

Obtenção do Grau FAIRness de Recursos Digitais Armazenados em Repositórios de Dados - Conceitos, Técnicas e Ferramentas



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

DOCUMENTOS 181

Obtenção do Grau FAIRness de Recursos Digitais Armazenados em Repositórios de Dados - Conceitos, Técnicas e Ferramentas

*Luiz Manoel Silva Cunha
Marcos Cezar Visoli
Luiz Antonio Falaguasta Barbosa*

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

Imagem de capa
worldclouds.com e unclpodger - stock.adobe.com

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

Cunha, Luiz Manoel Silva.

Obtenção do grau FAIRness de recursos digitais armazenados em repositórios de dados – conceitos, técnicas e ferramentas / Luiz Manoel Silva Cunha, Marcos Cezar Visoli, Luiz Antonio Falaguasta Barbosa. - Campinas: Embrapa Agricultura Digital, 2021.

PDF (35 p.) : il. color. - (Documentos / Embrapa Agricultura Digital, ISSN 2764-2488 ; 181).

1. Recursos digitais. 2. Repositórios de dados. 3. Princípios FAIR. 4. FAIRness. I. Visoli, Marcos Cezar. II. Barbosa, Luiz Antonio Falaguasta. III. Título. IV. Embrapa Agricultura Digital. V. Série.

CDD (21. ed.) 004.678

Carla Cristiane Osawa (CRB-8/10421)

© Embrapa, 2021

Autores

Luiz Manoel Silva Cunha

Estatístico, mestre em Ciência da Computação e Matemática Computacional, 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.

Luiz Antonio Falaguasta Barbosa

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

Apresentação

A publicação de dados abertos por parte de governos como forma de transparência e acesso à informação tem se intensificado nas últimas décadas, por diversas ações relacionadas ao movimento Open Government.

Concomitantemente, as iniciativas para promoção da Ciência Aberta, levaram inicialmente a disponibilização de publicações científicas no modo de acesso aberto, ou seja, com permissões de acesso, cópia e redistribuição, além de reuso em novos trabalhos científicos.

Diversos repositórios de publicações foram construídos e são mantidos pelas próprias instituições de pesquisa, além de repositórios sustentados para que pesquisadores e instituições publiquem e divulguem seus trabalhos. Na Empresa Brasileira de Pesquisa Agropecuária (Embrapa) temos os repositórios institucionais de acesso aberto, Alice e Infoteca-e com publicações técnico-científicas e tecnológicas.

As ferramentas tecnológicas voltadas para registros de recursos digitais em base de dados evoluíram a ponto de possibilitar catalogações destes com metadados padronizados, uso de identificadores únicos globais, mecanismos de busca indexada e acoplamento de licenças de uso, características fundamentais para possibilitar que recursos sejam encontrados, acessados, reutilizados e referenciados.

Embora as publicações de acesso aberto colaborarem imensamente para o avanço da ciência aberta, houve a necessidade de que progressivamente se tivesse acesso aos dados de pesquisa, ou seja, aqueles utilizados pelos pesquisadores. A complexidade em relação às publicações destes dados é relativamente mais alta por envolverem dados em formatos, tamanhos e contextos diversificados, que em muitas situações necessitam de descrições diferenciadas.

Os repositórios de dados de pesquisa, seguindo o caminho iniciado pelos repositórios de publicações, se expandiram e diversas instituições de pesquisa já os implantaram ou estão implementando. É o caso do Repositório de Dados de Pesquisa da Embrapa (Redape), que tem sua implantação prevista para dezembro de 2021.

Além dos esforços necessários para a devida catalogação dos dados no Repositório, ações para a medição de quanto estes estarão com características adequadas para serem encontráveis, acessíveis, integráveis e reutilizáveis são necessárias para prover um repositório cada vez mais utilizado, e fomentado com dados de pesquisa. Estas características são mensuradas a partir de indicadores de maturidade, que seguem os princípios *Findable*, *Accessible*, *Interoperable* e *Reusable* (FAIR). Os resultados obtidos neste trabalho como a identificação de indicadores de maturidade e seus respectivos testes de conformidades para verificação dos princípios FAIR, os métodos/abordagens e ferramentas para cálculos do grau de FAIRness são contribuições do Grupo de Pesquisa Engenharia da Informação (GPEI) da Embrapa, no tema de Ecossistema de Dados. A aferição do grau de FAIRness proverá informações importantes sobre aspectos relacionados à qualidade dos dados, em consonância com as ações da Embrapa em produzir e disseminar conhecimento para a agricultura brasileira.

Sílvia Maria Fonseca Silveira Massruhá

Chefe-geral

Sumário

Introdução	7
Princípios FAIR.....	7
Modelos de avaliação de maturidade FAIR de recursos digitais	9
Métodos de mensuração do grau de FAIRness	15
Ferramentas	18
Considerações finais	32
Referências	33

Introdução

Como apresentado por Wilkinson et al. (2016), os princípios FAIR (Go FAIR, 2018), cada vez mais, têm sido adotados pela comunidade científica como uma das formas de aperfeiçoar a infraestrutura de suporte à reutilização de dados de pesquisas.

Representantes da academia, da indústria, das agências de financiamento de pesquisa, editores acadêmicos e cientistas da computação, em conjunto, têm trabalhado no desenvolvimento de diferentes Tecnologias de Informação (TIs) (Clarke et al., 2019b; FAIRassist.org..., 2019; Research Data Alliance, 2020), de maneira a tornar os diferentes recursos digitais (APIs – conjunto de normas que possibilita a comunicação entre plataformas através de uma série de padrões e protocolos, Conjunto de Dados, *Workflow*, documentos digitalizados, etc) armazenados em repositórios de dados, FAIRness. Além das TIs, busca-se também por Indicadores de Maturidade FAIR (IM-FAIR) e seus respectivos Testes de Conformidades (Wilkinson et al., 2016; Research Data Alliance, 2020) e, por métodos mais aperfeiçoados de avaliação do grau de FAIRness.

Como forma de contribuir para que a Embrapa disponha de um processo robusto de aferição do grau de FAIRness dos recursos digitais armazenados em seus repositórios de dados, é necessária a absorção de conhecimentos de diversas naturezas, envolvendo o tema dados FAIR, Indicadores de Maturidade FAIR e Testes de Conformidades e ferramentas para automação deste processo.

Este trabalho tem como objetivo a formação de uma base de conhecimentos envolvendo conceitos, métodos, abordagens e softwares para subsidiar a construção de processo automatizado de obtenção do grau FAIRness de recursos digitais a serem disponibilizados no Redape quando este estiver em uso.

Para o alcance do objetivo definido foi realizado um levantamento bibliográfico não exaustivo, visando entender o que são dados FAIR, Indicadores de Maturidade FAIR e seus Testes de Conformidades e, também, identificados métodos e abordagens possíveis de serem aplicados no processo citado. Três softwares para a automação do processo de verificação do grau FAIRness foram identificados e investigados sob limitações relacionadas às configurações destes e a dados disponíveis para análises. Utilizando-se de discussões periódicas, envolvendo os autores deste trabalho, elaborou-se o conteúdo aqui apresentado.

Entende-se que este trabalho é um passo importante para a internalização, inicialmente, no âmbito do Grupo de Pesquisa Engenharia da Informação (GPEI), da Embrapa Agricultura Digital, de conhecimentos necessários para a construção do processo de obtenção de grau FAIRness de recursos digitais armazenados e repositório de dados.

Os princípios FAIR e seus componentes encontram-se descritos na seção 2. Na seção 3, apresentamos modelos de avaliação de maturidade FAIR de recursos digitais. Em seguida, seção 4, abordamos métodos utilizados para cálculos do grau FAIRness de recursos digitais armazenados em repositórios de dados. Três ferramentas para automação do processo de obtenção do grau FAIRness são relacionadas na seção 5, seguida das considerações finais.

Princípios FAIR

Os princípios FAIR surgiram para garantir que os recursos digitais possam ser encontráveis, acessíveis, interoperáveis e reutilizáveis por máquinas e humanos. Eles funcionam como um guia para os tipos de comportamento que pesquisadores e administradores de dados devem esperar dos

recursos digitais e, por sua vez, informar as condições para que os interessados em publicar seus resultados acadêmicos, torne-os mais FAIR. Estes princípios são crescentes, na medida em que não definem estritamente como atingir FAIRness. Em vez disso, eles descrevem um *continuum* de recursos, atributos e comportamentos que movem um recurso digital para mais perto desses princípios (Wilkinson et al., 2019). A seguir cada um desses princípios e seus elementos, seguidos de suas descrições, são apresentados (Wilkinson et al., 2016, p. 4 tradução nossa).

Findable – Encontrável (F) - Está relacionado à capacidade do dado e metadados serem encontrados, seja por ação humana ou por sistemas computacionais automatizados.

Elemento	Descrição
F1	Os metadados e dados são identificados por identificador digital persistente, de forma única e global.
F2	Os dados devem ser bem descritos com metadados enriquecidos, com o máximo de informação possível para compreensão.
F3	Os metadados incluem identificadores dos dados por eles descritos.
F4	Os metadados são indexados por sistemas de pesquisa.

Accessible – Acessível (A) - Está relacionado à capacidade de o recurso ser acessível e recuperado.

Elemento	Descrição
A1	Os metadados e dados são recuperáveis por meio de identificadores, via protocolos de comunicação padronizados. Ao ser informado um identificador, é possível recuperar os metadados e dados.
A1.1	O protocolo de comunicação é aberto e universalmente implementável.
A1.2	O protocolo de comunicação comporta procedimentos de autenticação e autorização, se necessário.
A2	Os metadados são acessíveis, mesmo quando os dados não estejam disponíveis.

Interoperable – Interoperável (I) - Está relacionado à capacidade do recurso ser interoperável com outras aplicações e sistemas para que possam ser analisados, armazenados, processados, descobertos e reusados.

Elemento	Descrição
I1	Os metadados e dados usam linguagem formal, acessível para representação de conhecimento.
I2	Os metadados e dados utilizam vocabulários que seguem princípios FAIR.
I3	Os metadados e dados incluem referências qualificadas a outros metadados.

Reusability – Reusabilidade (R) - Está relacionado à capacidade dos dados e metadados serem reusáveis, para replicarem e/ou se combinarem para que possam ser reusados.

Elemento	Descrição
R1	Metadados e dados são descritos com uma pluralidade de atributos precisos e relevantes.
R1.1	Metadados e dados são liberados para uso por meio de uma licença clara e acessível.
R1.2	Metadados e dados incluem referências qualificadas a outros metadados.
R1.3	Conteúdo dos metadados e os dados atendem aos padrões da comunidade relevante ao domínio.

