

Biblioteconomia
Representação
Interoperabilidade
E-Science
Tecnologia



Repositório FAIR

Luana Sales -PGCI IBICT-UFRJ

luanafsales@gmail.com

Luís Sayão – CIN - CNEN



BY

REVISITANDO O CONCEITO DE REPOSITÓRIOS

BANCOS DE DADOS CIENTÍFICOS DIGITAIS QUE GARANTEM O ACESSO A RESULTADOS DE PESQUISA AGORA E NO FUTURO, TENDO COMO PERSPECTIVA PRIMORDIAL O ACESSO ABERTO AO QUE É CONSIDERADO TAMBÉM UM PATRIMÔNIO DIGITAL DA HUMANIDADE (SURF FOUNDATION, 2013)



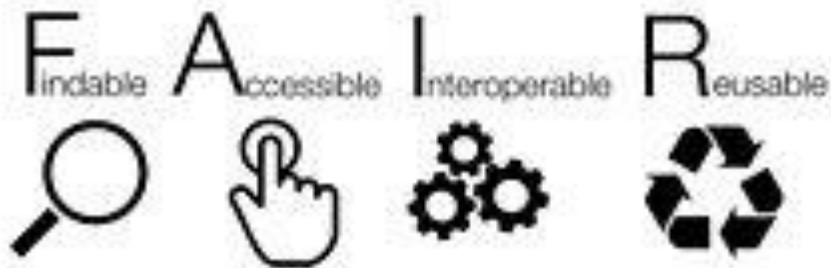
Qual a finalidade dos repositórios ?



COMO DEFINIR UM BOM REPOSITÓRIO?

- UM BOM REPOSITÓRIO É AQUELE QUE:

- É localizado
- É acessado
- É interoperável
- Seus dados são reusáveis



PRINCÍPIOS

FAIR





- Subutilizados por diversos motivos

- SÃO INVISÍVEIS
- SÃO MUITO GENERICOS
- NÃO TEM CONEXAO COM A COMUNIDADE
- NÃO POSSUI SERVICOS
- SÃO COMPULSORIOS
- PROBLEMAS TECNOLÓGICOS
- DUPLICAÇÃO DE ESFORÇOS
- NÃO DÃO FEEDBACK AO USUÁRIO
- NÃO ADOTAM PADRÕES

REPOSITÓRIO DE DADOS



CEMI⁺TÉRIO DE DADOS ?

+50%

DOS ACHADOS
NÃO FORAM
PUBLICADOS (FERGUNSON ET AL, 2014)



O COMPARTILHAMENTO PODE REVELAR VALORES IMPORTANTES OCULTOS NESSES DADOS

Afinal, o que é FAIR?

Princípios FAIR

Criados no *Jointly designing a data FAIRPORT Conference*, por especialistas de diversas áreas do conhecimento interessados no reuso de dados, no contexto da e-Science em 2014

Publicados em 2016

Lorentz center
Jointly Designing a Data FAIRPORT
Workshop: 13 – 16 January 2014, Leiden, the Netherlands

Scientific Organizations

- Scott Lusher, MASC, Amsterdam
- Ewald Moss, Leiden UMC

Topics

- Towards a Modular Blueprint: 7-year plan of a Safe and Fair Data Stewardship, Tracking and Routing Environment
- A Public-Private Partnership to Ensure Long Term Solutions for Data in the eScience Era.

Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly enforce a concise and

Lorentz center
www.lorentzcenter.nl

NEW **SCIENTIFIC DATA**

The FAIR Guiding Principles for scientific data management and stewardship

Van de Weide et al. | Published: 23 March 2016

Download PDF

754 Citations 1337 Downloads Article

Associated Content

Collection: Metadata Quality

Sections

Abstract

Comments

Additional Information

References

Acknowledgments



Os princípios FAIR servem para:

- Dados de pesquisa
- Dados de outras naturezas
- Objetos Digitais
- Repositórios
- Outros Bancos de Dados
- Serviços de Internet



DADOS FAIR

- Dados FAIR são aqueles que adotam os quinze princípios criados para as melhores práticas de gestão, compartilhamento e reuso respeitando todas as questões éticas, legais e restrições contratuais.

