 set. 2021.

DIAS, Guilherme Ataíde et al. Os princípios FAIR: viabilizando o reuso de dados científicos. *In*: DIAS, Guilherme Ataíde et al; OLIVEIRA, Bernadina Maria Juvenal de. **Dados científicos: perspectivas e desafios**. João Pessoa: Editora UFPB, 2019.

FAIR DATA POINT. **About**. 2020. Disponível em: <https://www.fairdatapoint.org/about/>. Acesso em: 20 out. 2021.

FENNER, Martin *et al.* A data citation roadmap for scholarly data repositories. **Scientific Data**, [s. l.], 2019, p. 1-9.

FIOCRUZ. FAIR: dos princípios à prática. 2020. Disponível em: <https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula4.html>. Acesso em: 13 dez. 2020.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOFAIR. **GO FAIR Initiative**. 2021. Disponível em: <https://www.go-fair.org/go-fair-initiative/>. Acesso em: 30 out. 2021.

GRAY, Jim. Jim gray on escience: a transformed scientific method. *In*: HEY, Tony; Tansley, Stewart; TOLLE, Kristin (ed.). **The Fourth Data-Intensive Scientific Discovery Paradigm**. Washington: Microsoft research, 2009. p. 17-31.

**METADATA** model specification. 2020. Disponível em <https://specs.fairdatapoint.org/>: Acesso em: 5 out. 2021.

PPBIO. **Repositório**. 2012. Disponível em: <https://ppbio.inpa.gov.br/repositorio/dados>. Acesso em: 30 out. 2021.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013. E-book. 276 p.

## Evaluation of metadata in research data repositories about biodiversity

**Abstract:** This article is about data repositories, in particular about metadata for a description of sets of data systems. It addresses the FAIR initiative and FAIR Data Point as a technology that helps to retrieve information from data repositories. It aims to assess, based on FAIR Data Point metadata, the quality of metadata in Repositories in the Biodiversity area. Analyzes how the description of data sets occurs in the Ecological Studies Data Repository and compares the metadata proposed by FAIR Data Point, a system that specifies a set of standardized metadata to be implemented in a Repository. As a methodology, it integrates bibliographic, exploratory, descriptive and documentary research with a qualitative approach. That said, it is concluded that some metadata identified in Ecological Studies Data Repository are common to FAIR Data Point, and others do not present similarity. Furthermore, improvements are proposed in the description of data sets, making their description unique and precise, especially regarding the representation of research data.

**Keywords:** Data repository; FAIR; FAIR Data Point; Metadata; Search data

Recebido: 18/08/2021

Aceito: 10/11/2021

### **Declaração de autoria**

**Concepção e elaboração do estudo:** Carla Beatriz Marques Felipe, Raimunda Fernanda dos Santos.

**Coleta de dados:** Carla Beatriz Marques Felipe, Raimunda Fernanda dos Santos.

**Análise e interpretação de dados:** Carla Beatriz Marques Felipe, Raimunda Fernanda dos Santos.

**Redação:** Carla Beatriz Marques Felipe, Raimunda Fernanda dos Santos.

**Revisão crítica do manuscrito:** Carla Beatriz Marques Felipe, Raimunda Fernanda dos Santos.

### **Como citar:**

FELIPE, Carla Beatriz Marques; SANTOS, Raimunda Fernanda dos. Avaliação de metadados em repositórios de dados de pesquisa sobre biodiversidade. **Em Questão**, Porto Alegre, v. 28, n. 3, e-117591, jul./set. 2022. <https://doi.org/10.19132/1808-5245283.117591>