Modelos de avaliação de maturidade FAIR de recursos digitais

Segundo o projeto *Atmosphären-Modelldaten: Datenqualität, Kurationskriterien und DOI-Branding (AtMoDat)*, maturidade descreve o grau de formalização e padronização de um objeto de dados (dados + metadados) com respeito à FAIRness e a qualidade dos metadados e dos dados (The AtMoDat..., 2021). Objetos de dados se aperfeiçoam à medida que passam por diferentes etapas de pós-produção dos dados. Quanto maior a maturidade, mais fácil é a reutilização dos dados (Heydebreck et al., 2020). Um indicador se refere a um elemento que aponta ou mostra algo. Os Indicadores de Maturidade (IMs) FAIR são especificações de autoria que delimitam um comportamento FAIR próprio, mensurável automaticamente (Wilkinson et al., 2019).

Modelo escalável, automático e gerenciado pela comunidade

Wilkinson et al. (2019) definiram um conjunto de 15 IMs visando à obtenção do grau de maturidade dos recursos digitais. Esses indicadores cobrem todos os princípios FAIR e seus respectivos elementos, exceto aqueles referidos como R1.2 (metadados voltados para proveniência) e R1.3 (metadados e dados que descrevem o padrão da comunidade).

Em alguns dos IMs são encontradas as palavras fraco ou forte associada, que revela o rigor de como o recurso digital será testado. Para cada um dos IMs, pelo menos, um Teste de Conformidade (TC) foi definido (Wilkinson et al., 2019). O teste é do tipo funcional, caixa-preta, no qual uma implementação (protocolo) é testada para conformidade, somente em relação aos requisitos mencionados explicitamente na sua especificação. Em determinadas situações, dois testes são aplicados: o primeiro para verificar a ocorrência da característica no recurso e o segundo para identificar a forma como ela acontece. Os Testes de Conformidade (TCs) podem ser agrupados em coleções customizáveis, que permitem que grupos e interessados possam criar a relevância de seus próprios domínios.

Os IMs definidos em Wilkinson et al. (2019) atendem a muitas comunidades científicas e são assimilados/utilizados como padrões. No entanto, outros IMs e seus respectivos TCs, podem ser criados para atender às demandas específicas. Essas novas criações devem seguir o padrão FAIR. Em

fairsharing.org¹, identifiers.org² e Bioportal³ são apresentados exemplos de registros de metadados para padrão de metadados incluindo formatos de arquivos, ontologias, esquema de identificadores, assim como para indicadores de maturidade.

Os IMs associados aos princípios FAIR e os TCs são apresentados na Tabela 1, maiores detalhes encontram-se em Wilkinson et al. (2019). Para cada indicador está associado o elemento vinculado ao princípio FAIR correspondente, informações estas apresentadas na seção 2.

Tabela 1. Indicadores de Maturidade FAIR.

Indicadores de Maturidade FAIR	
Indicador	Elemento/Princípio FAIR
Identificador único	F1
Identificador de persistência	F1
Metadado estruturado	F2
Metadado vinculado	F2
Uso de Global Unique Identifier (GUID) em metadados	F3
Metadado sendo indexado em recurso pesquisável	F4
Protocolo aberto para recuperação de dados e metadados	A1.1
Protocolo para suporte a autenticação/autorização	A1.2
Persistência de metadados	A2
Uso de uma linguagem de representação do conhecimento (fraca)	I1
Uso de uma linguagem para representação de conhecimento (robusta)	I1
Uso de vocabulários FAIR (fraco)	I2
Uso de vocabulários FAIR (robusto).	I2
Links externos qualificados	I3
Metadados contêm um link para uma licença	R1.1

Testes de Conformidades

A correspondência entre os IMs e os TCs não é única. Os Testes podem ser agrupados em classes e serem customizáveis, visando atender às necessidades de um domínio específico. Abaixo, são relacionados os TCs tidos como base, segundo a literatura consultada:

1. Identificador único.
2. Identificação de dados explicitamente em metadados.
3. Identificação de metadados explicitamente em metadados.
4. Identificador de persistência.

1 Disponível em: <<https://fairsharing.org/>>.

2 Disponível em: <<https://identifiers.org/>>.

3 Disponível em: <<https://bioportal.bioontology.org/>>.

5. Identificador de dados de persistência.
6. Metadados são estruturados.
7. Metadados são agrupados.
8. Principal recurso de localização definido.
9. Uso de protocolo aberto para recuperação de dados.
10. Uso de protocolo aberto para recuperação de metadados.
11. Autenticação e autorização de dados.
12. Autenticação e autorização de metadados.
13. Persistência de metadados.
14. Linguagem de representação de conhecimento de metadados (fraco).
15. Linguagem de representação de conhecimento de metadados (robusto).
16. Linguagem de representação de conhecimento de dados (fraco).
17. Linguagem de representação de conhecimento de dados (robusto)
18. Metadados usam vocabulários FAIR (fraco).
19. Metadados usam vocabulários FAIR (robusto).
20. Metadados contêm referências externas qualificadas.
21. Metadados incluem licença (forte).
22. Metadados incluem licença (robusto).

Modelo apoiado em software

Em Fairassist.org (FAIRassist.org..., 2019), são apresentados diferentes softwares que suportam modelos para o estabelecimento do grau de maturidade FAIR, para recursos digitais. Os modelos implementam uma das três abordagens abaixo apresentadas (Azevedo; Dumontier, 2020):

1. Avaliação baseada em questionário de respostas distintas – faz uso de questionários online tipo *self-report* (autorrelato) e *checklists* (lista de verificação), em que os entrevistados precisam indicar sua escolha de implementação (para um determinado princípio), a partir de um conjunto predefinido de respostas. As questões são agrupadas de acordo com cada princípio: F, A, I e R. O resultado da avaliação é uma pontuação que pode ser a soma, a soma ponderada ou uma pontuação visual, gerada automaticamente após a conclusão da avaliação. Essa abordagem traz como limitações: a) o conjunto de respostas não é extensível, o que significa que não é possível incluir padrões adicionais; b) os itens não fornecem evidências de uma implementação FAIR, mas apenas a intenção; e c) o cálculo da pontuação é arbitrário.
2. Avaliação baseada em questionário de respostas abertas - consiste em 14 métricas de “uso geral” refletindo os princípios FAIR e suas frações. Essas métricas foram renomeadas para IMs FAIR, pela capacidade de indicarem os pontos específicos que podem ser melhorados, além de “medir” o FAIRness. Como pontos fracos, esse tipo de avaliação apresenta limitações para responder virtuosamente a questionários dessa categoria, consumindo muito tempo e sujeito ao

viés do entrevistado. A falta de material instrucional para auxiliar os respondentes na condução das avaliações é outra limitação encontrada. Além disso, as propriedades psicométricas dessa ferramenta também não foram avaliadas, o que compromete sua validade e confiabilidade, portanto, qualquer tentativa de calcular os escores totais deve ser considerada experimental. Os pontos fortes são: a) as respostas precisam incluir uma declaração que evidencie a implementação FAIR (por exemplo, um URI de um registro de metadados do padrão adotado); b) permite às comunidades científicas criarem IMs adicionais; e c) podem ser preenchidos em múltiplas ocasiões (versão em planilha).

3. Avaliação semiautomática - a abordagem FAIR *Maturity Evaluation Service* consiste em um coletor de metadados baseado na web, que analisa os metadados de um recurso digital atribuído a um identificador único global (GUID). Uma breve descrição do recurso a ser avaliado, o GUID do recurso e o *Open Research and Contributor ID* (ORCID) devem ser fornecidos antes do início do processo. Anterior a tudo isso, um grupo de IMs deve ser selecionado para conduzir a avaliação. Por ser uma abordagem automatizada reduz o viés do respondente, é de código aberto e garante transparência, pois as avaliações são abertas e os avaliadores identificáveis. Como limitações, essa abordagem requer que o recurso tenha algum tipo de provedor de metadados disponível, depende da compatibilidade entre software e provedor de metadados e tem um desempenho diferente ao comparar dois identificadores diferentes do mesmo recurso.

Se a intenção é unicamente avaliar o FAIRness do recurso digital, de forma restrita, recomenda-se iniciar pelo modo semiautomático, uma vez que o resultado emitido pode não refletir as necessidades de outros interessados. As diferentes abordagens para mensuração devem ser complementares, em vez de simultâneas. O resultado dos questionários de respostas abertas pode ser útil para capturar informações sobre os padrões que ainda precisam ser implementados para os testes de maturidade automatizados. Além disso, pode também auxiliar na coleta de dados destinados à implementação de opções tidas como importantes para uma determinada comunidade científica.

Modelo Research Data Alliance (RDA)

O modelo de maturidade proposto pela RDA (Research Data Alliance, 2020) apresenta um conjunto de indicadores apoiados nos princípios FAIR, em prioridades e nos métodos de avaliação FAIRness para os recursos digitais. Ele busca com que os indicadores, os níveis de maturidade e a priorização sejam entendidos da mesma forma. Esse modelo não tem a intenção de ensinar “como fazer a avaliação” e, sim, apresenta uma forma de normalizá-la. Os elementos que compõem o modelo são:

1. **Indicadores:** são derivados dos princípios FAIR e dos aspectos individuais do FAIRness, que são avaliados. Eles buscam responder um único propósito: **O que** precisa ser medido para avaliar o FAIRness e não como medir o FAIRness do recurso digital. Os indicadores podem ser aplicados em diferentes contextos, envolvendo algoritmos, ferramentas, fluxos de trabalho, protocolos e outros serviços relacionados a dados que são produzidos ou gerenciados, na medida em que são disponibilizados como objetos digitais. Para cada aspecto distinto, que pode ser medido pela descrição do princípio, há um indicador.
2. **Prioridades:** indica a ordem de aplicação do indicador para a medição do FAIRness. Os níveis de prioridades são:
3. **Essencial:** aspecto de extrema importância para o alcance do FAIRness na maioria das circunstâncias; sem utilizá-lo é praticamente impossível de alcançá-lo se o indicador não for satisfeito.
4. **Importante:** significa que o indicador aborda um aspecto que não é de máxima importância,

mas quando apresentado contribui para o aumento do FAIRness.

5. **Útil:** informação de que o indicador aborda um aspecto que é bom estar presente, mas o indicador pode não ser atendido. Não é necessariamente indispensável.

Na Tabela 2 estão relacionados os indicadores propostos no Modelo de Maturidade de Dados FAIR – RDA, agrupados por princípios FAIR e, registrando o nível de prioridade.