REPOSITÓRIOS FAIR

- Repositórios FAIR são aqueles que se adaptam aos quinze princípios FAIR e abrigam dados, dados de pesquisa ou outros objetos digitais a partir de práticas de gestão que tornam seus objetos localizáveis, acessáveis, interoperáveis e reusáveis respeitando as questões éticas, legais de compartilhamento



REPOSITÓRIO LOCALIZADO

- Tem que estar visível
- Tem que ser indexado por metabuscadores
- Tem que ter um trabalho forte de divulgação
- Seus Objetos precisam ter identificação única

Dados de pesquisa
precisam estar visíveis

Dados de pesquisa se tornam
visíveis a partir do momento que
são gerenciados, preservados e
disseminados por meio de
serviços criados sobre
plataformas

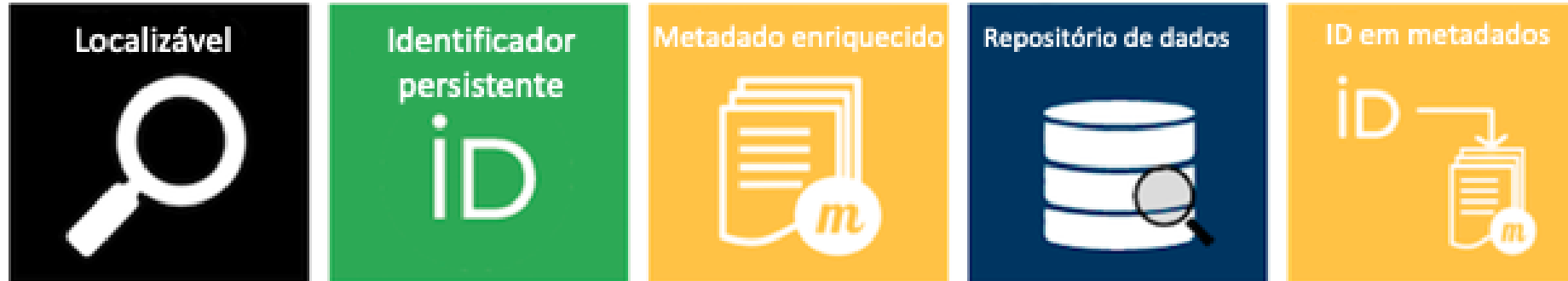




LOCALIZÁVEIS

- Os dados devem ser fáceis de serem encontrados.
- **Metadados** ricos devem estar disponíveis em uma ferramenta de busca online
- Devem ser associados a um identificador persistente (DOI)

LOCALIZÁVEL



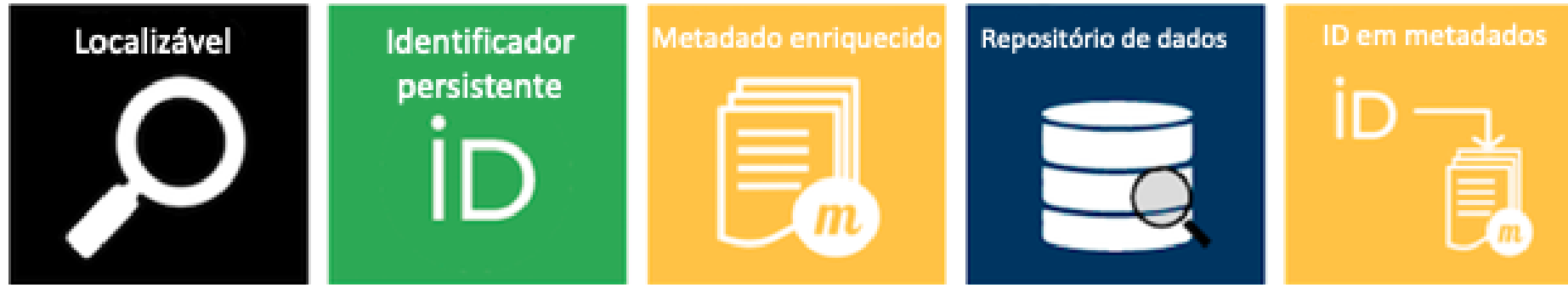
F1. (meta)dados devem ter identificadores globais, únicos e persistentes