Tabela 2. Indicadores de Maturidade FAIR – Modelo RDA

Elemento	ID	Descrição	Nível de Prioridade
F1	RDA-F1-01M	O metadado é identificado por um identificador de persistência.	Essencial
F1	RDA-F1-01D	O dado é identificado por um identificador persistente.	Essencial
F1	RDA-F1-02M	O metadado é identificado por um identificador global único.	Essencial
F1	RDA-F1-02D	O dado é identificado por um identificador global único.	Essencial
F2	RDA-F2-01M	Um metadado robusto é fornecido para permitir a descoberta.	Essencial
F3	RDA-F3-01M	O metadado inclui o identificador para o dado.	Essencial
F4	RDA-F4-01M	O metadado é oferecido de forma tal que possa ser coletado e indexado.	Essencial
A1	RDA-A1-01M	O metadado contém informações que habilitam os usuários a terem acesso ao dado.	Importante
A1	RDA-A1-02M	O metadado pode ser acessado manualmente (com intervenção humana).	Essencial
A1	RDA-A1-02D	O dado pode ser acessado manualmente (com intervenção humana).	Essencial
A1	RDA-A1-03M	O identificador endereça a um registro de metadados.	Essencial
A1	RDA-A1-03D	O identificador do dado endereça a um objeto digital.	Essencial
A1	RDA-A1-04M	O metadado é acessado através de um protocolo padrão.	Essencial
A1	RDA-A1-04D	O dado é acessível por meio de um protocolo padrão.	Essencial
A1	RDA-A1-05D	O dado pode ser acessado automaticamente (por software).	Importante

A1.1	RDA-A1.1-01M	O metadado é acessível por meio de protocolo de acesso aberto.	Essencial
A1.1	RDA-A1.1-01D	O dado é acessível por meio de um protocolo livre.	Importante
A1.2	RDA-A1.2-02D	O dado é acessível por meio de um protocolo de acesso que suporta autenticação e autorização. É garantido que os metadados permanecerão disponíveis depois que os dados não estiverem mais disponíveis.	Útil
A2	RDA-A2-01M	Os metadados usam representação do conhecimento expressado no formato padrão.	Essencial
I1	RDA-I1-01M	Os dados usam representação do conhecimento expressado no formato padrão.	Importante
I1	RDA-I1-01D	Os metadados usam representação do conhecimento reconhecido por máquina.	Importante
I1	RDA-I1-02M	Os dados usam representação de conhecimento reconhecido por máquina.	Importante
I1	RDA-I1-02D	Os metadados usam vocabulário FAIR controlados.	Útil
I2	RDA-I2-01M	Os metadados usam vocabulários FAIR controlados.	Útil
I2	RDA-I2-01D	Os metadados incluem referências para outros metadados.	Importante
I3	RDA-I3-01M	Os dados incluem referências para outros dados	Útil
I3	RDA-I3-01D	Os metadados incluem referências para outros dados.	Útil
I3	RDA-I3-02M	Os dados incluem referências qualificadas para outros dados.	Útil
I3	RDA-I3-02D	Os metadados incluem referências qualificadas para outros metadados.	Importante
I3	RDA-I3-02M	Os metadados incluem referências para outros dados.	Útil
R1	RDA-R1-01M	A variedade de atributos precisos e relevantes são fornecidos para permitir a reutilização	Essencial
R1.1	RDA-R1.1-02M	Os metadados incluem informações sobre a licença a qual os dados podem ser reusados.	Essencial
R1.1	RDA-R1.1-02M	Os metadados referem-se a um padrão de licença de reuso.	Importante

R1.1	RDA-R1.2-03M	Os metadados referem-se à licença de reuso compreensível por máquina.	Importante
R1.2	RDA-R1.2-02M	Os metadados incluem informações de proveniência de acordo com a referência cruzada.	Útil
R1.3	RDA-R1.3-01M	Os dados estão em conformidade com o padrão pré-estabelecido.	Essencial
R1.3	RDA-R1.3-02M	Os metadados são expressados em conformidade com o padrão pré-estabelecido.	Essencial
R1.3	RDA-R1.3-02D	Os dados são expressados em conformidade com o padrão pré-estabelecido.	Importante

A distribuição das prioridades por princípios FAIR apresentados na Tabela 2, é apresentada na Tabela 3. Como observado, entre os indicadores utilizados para encontrar o FAIRness dos recursos digitais, 20 deles são classificados como essenciais, 14 como importantes e 7 como úteis.

Tabela 3. Distribuição de prioridades por princípios FAIR.

Prioridade	Princípios				
	F	A	I	R	Total
Essencial	7	8	0	5	20
Importante	0	3	7	4	14
Útil	0	1	5	1	7
Total	7	12	12	10	41

Fonte: Adaptado de Research Data Alliance (2020).

Métodos de mensuração do grau de FAIRness

Como apresentado em FAIRassist.org (FAIRassist.org..., 2019), variações nos métodos de mensuração do grau de FAIRness encontram-se implementados em diferentes softwares. Eles permitem identificar se o padrão declarado nos metadados foi devidamente seguido ou se ainda não foi oferecido por meio de uma funcionalidade. No modelo proposto pela RDA, os métodos são:

- 1. Mensuração do progresso** - foca em até que ponto um recurso, sob avaliação, atende aos requisitos do indicador para responder a pergunta: “Como pode ser aperfeiçoado o FAIRness deste dado?”. Cinco níveis de maturidade foram definidos para o indicador:
 - 0 – Não aplicável.
 - 1 – Não considerado ainda.
 - 2 – Em consideração ou em fase de planejamento.
 - 3 – Em fase de implementação.
 - 4 – Totalmente implementado.

O progresso FAIRness, por indicador, é obtido avaliando cada indicador contra os cinco níveis apresentados acima. Assim, é possível descartar o indicador que se mostra irrelevante. Esta abordagem

pode ser útil para que provedores de dados e editores realizem testes de avaliação para terem uma melhor ideia de onde concentrar esforços para tornarem seus recursos mais FAIR.

2. **Mensuração da aprovação ou reprovação** - foca em determinar como um recurso, sob avaliação, atende aos requisitos definidos para os indicadores FAIR. Trata-se de uma avaliação mais restrita, atuando numa escada binária, nesse sentido, somente são contabilizados os indicadores que alcançarem nível 4, definido na abordagem anterior. Além de medir a aprovação ou reprovação dos princípios que compõem o agrupamento em que eles se encontram, mede-se também o FAIRness deste agrupamento, com base no alcance de prioridades. Esta abordagem é útil para os avaliadores externos ou agência de fomento que queiram verificar se os níveis de FAIRness predefinidos para os recursos, que eles gerenciam, estão sendo cumpridos.
3. **Combinação das abordagens** – esta estratégia busca extrair os benefícios das abordagens anteriores. Aplicando a primeira abordagem, gráficos de radar são construídos. Este tipo de gráfico é uma forma de apresentar vários dados que possuem mais de um eixo em um único lugar. Dessa forma não é necessário um gráfico tridimensional ou algo parecido (Gráficos..., 2021). A segunda abordagem, sumariza todos os indicadores de maior nível, isto é, aqueles que tiveram seus requisitos totalmente implementados.

Segundo *Consultative Group on International Agricultural Research* (2021b), um dos grandes desafios da ciência apoiada em grandes volumes de dados, para humanos e máquinas, é facilitar a descoberta de conhecimentos. Na tentativa de facilitar este processo, incorporando os princípios FAIR, *Consultative Group on International Agricultural Research* (2021b) adotou os princípios utilizados pelo *Netherlands Institute for Permanent Access to Digital Research Resources* (DANS) para verificar o grau de maturidade FAIRness para os recursos digitais armazenados em repositórios. Para cada princípio FAIR, níveis foram propostos e, para cada um deles, uma pontuação definida. De posse da pontuação alcançada tem-se a ideia de onde melhorar para o alcance do maior nível possível de FAIRness. A abordagem aqui apresentada encontra-se publicada em *Consultative Group on International Agricultural Research* (2021a). A seguir, um resumo dessa abordagem é visualizada abaixo e nas Figuras 1, 2 e 3.

Princípio Findable (F)

- Nível 0: Não há Indicador de Persistência (PID), sem metadado e/ou documentação (Pontos FAIR= 0).
- Nível 1: Não há PID, metadados e/ou documentação insuficiente (Pontos FAIR = 1).
- Nível 2: Não há PID mas metadados e/ou documentação suficiente (Pontos FAIR = 2).
- Nível 3: Não existe PID mas extensiva quantidade de metadados e/ou documentação (Pontos FAIR = 4).
- Nível 4: Existência do PID com extensa quantidade de metadados e documentação (Pontos FAIR = 4.5).
- Nível 5: Metadados e dados registrados ou indexados em um recurso pesquisável.

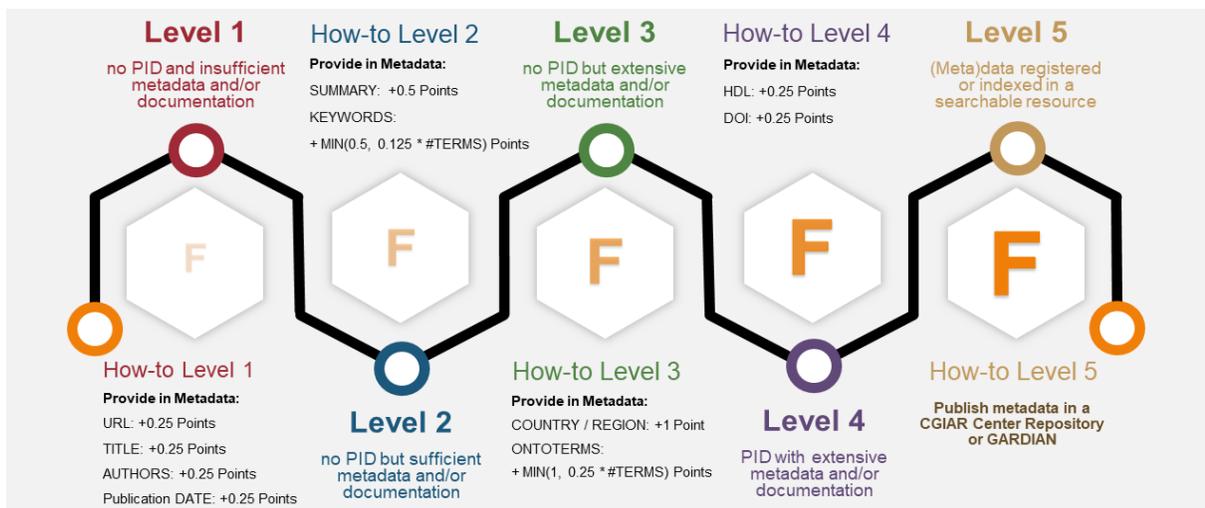


Figura 1. Níveis do princípio *Findable* (F).

Fonte: Consultative Group on International Agricultural Research (2021b).