Adotar identificador único persistente tanto para o conjunto de dados quanto para os metadados (ex: DOI, handle, orcid, PID)

F2. dados devem ser descritos utilizando metadados ricos (impacta diretamente R1)

O conjunto de dados deve ser descrito por metadados ricos o suficiente para que, uma vez indexados em um mecanismo de busca, possam ser encontrados mesmo sem o seu identificador único persistente

LOCALIZÁVEL



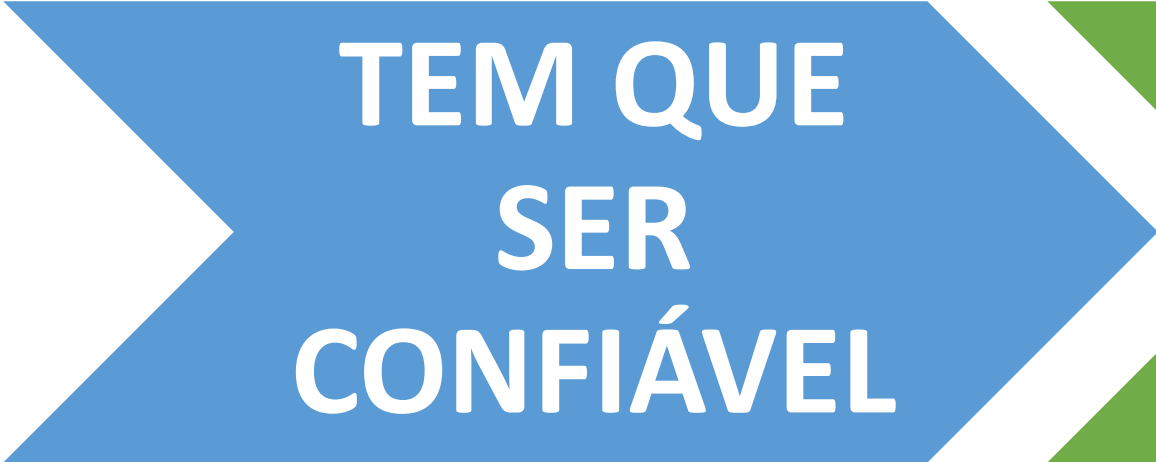
F3. (meta)dados devem incluir clara e explicitamente os identificadores dos dados que descrevem

Como não podemos prever que os dados e seus metadados estejam sempre juntos, a associação entre eles deve ocorrer pela inclusão do identificador persistente dos dados nos metadados.

F4. (meta)dados devem ser registrados ou indexados em mecanismos de busca

Para que os dados sejam encontrados, seus metadados devem ser indexados em mecanismos de busca (search engine), que possibilitem aos computadores e usuários encontrá-los com facilidade.

REPOSITÓRIO ACESSAVEL E ACESSADO



TEM QUE
SER
CONFIÁVEL



TEM QUE
TER
SERVIÇOS

ACESSÁVEL



Homens e máquinas devem ter acesso aos dados sob condições específicas ou restritas, quando for apropriado.

Tão aberto quanto possível e tão fechado quanto necessário

ACESSÁVEL



A1. (meta) dados devem ser recuperáveis pelos seus identificadores usando protocolo de comunicação padronizado

Com o identificador persistente do conjunto de dados e/ou de seus metadados, o usuário deverá recuperá-los mais facilmente por meio de protocolos de comunicação padronizados. (ex: HTTP ou FTP)

A1.1 o protocolo deve ser aberto, gratuito e universalmente implementável

Independente de licenciamento dos dados e dos metadados, o protocolo de comunicação usado para dar acesso a eles deve ser aberto, gratuito e passível de ser implementado por qualquer interessado. (ex: HTTP ou FTP)

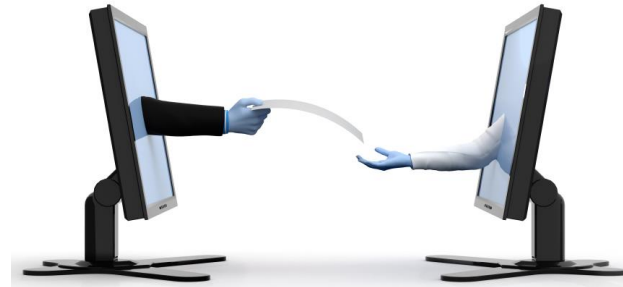


A1.2. o protocolo deve permitir procedimentos de autenticação e autorização, quando necessário

Dependendo das restrições de acesso aos dados e/ou metadados, um mecanismo de autenticação e autorização para o acesso deve ser liberado pelo protocolo de comunicação. (Ex: os repositórios confiáveis oferecem essa opção)

A2. metadados devem ser acessíveis, mesmo quando os dados não estiverem mais disponíveis.