Princípio *Accessible* (A)

- Nível 0: Não há licença para o usuário/condições de reuso pouco esclarecedoras (Pontos FAIR = 0).
- Nível 1: Acesso limitado (Pontos FAIR = 1)
- Nível 2: Acesso aberto (com restrições) (Pontos FAIR = 2)
- Nível 3: Acesso aberto (sem restrições) (Pontos FAIR = 3.5)
- Nível 4: Acesso fornecido aos arquivos físicos (Pontos FAIR = 4.5)
- Nível 5: Metadados acessíveis (mesmo quando os dados não estão ou estão indisponíveis) (Pontos FAIR = 5)

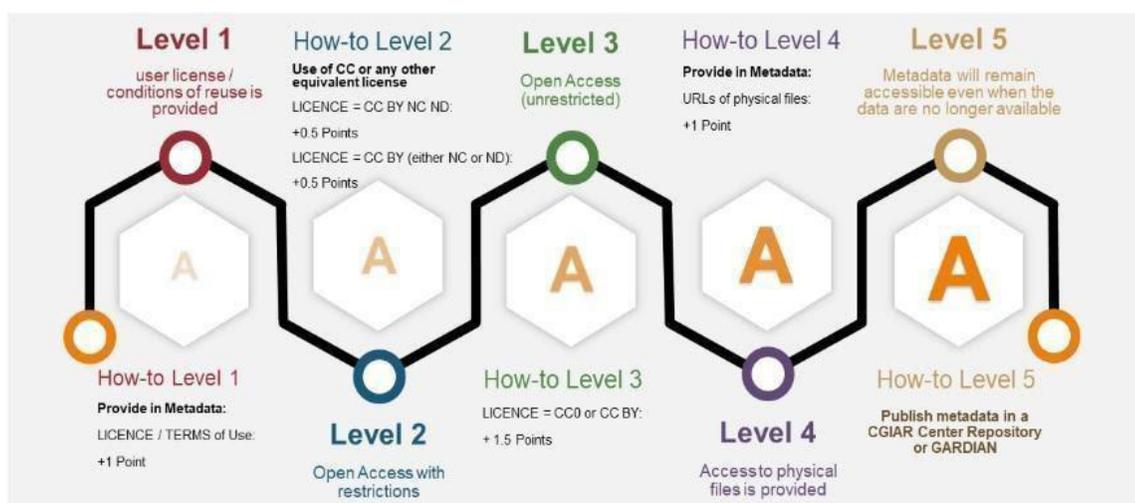


Figura 2. Níveis do princípio *Accessible* (A).

Fonte: Consultative Group on International Agricultural Research (2021b).

Princípio *Interoperable* (I)

Nível 0: Formato de dados proprietário, não aberto (Pontos FAIR = 0).

- Nível 1: Formato proprietário, aceito por Certified Trusted Data Repository (Pontos FAIR = 1).
- Nível 2: Formato aberto e não proprietário (Pontos FAIR = 2.5).
- Nível 3: Dados são uniformizados, usando vocabulários padrão (Pontos FAIR = 3.5).
- Nível 4: Metadados e dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento (Pontos FAIR = 4.5).
- Nível 5: Dados são ligados a outros dados para fornecer o contexto (Pontos FAIR = 5).

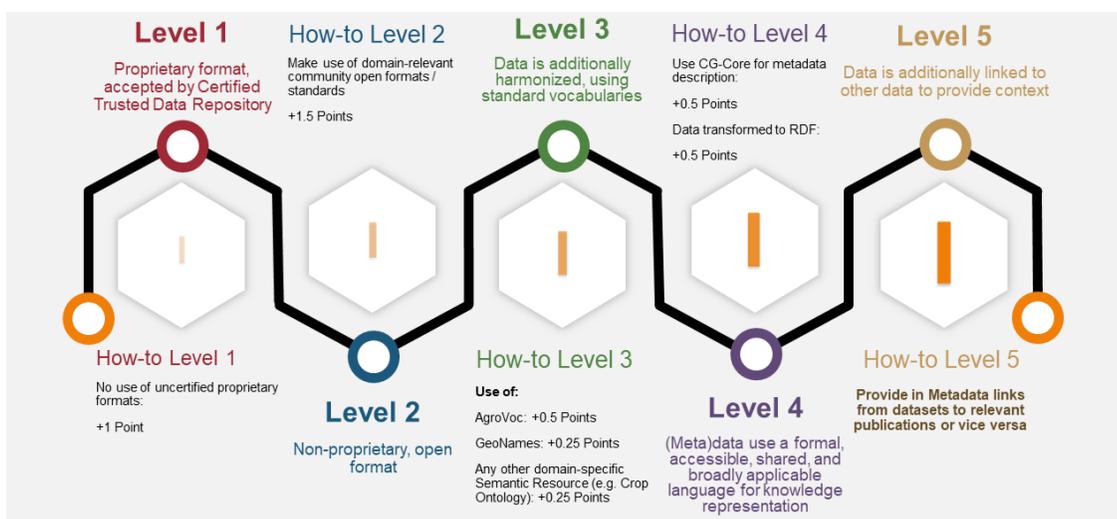


Figura 3. Níveis do princípio Interoperable (I).

Fonte: Consultative Group on International Agricultural Research (2021b).

Princípio Reusable (R)

Para ser *Reusable*, os princípios se sobrepõem aos princípios *Accessible* e *Interoperable*.

- R1. metadados e dados têm uma pluralidade de atributos precisos e relevantes.
- R1.1. metadados e dados são liberados com uma licença de uso clara e acessível.
- R1.2. metadados e dados estão associados à sua proveniência.
- R1.3. metadados e dados atendem aos padrões definidos para a comunidade pertencente a um domínio.

O resultado do princípio *Reusable* é estabelecido aplicando a fórmula: $R = (F+A+I)/3$.

Ferramentas

Cada vez mais recursos digitais, tais como, aplicativos, sites na internet, softwares, imagens, entre outros são produzidos e, torná-los FAIRness tem sido um desafio. Nesta seção apresentaremos os softwares FAIR-Aware, FAIR Evaluation Services e FAIRshake que implementam os seguintes tipos de execução: a) Manual - questionnaire; b) Automated; e c) Manual - questionnaire, Semi-manual, respectivamente. Como ponto comum, todos eles além de apresentarem a informação referente ao

grau de FAIRness, também emitem um relatório contendo informações que podem ser utilizadas para a elevação deste grau.

FAIR-Aware

FAIR-Aware (Data Archiving and Networked Services, 2021) se apresenta de forma on-line, foi desenvolvida no escopo do projeto FAIRsFAIR⁴, cujo objetivo é auxiliar no entendimento dos princípios FAIR e utilizá-los para tomar os dados FAIR, aumentando o valor potencial e o impacto de seus dados. Para isso, é desejável ter em mãos um conjunto de dados que seja capaz de responder às perguntas e concluir a avaliação. A estratégia empregada por esta ferramenta para o alcance de seu objetivo, passa por:

- Capturar informações que caracterizam o responsável pelo conjunto de dados a ser avaliado, conforme a Figura 4.
- Responder um questionário contendo 10 questões, envolvendo os princípios FAIR, dividido da seguinte forma: a) *Findable* (3); b) *Accessible* (2); c) *Interoperable* (1); e d) *Reusable* (4), no formato sim ou não, para cada princípio, subquestões são apresentadas e, a elas, uma nota na escala de 5 a 1 é atribuída. O valor 5 indica muito provavelmente, e o valor 1 significa muito improvável. Essa nota caracteriza o quanto a resposta Sim ou Não, apresentada à questão, será efetivada. Em seguida, quatro outras questões são apresentadas com intuito de levantar informações visando incorporar melhorias na ferramenta. A primeira delas é composta por 10 itens. A Figura 5 apresenta o formulário utilizado na coleta de dados para o princípio *Findable*. Este modelo de formulário se repete para os demais princípios. A Figura 6 apresenta as questões para o aperfeiçoamento da ferramenta FAIR-Aware.

About you

Which research domain do you work in? 

Domain ▾

Which of the following describes your role? Please select all that apply.

- | | |
|--|------------------------------------|
| <input type="checkbox"/> Researcher | <input type="checkbox"/> Funder |
| <input type="checkbox"/> Policy maker | <input type="checkbox"/> Publisher |
| <input type="checkbox"/> Research support (e.g. data steward, curator, data manager, librarian, information technology professional) | <input type="checkbox"/> Other |

Which of the following types of organisations best describe your employer? Please select all that apply.

- | | |
|--|---------------------------------------|
| <input type="checkbox"/> Research Infrastructure/eInfrastructure (e.g. data repository, service provider, library) | <input type="checkbox"/> Funding Body |
| <input type="checkbox"/> University | <input type="checkbox"/> Publisher |
| <input type="checkbox"/> Research Performing Organisation | <input type="checkbox"/> Industry |
| <input type="checkbox"/> Government | <input type="checkbox"/> Other |
| <input type="checkbox"/> eInfrastructure (e.g. repository or scientific data provider) | |

Figura 4. Formulário de caracterização do responsável pelo conjunto de dados.

Fonte: Data Archiving and Networked Services (2021).

FINDABLE

F

1. Are you aware that a dataset should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository? i

Yes To what degree do you intend to comply with this?
 No Very likely 5 4 3 2 1
Very unlikely

2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others? i

Yes To what degree do you intend to comply with this?
 No Very likely 5 4 3 2 1
Very unlikely

3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans? i

Yes To what degree do you intend to comply with this?
 No Very likely 5 4 3 2 1
Very unlikely

Figura 5. Formulário para captura de dados para o princípio *Findable*.

Fonte: Data Archiving and Networked Services (2021).

Feedback

Please answer the four questions below to help us improve the tool and make it even more relevant for the community

- Which of the following issues do you find hard to understand and/or to answer?

1. Globally unique persistent identifier (PID) i
2. Metadata for citation and discovery including PID i
3. Metadata available online is readable by humans and machines i
4. Metadata includes licence, level of access and conditions to access the data i
5. Persistence of metadata i
6. Use of controlled vocabularies in metadata i
7. Metadata includes provenance i
8. Community-endorsed metadata i
9. Data in a preferred format for reuse & preservation i
10. Digital curation and preservation i

- Are there any issues relevant to your discipline and/or needed for enabling reusability of your dataset that are missing from this assessment? (Please do not include any personal data, such as your name and email address in your response.)

- Please submit any other feedback on how we might improve the FAIR-Aware assessment tool, including possible additional guidance. (Please do not include any personal data, such as your name and email address in your response.)

- Undertaking this assessment helped raise my awareness and my understanding of the FAIR data Principles.

- Strongly Agree Agree Neutral Disagree Strongly Disagree

Submit

Figura 6. Formulário para captura de dados para aperfeiçoamento da ferramenta FAIR-Aware.

Fonte: Data Archiving and Networked Services (2021).

Para compreender o funcionamento da ferramenta, um exercício foi realizado. Das dez perguntas, seis foram respondidas como Sim, com grau de caracterização 4 ou 5. As 4 perguntas restantes foram marcadas como Não, recebendo graus variando de 3 a 5. A Tabela 4 apresenta um resumo das informações utilizadas para definição do nível de conhecimento, por parte do usuário sobre os princípios FAIR e o nível de comprometimento em manter as respostas fornecidas para cada uma das questões.

Tabela 4. Resumo das informações para definição do nível de conhecimento dos princípios FAIR.

Perguntas	Respostas	Grau de Caracterização	Cálculo
6	Sim	4, 5	$4 \times 5 + 2 \times 4 = 28$
4	Não	3, 4, 5	$2 \times 4 + 1 \times 3 + 1 \times 5 = 16$

A Figura 7 apresenta o nível e a nota atribuída ao grau de conhecimento sobre os princípios FAIR e o compromisso em mantê-los por parte do usuário.



Figura 7. Nível de conhecimento sobre os princípios FAIR e o comprometimento em mantê-los por parte do usuário.

Fonte: Data Archiving and Networked Services (2021).

Neste exercício, o grau de conhecimento demonstrado sobre os princípios FAIR foi classificado como Moderado, das dez questões respondidas, seis delas foram marcadas como Sim. Em relação ao item *Willingness to comply* (manter as respostas destinadas aos princípios FAIR), o grau apresentado foi considerado alto, 44 pontos alcançados em 50 possíveis.

A segunda parte do relatório, Figura 8 gerado pelo FAIR Aware, traz um conjunto de informações visando aperfeiçoar os conhecimentos sobre os princípios FAIR, já adquiridos pelo usuário. Estas informações são apresentadas após a execução do software.

Guidance:

Based on your answers, you can find the guidance below to improve your awareness on some FAIR issues.

3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans?

This question refers to the capability to make the metadata accessible online in a standard and machine-readable format. Machine-readability refers to a structured format that can be read and processed by a computer.

Advantages of machine-readable formats

By ensuring that the metadata describing your data deposit is machine-readable, your data deposit will be visible to search engines and aggregators which can improve your chances of having your data cited and reused. This capability is normally offered by the data repository that holds the data and makes the essential details about it available online. For example, some repositories expose the metadata online to allow access or harvesting through a standard protocol (e.g. OAI-PMH) and/or a web service. Your choice of a data repository will be critical to meet this requirement. You can select a repository offering such capability by searching for example in the Registry of Research Repositories re3data.org

Figura 8. Orientações para melhoria do conhecimento sobre os princípios FAIR.

Fonte: Data Archiving and Networked Services (2021).

Em relação às orientações descritas na Figura 8 elas estão voltadas para as questões respondidas como Não, são elas:

- Você está ciente de que o repositório que fornece acesso ao seu conjunto de dados deve disponibilizar os metadados que descrevem este conjunto em um formato legível por máquinas e, também, por humanos?
- Você está ciente de que os metadados que descrevem seus conjuntos de dados devem usar vocabulários semânticos?
- Você está ciente de que as informações de proveniência sobre a coleta e/ou geração de dados devem ser incluídas nos metadados?
- Você está ciente de que manter seu conjunto de dados FAIR ao longo do tempo requer curadoria e preservação profissional de dados?

FAIR Evaluation Services

Esta ferramenta é uma evolução do FAIR-Aware. Wilkinson et al. (2019) propuseram um modelo escalável e automatizado para avaliação de recursos digitais, incorporando: a) IM – especificações canceladas por especialistas do domínio que delimitam um recurso FAIR; b) Testes de Conformidades – pequenas aplicações para web que testam os recursos digitais aplicando os IMs, de forma individualizada; e c) Avaliador - uma aplicação web que registra, organiza em coleções um

conjunto de Testes de Conformidades de interesse para o usuário, em relação a um recurso digital, e fornece um relatório sobre o que a máquina “visualizou” após checar o recurso. No processo de estruturação do FAIR Evaluation Services, algumas premissas voltadas para a avaliação FAIRness foram estabelecidas:

- Qualquer indivíduo deve ser capaz de avaliar qualquer objeto digital.
- Um indivíduo ou preferencialmente uma comunidade pode criar e publicar IM ou uma coleção de IMs.
- Avaliações devem contemplar domínios relevantes e testes de conformidade definidos pela comunidade.
- Todos os componentes do modelo de avaliação devem ser, eles mesmos, FAIR.
- A avaliação deve ser mais objetiva quanto possível; o código de avaliação deve evitar acomodar comportamentos fora dos padrões, se possível.
- A avaliação deve fornecer ao usuário não apenas o resultado, mas também orientações claras para a elevação do nível de FAIRness.
- O modelo de avaliação deve ser modular e expansível.
- O modelo pode ser automaticamente atualizado para acomodar novos padrões ou tecnologias, se possível, para minimizar a manutenção e ser sustentável no tempo.

Além destas premissas, um *workflow* do modelo conceitual para a avaliação dos recursos digitais foi estabelecido, apresentado na Figura 9. Os dois únicos processos obrigatórios são os assinalados em “Verde”. Os itens em “Amarelo” são inseridos somente se um conjunto de IMs e/ou Coleções determinadas não forem suficientes.

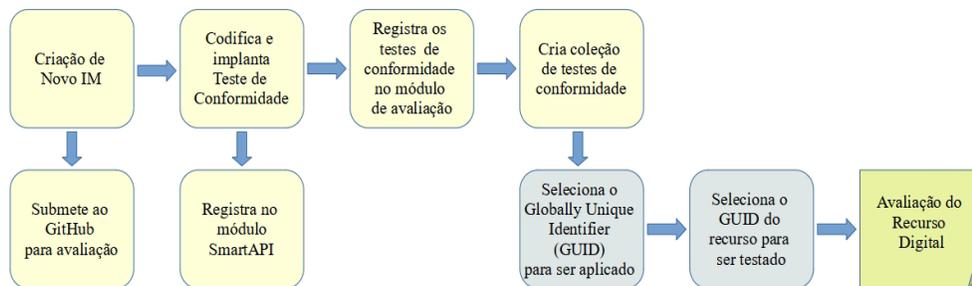


Figura 9. *Workflow* do modelo conceitual para avaliação de recursos digitais.

Fonte: Adaptado de Wilkinson et al. (2019).

A Figura 10 exibe a janela principal da ferramenta FAIR Evaluation Services. Na parte superior da janela estão dispostas as opções para: a) disparar o processo de avaliações (EVALUATIONS); b) registro e visualização de Testes de Conformidades (MI Tests); c) inserções, visualizações e testes das Coleções de IM estabelecidas (COLLECTIONS). O acesso ao mecanismo de busca e a opção ABOUT US encontram-se também disponibilizados. Na parte central dessa janela encontramos as opções: a) Import MI Tests – que permite importar os testes para os indicadores de maturidade, como anotações de interface YAML smartAPI; b) Create collections – reúne os testes de maturidade, em coleções, estas direcionadas para a comunidade; c) Evaluate resources – Avalia o

FAIRness dos recursos digitais utilizando as coleções de testes de maturidade estabelecidos. Logo abaixo, são fornecidas informações sobre a filosofia do teste FAIR, de como produzir e registrar novos indicadores de maturidade e como criar os testes para os indicadores.

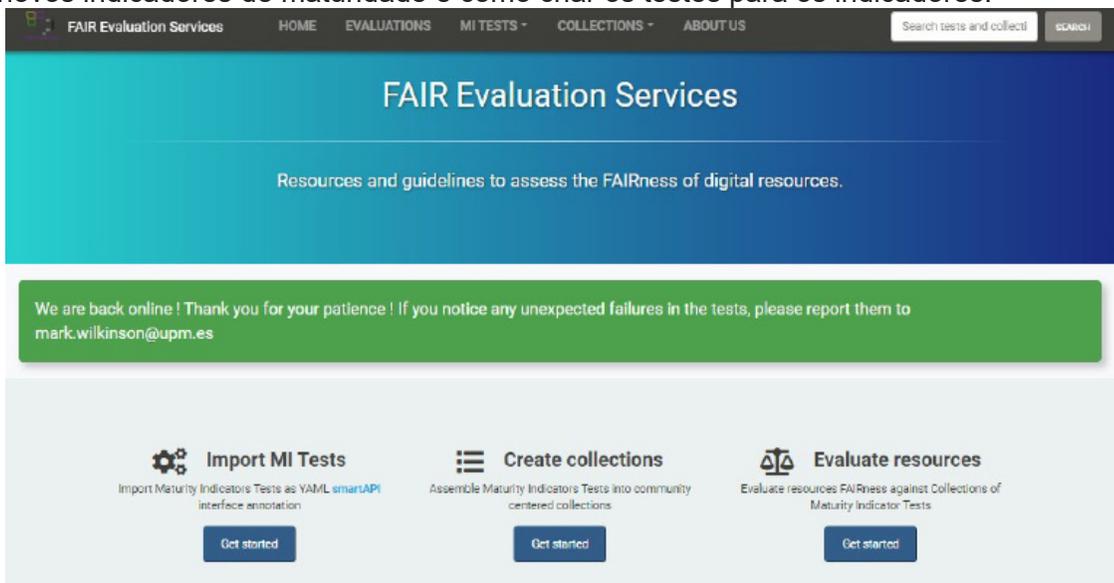


Figura 10. Janela principal da ferramenta FAIR Evaluation Services.

Fonte: FAIR Evaluation Services (2021).

Ao ativar o módulo EVALUATIONS, o sistema exibe como resposta do processo de avaliação do grau FAIRness de um recurso digital, as informações dispostas na Figura 11.

FAIR Maturity Indicator - Evaluations

Search terms

< Previous 1 2 3 ... 111 112 113 Next >

Collection	Title	Description	Executor	View
6	Human Protein Atlas FAIRsharing record against all MIs	FAIR Metrics Evaluation: Human Protein Atlas FAIRsharing record against all MIs; Tested identifier: 10.25504/FAIRsharing.j0t0pe; generated by https://orcid.org/0000-0001-6960-357X	ID	
6	Zenodo record of the UPM Data Science Virtual Machine evaluated against all Maturity Indicators	FAIR Metrics Evaluation: Zenodo record of the UPM Data Science Virtual Machine evaluated against all Maturity Indicators; Tested identifier: 10.5281/zenodo.1147435; generated by https://orcid.org/0000-0001-6960-357X	ID	
1	Zenodo record of the UPM Data Science Virtual Machine evaluated against all Maturity Indicators	FAIR Metrics Evaluation: Zenodo record of the UPM Data Science Virtual Machine evaluated against all Maturity Indicators; Tested identifier: 10.5281/zenodo.1147435; generated by https://orcid.org/0000-0001-6960-357X	ID	

Figura 11. Resultado da avaliação.

Fonte: FAIR Evaluation Services (2021).

Nesta janela encontram-se informações referentes ao: a) identificador da coleção; b) título da coleção; c) uma breve descrição da coleção; d) botão para acesso às informações do responsável pelos dados; e) um ícone que ao ser acionado leva ao relatório exibido na Figura 12.

Human Protein Atlas FAIRsharing record against all MIs



Summary:

Description: FAIR Metrics Evaluation: Human Protein Atlas FAIRsharing record against all MIs; Tested identifier: 10.25504/FAIRsharing.j0t0pe; generated by <https://orcid.org/0000-0001-6960-357X>
Resource: 10.25504/FAIRsharing.j0t0pe
Collection: 6
Observations: Ran 22 tests (14 succeeded, 8 failed).
JSON response: https://w3id.org/FAIR_Evaluator/evaluations/1.json

Tests passing and failing



Figura 12. Relatório detalhado da avaliação de recursos digitais.

Fonte: Human... (2021).

Na parte superior esquerda do relatório, é apresentado um sumário da coleção de dados. A direita, um gráfico é exibido informando as quantidades de Testes de Conformidade que foram aprovados e rejeitados. Nesse caso, 22 Testes foram realizados, 14 tiveram êxito e 8 foram rejeitados. Os Testes aprovados encontram-se assinalados em verde (+) e, os rejeitados, marcados em vermelho (-). Acionando o botão +, do primeiro Teste, a Figura 13 é apresentada.