É preciso existir um conjunto de estratégias de preservação para dados e metadados. Os metadados devem ser sempre acessíveis, possibilitando a criação de índices para o conjunto de dados atuais vigentes e aqueles não mais disponíveis.

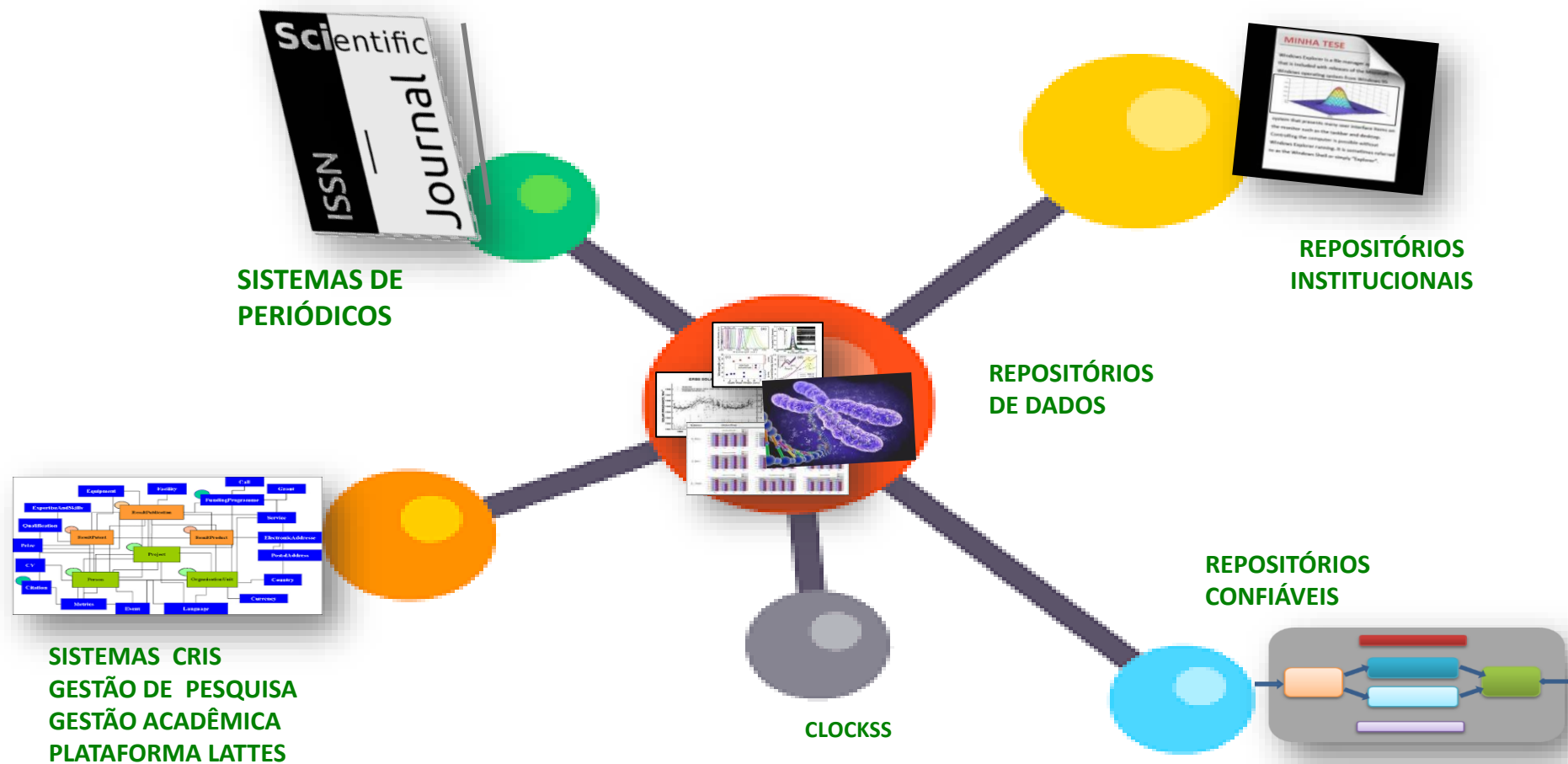


Os dados só atingirão seu potencial máximo de reuso, gerando inovação, se puderem transitar entre plataformas