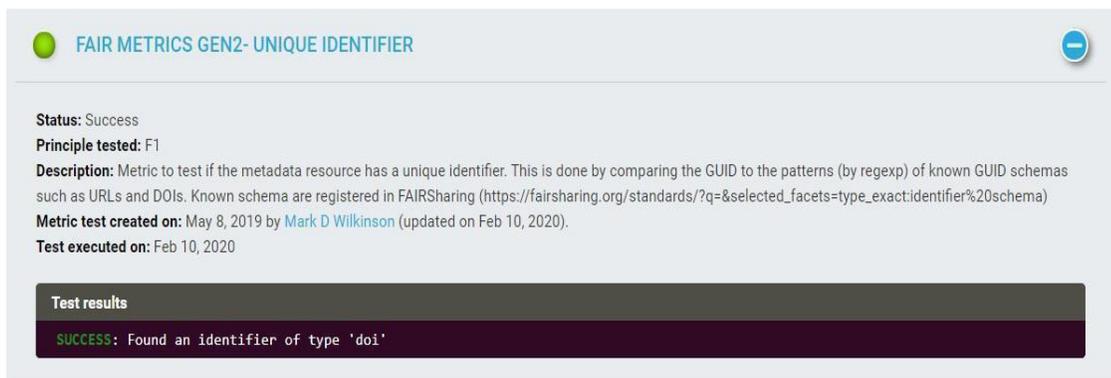


Figura 13. Síntese do Teste de Conformidade positivo.

Fonte: FAIR metrics... (2021).

A Figura 14 exibe o resultado de um Teste de Conformidade rejeitado. O nome do Teste é apresentado no alto da janela, seguido de um resumo e, por fim, são relacionadas informações decorrentes dos processos executados. De posse dessas informações, o usuário terá condições de ajustar as descrições de seu recurso digital para que ele se torne FAIR.

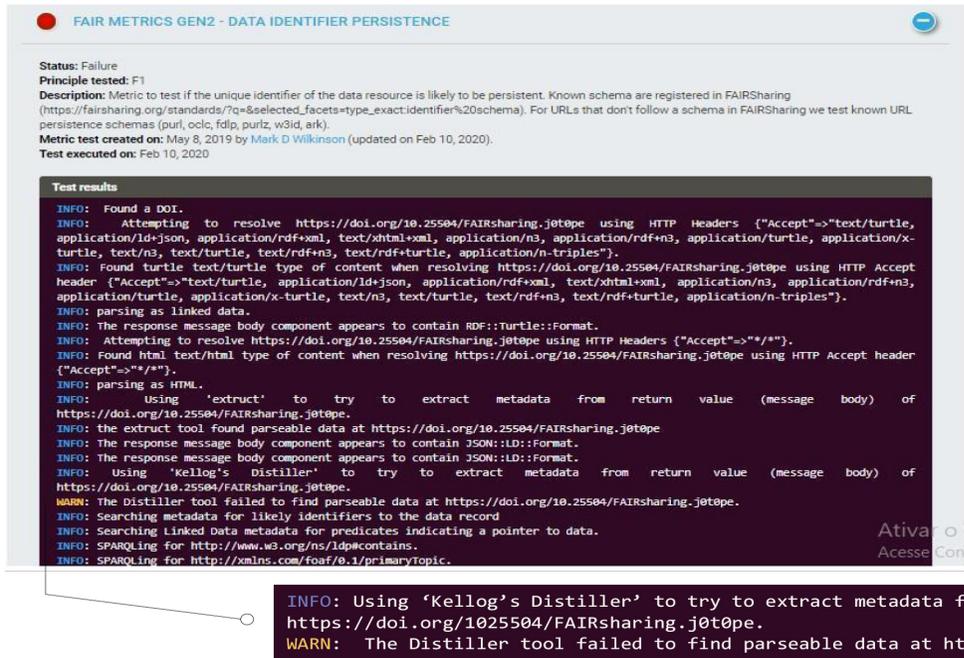


Figura 14. Síntese do Teste de Conformidade negativo.

Fonte: FAIR metrics... (2021).

Toolkit FAIRshake

O *toolkit* FAIRshake (Clarke et al., 2019a, 2019b) foi concebido com objetivo de permitir a construção de projetos registrando métricas e rubricas para a avaliação do grau FAIRness de diferentes recursos digitais, pertencentes a diferentes comunidades (grupos). Rubricas são coleções de métricas FAIR estabelecidas para verificar o FAIRness. As avaliações dos recursos digitais ocorrem de forma manual, semi e totalmente automatizadas. O *toolkit* FAIRshake fornece:

- Catálogo de objetos digitais – pode ser formado por conjuntos de dados, APIs, *workflow* etc, cada um com identidade única.
- Catálogo de projetos – um projeto pode registrar conjuntos de objetos digitais, agrupados para fins analíticos e de localização. Todos os objetos fazem parte de um processo específico. Se a intenção é a automação das avaliações FAIR, faz sentido esse agrupamento, dando a possibilidade de comparação entre as avaliações dentro do projeto.
- Catálogo de métricas – trata-se de qualquer critério único do FAIR ou uma questão de conformidade FAIR, que muitas vezes pode ser respondida como sim/não/percentagem. Na avaliação manual, a resposta é no formato sim/não, enquanto que nos demais tipos de avaliação a resposta é expressada em percentagem. A avaliação automática faz uso de API.
- Catálogo de rubricas – conjuntos/pacotes de métricas FAIR destinadas a serem respondidas em conjunto, por exemplo, uma API FAIR deve satisfazer várias métricas independentes já registradas no FAIRshake, podendo fazer parte de uma ou mais métricas.
- Facilidades de avaliações FAIR – qualquer projeto digital pode ser avaliado por uma determinada rubrica, manualmente por meio do site FAIRshake ou automaticamente ativando uma API.

- Agregações de avaliações – a ferramenta fornece uma Insígnia (Mostrador) FAIR, que fornece uma visão das avaliações médias de um determinado objeto digital, projeto ou rubrica. O Mostrador utiliza um gradiente de cores que indica o quanto o recurso digital avaliado atende às métricas FAIR. Os resultados podem ser visualizados na forma de um relatório com gráficos de estatísticas resumidas.

O diagrama abaixo, Figura 15, apresenta o *workflow* do FAIRshake.

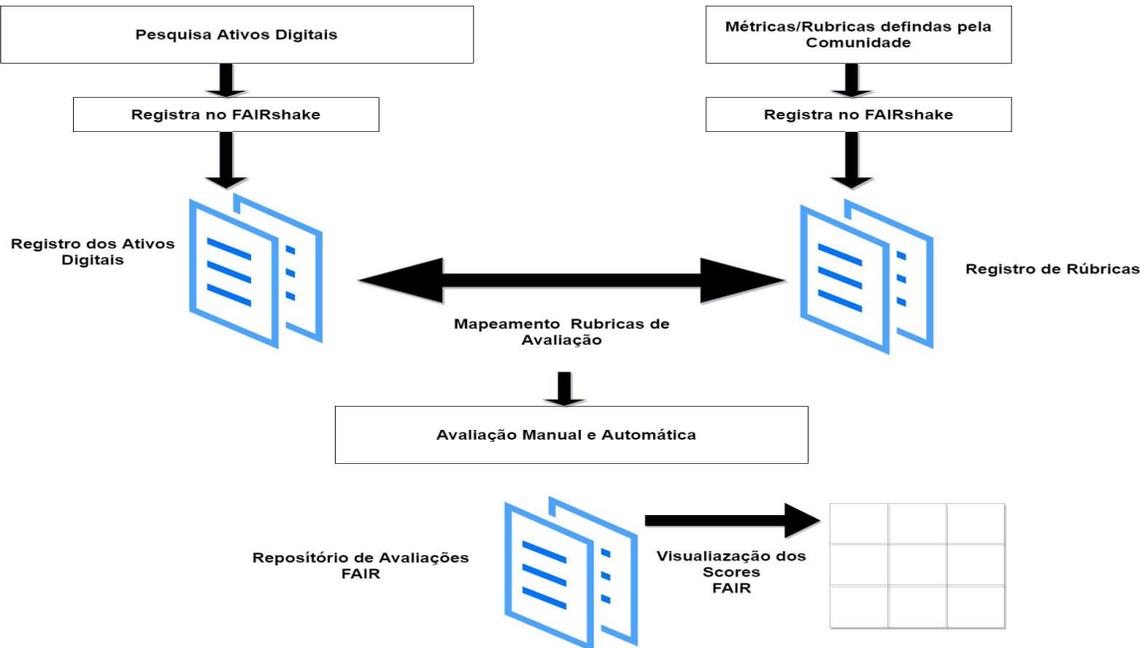


Figura 15. *Workflow* FAIRshake.

Fonte: Adaptado de Clarke et al. (2019a).

Outros dois recursos encontrados no FAIRshake são: a) *Bookmarklet* – usuários podem instalar este recurso que habilita avaliações FAIR de recursos digitais listados em qualquer site; e b) *Browser extension* – usuários podem instalar em um *browser* para habilitar avaliações FAIR de objetos digitais relacionados em qualquer site.

Os resultados das avaliações são exibidos no Mostrador, visualizado no site do FAIRshake ou incorporado em sites de hospedagem. Como um mesmo recurso pode ser avaliado por diferentes rubricas, o Mostrador dinamicamente se expande para ajustar todas as avaliações, como mostrado na Figura 16.

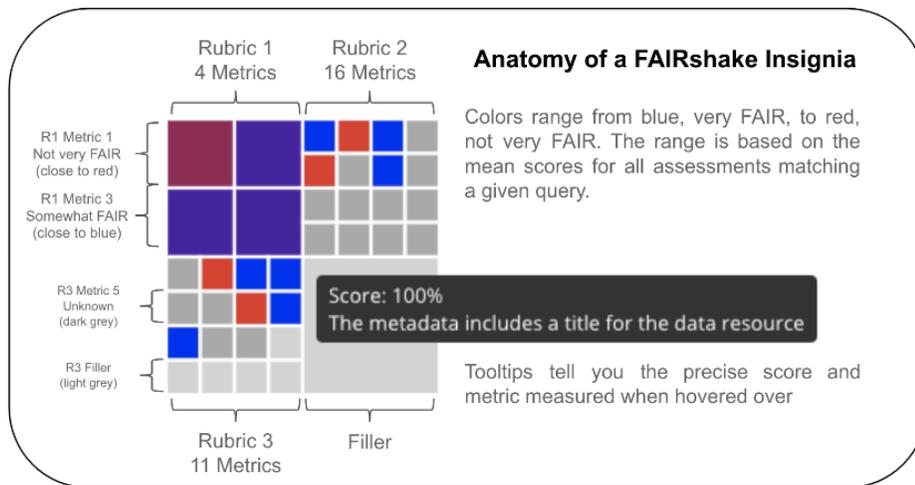


Figura 16. Exemplo de Mostradores FAIRshake reunidos.