REPOSITÓRIO INTEROPERÁVEL

- Usar padrões internacionais genéricos, sem esquecer dos modelos disciplinares
 - Metadados
 - Taxonomias
 - Ontologias

OUTRAS INTEROPERABILIDADES POSSÍVEIS



INTEROPERÁVEIS



Os dados e os metadados devem estar em conformidade com formatos e padrões reconhecidos para permitir que sejam combinados e trocados entre sistemas.

INTEROPERÁVEL



I1. (meta) dados devem ser representados por meio de uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento

Para que se possa representar dados e metadados devem ser adotadas linguagens de representação do conhecimento que sejam padronizadas, acessíveis e amplamente aplicáveis. (Ex: RDF, XML, DICOM, etc.)

I2. (meta) dados devem usar vocabulários de acordo com os princípios FAIR

Dados e metadados devem possuir referências a vocabulários e/ou ontologias que os descrevem. Devemos garantir que esses também sigam os princípios FAIR.

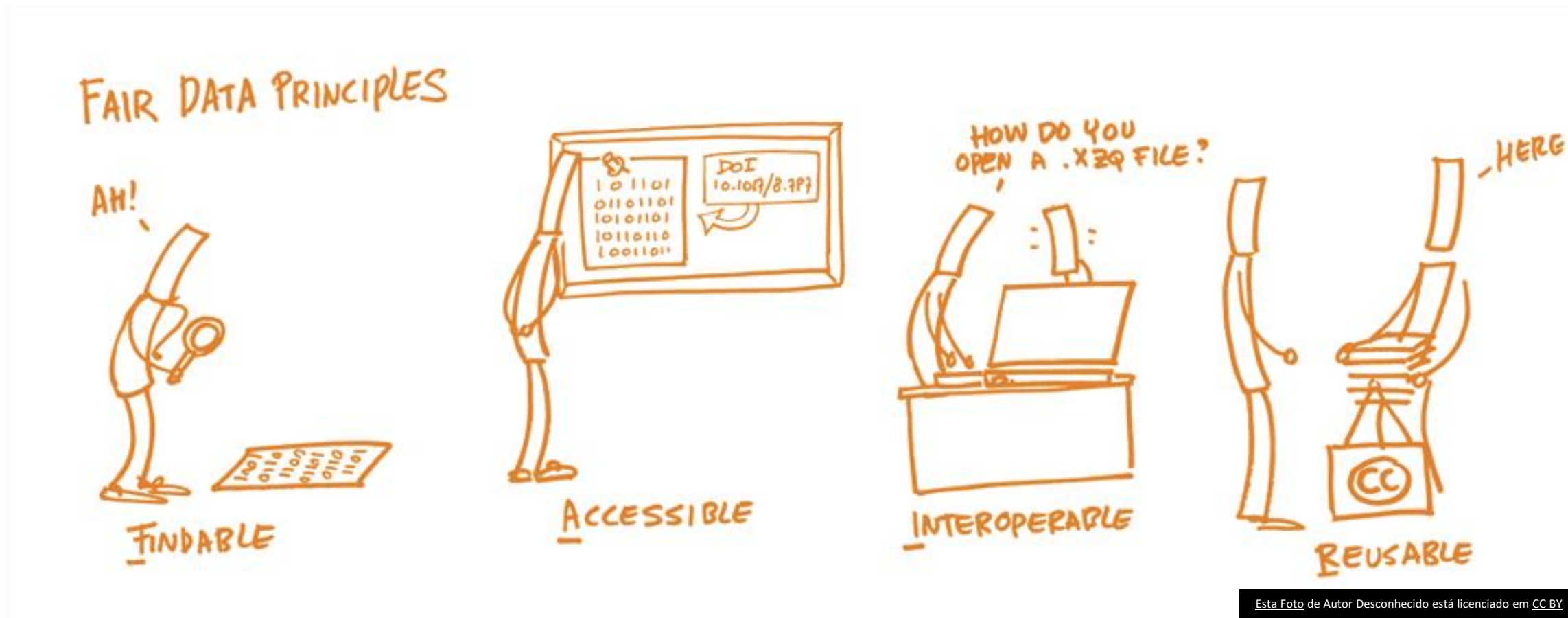
INTEROPERÁVEL



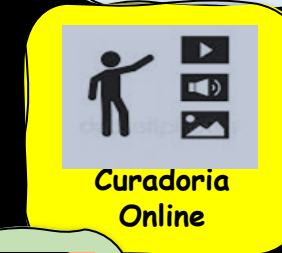
13. (meta) dados devem incluir referências qualificadas para outros (Meta) dados

É necessário referenciar o conjunto de dados, possibilitando que aqueles gerados a partir de outros conjuntos, sejam interligados. Assegurando a ligação semântica entre eles.

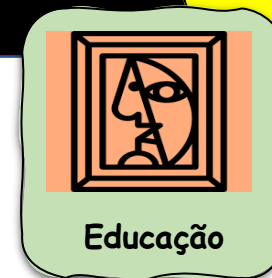
O CONTEÚDO DO REPOSITÓRIO TEM QUE SER PASSÍVEL DE SER REUSADO



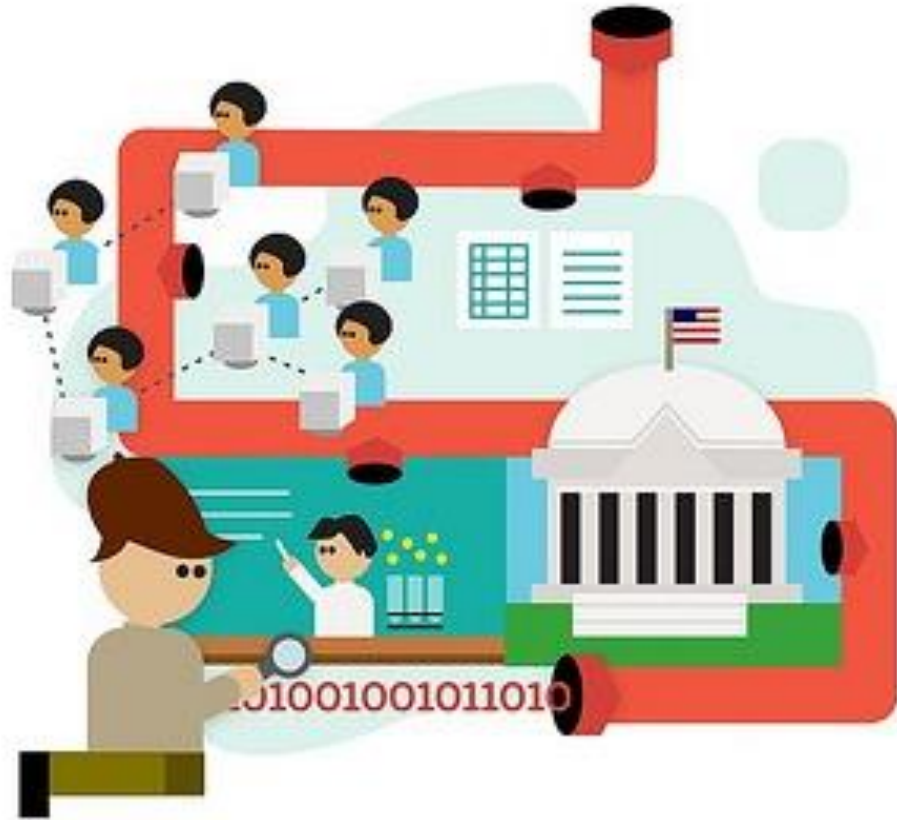
Esta Foto de Autor Desconhecido está licenciado em [CC BY](#)



POSSÍVEIS
CATEGORIAS DE
REUSOS



REUSÁVEIS



É necessário documentação apropriada para apoiar a interpretação e reutilização de dados.

Os dados devem estar em conformidade com as normas da comunidade e ter uma licença clara para que outros saibam quais tipos de reuso são permitidos.

REUSAVEL



R1. (meta) dados são descritos com uma pluralidade de atributos precisos e relevantes.

Prover metadados descritos com alto nível de detalhes que permita ao pesquisador avaliar a possibilidade do seu reuso bem como adequação às suas necessidades.

R1.1. (meta) dados devem ser disponibilizados com licenças de uso claras e acessíveis

É fundamental que o responsável pelos dados e metadados defina explicitamente quem pode ter acesso a eles, com que finalidade e sob quais condições. Essas informações são definidas por meio de licenças de uso.

REUSAVEL



R1.2. (meta) dados devem estar associados à sua proveniência

Especificar a proveniência (linhagem) dos dados é importante não só para que o pesquisador possa avaliar a utilidade dos dados ou metadados, mas também para que possa atribuir o devido crédito a quem produziu, manteve ou editou esses dados. Dentre as informações relativas à proveniência destacam-se:

- a) A linhagem dos dados, ou seja, o processo de obtenção dos dados (gerado ou coletado);
- b) Particularidades ou limitações sobre os dados que outros pesquisadores devem conhecer;
- c) Data da geração do conjunto de dados, condições de laboratório, quem preparou os dados, configurações de parâmetros, nome e versão do software utilizado;
- d) Explicitar se são dados brutos ou processados;
- e) A versão dos dados arquivados e/ou reutilizados deve ser claramente especificada e documentada.

REUSAVEL



R1.3. (meta) dados devem estar alinhados com padrões relevantes do seu domínio

Além de atender aos padrões específicos da área de cada comunidade deve-se dar atenção as boas práticas de arquivamento e compartilhamento específicos da área de pesquisa.



Fairificação de Repositórios

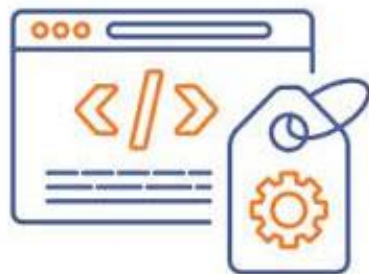
- Fairificar repositórios significa torná-los compatíveis com os princípios FAIR
- A maioria dos requisitos de localização e acessibilidade (F e A) aconteçam no âmbito dos metadados
- A interoperabilidade e o reuso (I e R), no nível dos dados.

FAIR é sobre...



ACIONALIDADE POR MÁQUINA

- Os seres humanos não são os únicos interlocutores críticos no ecossistema de dados, o **FAIR é principalmente para máquinas**;
- Escopo, escala e velocidade requisitada pelo nível de complexidade da ciência contemporânea;
- Os computadores devem ser capazes de acessar os dados de forma autônoma;
- Devem encontrar e usar dados e apoiar o reuso por humanos;
- Os “**stakeholders computacionais**” são exploradores que agem em nosso nome: agentes, programas de aplicação etc.



METADADOS

- A acionabilidade por máquina coloca em destaque **a importância dos metadados** que estão presentes nos 15 princípios.
- O objeto digital deve fornecer informações cada vez mais detalhada para um explorador computacional;
- O que é o objeto, contexto, estrutura e intenção, utilidade no contexto, licença, consentimento, nível de sensibilidade;