Fonte: Clarke (2021).

Os resultados exibidos no Mostrador não têm por objetivo o ranqueamento das métricas e, sim, fornecer informações que permitam com que o recurso digital, com score FAIRness baixo, possa ser trabalhado para elevá-lo. O grau de satisfação de atendimento de uma métrica pode ser representado como percentagem ou um número entre 0 e 1. O aumento do score, que representa o grau de satisfação, se dá por meio da melhoria das métricas, que envolve um maior aprendizado sobre a cobertura da métrica. Este aprendizado envolve os aspectos: a) teórico (o que a métrica expressa) e o concreto (como o recurso digital foi avaliado). Além do Mostrador, o FAIRshake apresenta outras formas de externar resultados de suas avaliações, Figura 17.



Figura 17. Formas de apresentação de resultados da avaliação de recursos digitais.

Fonte: Adaptado de Clarke et al. (2019b).

A seguir é apresentado o funcionamento do FAIRshake por meio de um exemplo descrito em Clarke (2021). Este, apesar de simples, orienta como proceder para o desenvolvimento de avaliações mais completas. Maiores detalhes encontram-se em Clarke (2021). Para a utilização da ferramenta FAIRshake é necessário a realização de um cadastro no site <https://fairshake.cloud/>.

O exemplo foca na avaliação manual de um *dataset* já cadastrado no *toolkit* FAIRshake. A rubrica FAIR metrics by fairmetrics.org (25) foi acionada, assim como, o projeto FAIRshake Testing (55). Os dados utilizados encontram-se registrados em Common Fund Data Ecosystem - CFDE (About..., 2021).

Após entrar no sistema, o usuário se depara com a janela exibida na Figura 17. No mecanismo de busca (Search) foi inserido a URL (<http://lincportal.ccs.miami.edu/datasets/view/LDS-1293>) para localização do recurso digital a ser avaliado, vide Figura 18.



Figura 18. Localização dos conjuntos de dados em LDS-193.

Fonte: Common Fund Data Ecosystem Documentation (2021a).

Uma vez localizado o arquivo de dados desejado (L1000 Dataset-small molecule CRISPR perturbagens, LINCS Phase 2 (December 2015), Figura 19, o botão Assess (Avaliação) foi acionado.

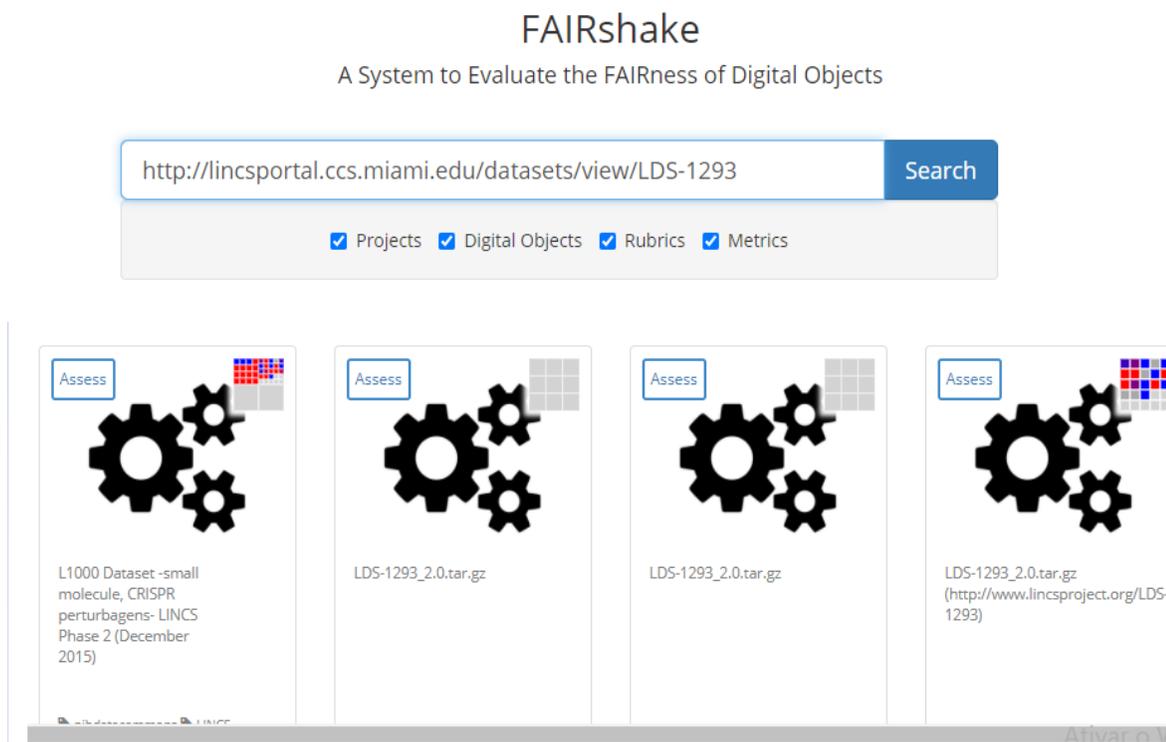


Figura 19. Seleção do arquivo de dados para avaliação.

Fonte: Common Fund Data Ecosystem Documentation (2021b).

Após a avaliação ser ativada, é apresentada a janela exibida na Figura 20. Através dela foi possível definir a rubrica, o projeto de avaliação e o objeto alvo a ser analisado. Neste caso, a ferramenta destaca de forma automática, as informações referentes ao projeto de avaliação, definido pelo CFDE.

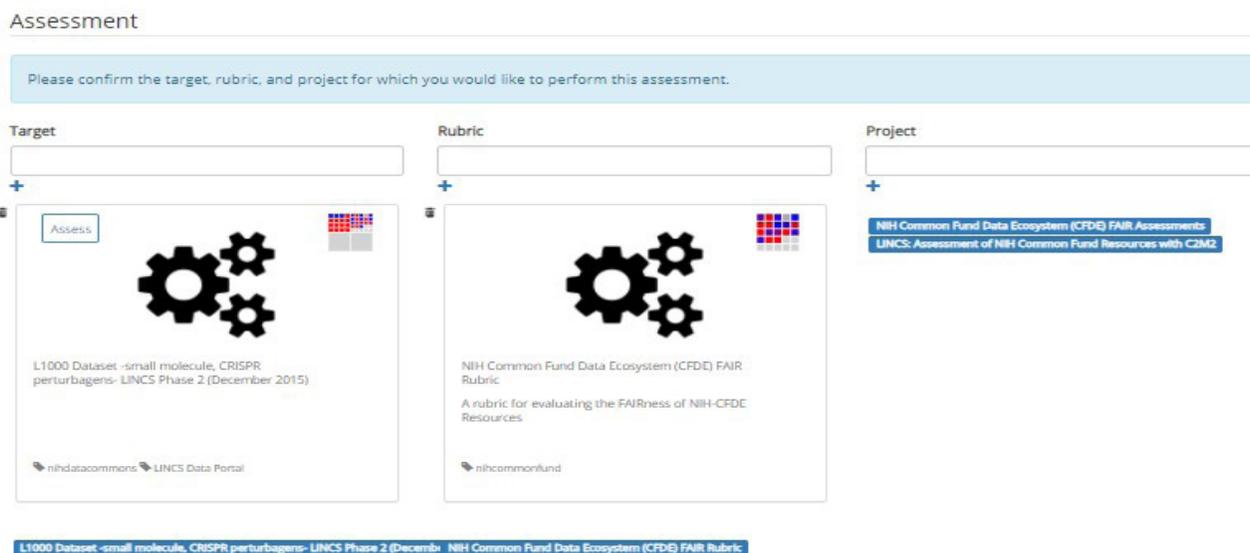


Figura 20. Definição dos parâmetros para avaliação.

Fonte: Common Fund Data Ecosystem (2021a).

Por se tratar de um exemplo de uso do FAIRshake, optou-se por utilizar um projeto teste, digitando na caixa Project “FAIRshake Testing” e acionando em seguida o botão *Confirm*. O não fornecimento dessa informação indicará que a avaliação estará desvinculada de qualquer projeto.

Na caixa Rubric, foi inserido FAIR metrics by fairmetrics.org e, na sequência, acionado o botão *Confirm*. Logo após, o sistema apresenta o formulário, Figura 21, com as 16 métricas disponíveis para uso.

Assessment of L1000 Dataset -small molecule, CRISPR perturbagens- LINCS Phase 2 (December 2015) with FAIR metrics by fairmetrics.org



Globally unique identifier

Provide an URL to a registered scheme that defines the globally-unique structure of the identifier(s).

Yes
 No

This field is required.



Persistent identifier

Provide a URL to a document that defines the policy (data management plan) with respect to long term support for persisting the identifier.

Yes
 No

This field is required.



Machine-readable metadata

Provide the URL to a document that contains machine-readable metadata for the digital resource.

Yes
 No

This field is required.

Figura 21 Parametrização para verificação das métricas.

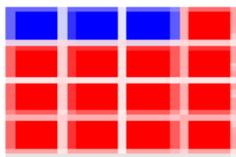
Fonte: Common Fund Data Ecosystem (2021a).

Neste exemplo, apenas as três primeiras métricas foram consideradas, sendo confirmadas através da resposta *Yes* e, para cada uma delas, fornecido o(s) endereço(s) para sua verificação. As 13 métricas restantes foram marcadas como *No*, em virtude da falta de informações para referenciá-las. Na Tabela 5, as métricas utilizadas e seu(s) respectivo(s) endereço(s) são apresentadas.

Tabela 5. Métricas e seus respectivos endereços Web.

Métricas	Endereço Web
Globally unique identifier	http://fairsharing.org/bsg-s001186
Persistent identifier	https://lincsportal.ccs.miami.edu/datasets/terms https://registry.identifiers.org/registry/lincs.data
Machine-readable metadata	https://lincsportal.ccs.miami.edu/apps/#/ https://smart-api.info/ui/1ad2cba40cb25cd70d00aa8fba9cfaf3

As definições acima podem ser salvas, publicadas ou removidas. Uma vez publicadas, a avaliação e as respostas obtidas estarão associadas ao objeto digital avaliado. As informações resultantes serão utilizadas na confecção do Mostrador, que após ser examinado indicará quais métricas foram atendidas para o recurso digital selecionado. A Figura 22 exibe o Mostrador gerado neste exemplo. Os retângulos pintados em azul registram os resultados das métricas avaliadas.

**Figura 22.** Mostrador resultante do processo de avaliação realizada.

Fonte: Common Fund Data Ecosystem (2021a).

Todas as métricas verificadas foram contempladas 100%, o que significa que o recurso digital apresentou o Indicador Único Global, o Indicador de Persistência e os metadados reconhecidos por máquina. Para outras 13 métricas os resultados foram de 0% em virtude delas não terem sido avaliadas. Quanto mais aprofundado for o nível de avaliação, maior será a quantidade de respostas emitida pelo sistema. Isso ocorrerá se os dados básicos para a parametrização do processo de avaliação forem disponibilizados.

Considerações finais

O processo de obtenção do grau FAIRness, para os recursos digitais, não é uma tarefa que se dá de forma isolada. Para o atingimento de grau elevado, é desejável um plano de gestão de dados bem estruturado, podendo este ser construído com auxílio de um sistema inteligente (Pergl et al., 2019). Além disso, os dados devem apresentar um bom nível de qualidade. Também se faz necessário ter conhecimento dos IM e seus respectivos Testes de Conformidade (Wilkinson et al., 2019), para verificar se eles estão adequados às necessidades. Caso não atendam, novos IM e Testes de Conformidades poderão ser estabelecidos. Em relação aos métodos para definição do grau FAIRness, alguns deles foram apresentados, no entanto, outros métodos poderão ser criados visando aperfeiçoar os já existentes.

A ferramenta *toolkit* FAIRshake, dentre as relacionadas, mostrou-se bastante completa para avaliação de recursos digitais armazenados em repositórios. Trata-se de uma ferramenta de código livre e encontra-se em <https://github.com/MaayanLab/FAIRshake>. A avaliação dos recursos digital pode ser executada na forma manual, semi e automática. Diversos tipos de relatórios são emitidos fornecendo informações que podem ser utilizadas para a elevação do grau FAIRness. Para utilizar o FAIRshake é necessário organizar os dados em um repositório, seguindo de preferência os

princípios descritos na definição do ciclo de vida de dados, como apresentado em DataONE (2021), configurá-lo adequadamente para acesso aos armazenados no repositório e para ativação dos re-cursos necessários para o cálculo do FAIRness.

Em FAIRassist.org (FAIRassist.org..., 2019), diversas ferramentas para automação do processo de obtenção do grau de FAIRness encontram-se relacionadas, elas encontram-se em diferentes estágios de maturação. Com base nas características das ferramentas apresentadas, vemos a possibilidade de uso delas como um pacote de apoio ao processo de Gestão de Dados de Pesquisa, envolvendo os aspectos relacionados à compreensão dos princípios FAIR, dos IMs e seus respectivos Testes de Conformidade, e obtenção do grau de FAIRness.

Esse trabalho buscou a construção de uma base de conhecimento visando dotar a Embrapa de meios cada vez mais aperfeiçoados, para a recuperação e a reutilização dos recursos digitais, registrados nos repositórios de dados da Empresa, permitindo a geração de novos conhecimentos e avanços na agricultura brasileira. As próximas etapas, na linha deste trabalho, será selecionar uma das ferramentas apresentadas e aplicá-la sobre o Redape, cuja implantação está em curso. Inicialmente, está aplicação poder-se-á avaliar quão FAIR são os datasets catalogados no Redape.

Referências

ABOUT CFDE. In: COMMON Fund Data Ecosystem Documentation. Disponível em: <<https://docs.nih-cfde.org/en/latest/about/about/>>. Acesso em: 26 nov. 2021.

AZEVEDO, R. de M.; DUMONTIER, M. Considerations for conduction and interpretation of FAIRness evaluations. **Data Intelligence**, v. 2, n. 1-2, p. 285–292, Winter-Springer 2020. DOI: 10.1162/dint_a_00051.

CLARKE, D. J. B. **Evaluating FAIRness with FAIRshake**: a tutorial that demonstrates how to use FAIRshake to perform FAIR evaluations of DATS serialized metadata in the context of the CFDE. Disponível em: <<https://docs.nih-cfde.org/en/latest/the-fair-cookbook/content/recipes/Compliance/fairshake/>>. Acesso em: 14 dez. 2021.

CLARKE, D. J. B.; WANG, L.; JONES, A.; WOJCIECHOWICZ, M. L.; TORRE, D.; JAGODNIK, K. M.; JENKINS, S. L.; MCQUILTON, P.; FLAMHOLZ, Z.; SILVERSTEIN, M. C.; SCHILDER, B. M.; ROBASKY, K.; CASTILLO, C.; IDASZAK, R.; AHALT, S. C.; WILLIAMS, J.; SCHURER, S.; COOPER, D. J.; AZEVEDO, R. de M.; KLENK, J. A.; HAENDEL, M. A.; NEDZEL, J.; AVILLACH, P.; SHIMOYAMA, M. E.; HARRIS, R. M.; GAMBLE, M.; POTEN, R.; CHARBONNEAU, A. L.; LARKIN, J.; BROWN, C. T.; BONAZZI, V. R.; DUMONTIER, M. J.; SANSONE, S. A.; MA'AYAN, A. FAIRshake: toolkit to evaluate the FAIRness of research digital resources. **Cell Systems**, v. 9, n. 5, p. 417–421, Nov. 2019a. DOI: 10.1016/j.cels.2019.09.011.

CLARKE, D. J. B.; WANG, L.; JONES, A.; WOJCIECHOWICZ, M. L.; TORRE, D.; JAGODNIK, K. M.; JENKINS, S. L.; MCQUILTON, P.; FLAMHOLZ, Z.; SILVERSTEIN, M. C.; SCHILDER, B. M.; ROBASKY, K.; CASTILLO, C.; IDASZAK, R.; AHALT, S. C.; WILLIAMS, J.; SCHURER, S.; COOPER, D. J.; AZEVEDO, R. de M.; KLENK, J. A.; HAENDEL, M. A.; NEDZEL, J.; AVILLACH, P.; SHIMOYAMA, M. E.; HARRIS, R. M.; GAMBLE, M.; POTEN, R.; CHARBONNEAU, A. L.; LARKIN, J.; BROWN, C. T.; BONAZZI, V. R.; DUMONTIER, M. J.; SANSONE, S. A.; MA'AYAN, A. **FAIRshake**: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources, 2019b. Disponível em: <<https://www.biorxiv.org/content/10.1101/657676v1>>. Acesso em: 18 jun. 2021.

COMMON FUND DATA ECOSYSTEM. Coordination Center. **FAIRshake**: a system to evaluate the FAIRness of digital objects. Disponível em: <<https://fairshake.cloud/>>. Acesso em: 26 nov. 2021a.

COMMON FUND DATA ECOSYSTEM. Coordination Center. **FAIRshake**: a system to evaluate the FAIRness of digital objects. Disponível em: <<https://fairshake.cloud/?q=http%3A%2F%2Flincsportal.ccs.miami.edu%2Fdatasets%2Fview%2FLDS-1293&projects=1&digitalobjects=1&rubrics=1&metrics=1>>. Acesso em: 26 nov. 2021b.

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH. **FAIR metrics**. Disponível em: <<https://gardian.bigdata.cgiar.org/metrics.php#!/>>. Acesso em: 15 dez. 2021a.

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH. **Gardian FAIR metrics**. Disponível em: <https://gardian.bigdata.cgiar.org/files/GARDIAN_FAIR_metrics_guide.pdf>. Acesso em: 14 jun. 2021b.

DATA ARCHIVING AND NETWORKED SERVICES. **FAIR aware**: your first step towards your FAIR data(set). Disponível em: <<https://fairaware.dans.knaw.nl/>>. Acesso em: 26 nov. 2021.

DATAONE. **Data life cycle**. Disponível em: <<https://old.dataone.org/data-life-cycle>>. Acesso em: 29 nov. 2021.

FAIRASSIST.ORG: help you discover resources to measure and improve FAIRness. 2019. Disponível em: <<https://fairassist.org#!/>>. Acesso em: 29 mar. 2021.

FAIR EVALUATION SERVICES. **Resources and guidelines to assess the FAIRness of digital resources**. Disponível em: <[https://fairsharing.github.io/FAIR-Evaluator-FrontEnd#!/#%2F!](https://fairsharing.github.io/FAIR-Evaluator-FrontEnd#!/#%2F!>)>. Acesso em: 26 nov. 2021.

FAIR maturity indicator – evaluations. In: FAIR EVALUATION SERVICES. **Resources and guidelines to assess the FAIRness of digital resources**. Disponível em: <<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd#!/evaluations>>. Acesso em: 26 nov. 2021.

FAIR metrics GEN 2 – unique identifier. In: FAIR EVALUATION SERVICES. **Resources and guidelines to assess the FAIRness of digital resources**. Disponível em: <<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd#!/evaluations/1>>. Acesso em: 26 nov. 2021.

GOFAIR. **FAIR principles**. [Hamburg, 2018]. Disponível em: <<https://www.go-fair.org/fair-principles/>>. Acesso em: 25 mar. 2021.

GRÁFICOS de radar. In: IBM. **Cognos Analytics 11.0x**. 2021. Disponível em: <https://www.ibm.com/docs/pt-br/cognos-analytics/11.0.0?topic=SSEP7J_11.0.0/com.ibm.swg.ba.cognos.ug_cr_pps.doc/c_as_ti_charts_radr.html>. Acesso em: 26 nov. 2021.

HEYDEBRECK, D.; GANSKE, A.; KRAFT, A.; KAISER, A. **The data maturity indicator concept**. 2020. Disponível em: <<https://av.tib.eu/media/47696>>. Acesso em: 14 dez. 2021.

HUMAN protein atlas FAIRsharing record against all MIs. In: FAIR EVALUATION SERVICES. **Resources and guidelines to assess the FAIRness of digital resources**. Disponível em: <<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd#!/evaluations/1>>. Acesso em: 26 nov. 2021.

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, n. 1, p. 1-18, 2019. DOI: 10.5334/dsj-2019-059.

RESEARCH DATA ALLIANCE. **FAIR data maturity model**: specification and guidelines. 2020. Disponível em: <<https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0>>. Acesso em: 24 nov. 2021.

THE ATMODAT project: atmospheric model data: data quality, curation criteria and DOI branding. Disponível em: <<https://www.atmodat.de/>>. Acesso em: 15 dez. 2021.

WILKINSON, M. D.; DUMONTIER, M.; AALBERSBERG, I. J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J. W.; SANTOS, L. B. da S.; BOURNE, P. E.; BOUWMAN, J.; BROOKES, A. J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C. T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A. J. G.; GROTH, P.; GOBLE, C.; GRETHE, J. S.; HERINGA, J.; 't HOEN, P. A. C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S. J.; MARTONE, M. E.; MONS, A.; PACKER, A. L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAIK, R. van; SANSONE, S. A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M. A.; THOMPSON, M.; VAN DER LEI, J.; VAN MULLIGEN, E.; VELTEROP, J.; WAAGMEESTER, A.; WITTENBURG, P.; WOLSTENCROFT, K.; ZHAO, J.; MONS, B. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 13, p. 1-8, Mar. 2016. DOI: 10.1038/sdata.2016.18.

WILKINSON, M. D.; DUMONTIER, M.; SANSONE, S. A.; SANTOS, L. O. B. da S.; PRIETO, M.; BATISTA, D.; MCQUILTON, P.; KUHN, T.; ROCCA-SERRA, P.; CROSAS, M.; SCHULTES, E. Evaluating FAIR maturity through a scalable, automated, community-governed framework. **Scientific Data**, v. 6, n. 174, p. 1–12, 2019. DOI: 10.1038/s41597-019-0184-5.