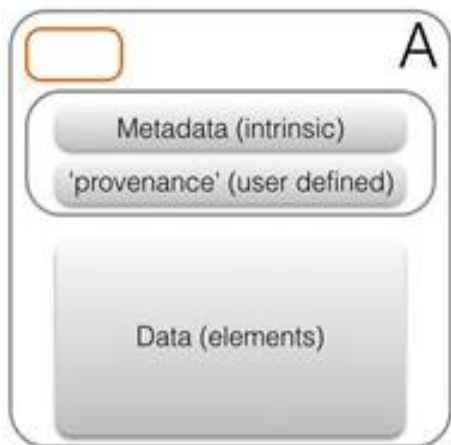


ACESSO SOB CONDIÇÕES BEM DEFINIDAS

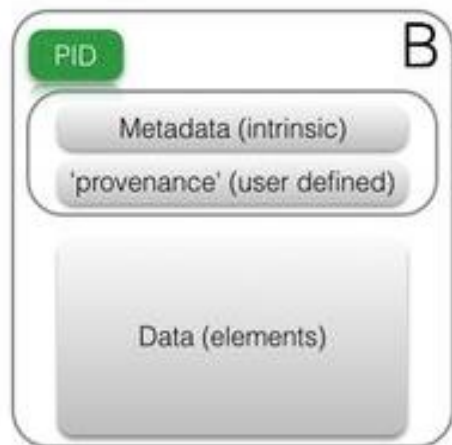
- **FAIR não é igual a aberto**;
- Há razões legítimas para blindar os dados e serviços gerados com fundos públicos que requerem medidas adicionais de autorização e autenticação tanto para humanos quanto para máquinas;
- FAIR não endereça questões e morais sobre a abertura dos dados: critério do custodiante.

NIVEIS DE FAIRNESS X ABERTURA DOS DADOS

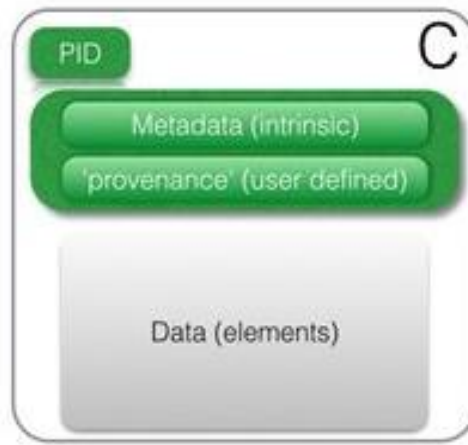
Re-useless data (80%)



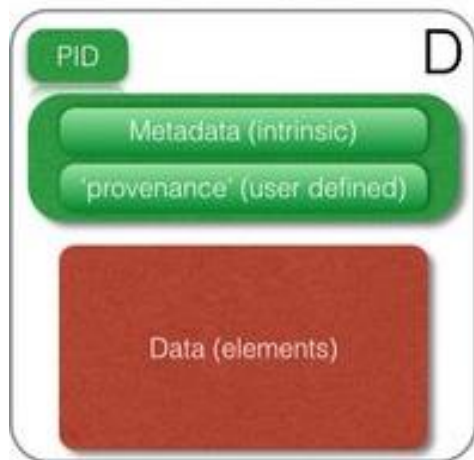
Findable



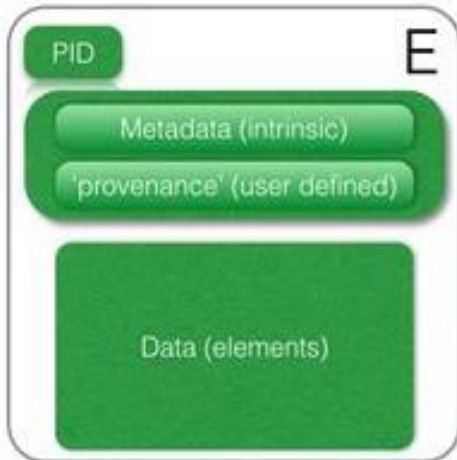
FAIR metadata



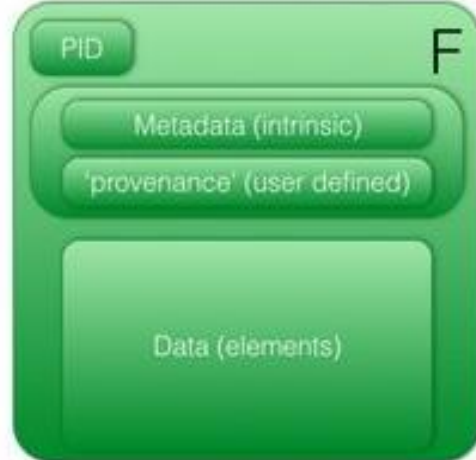
FAIR data-
restricted access



FAIR data-
Open Access



FAIR data-
Open Access/Functionally Linked



A

80% dos datasets está indisponível para o reuso;

B

Primeiro passo: PID – identificador persistente;

C

Metadados legíveis por máquina: intrínsecos e definidos pelo usuários (contexto e proveniência)

D

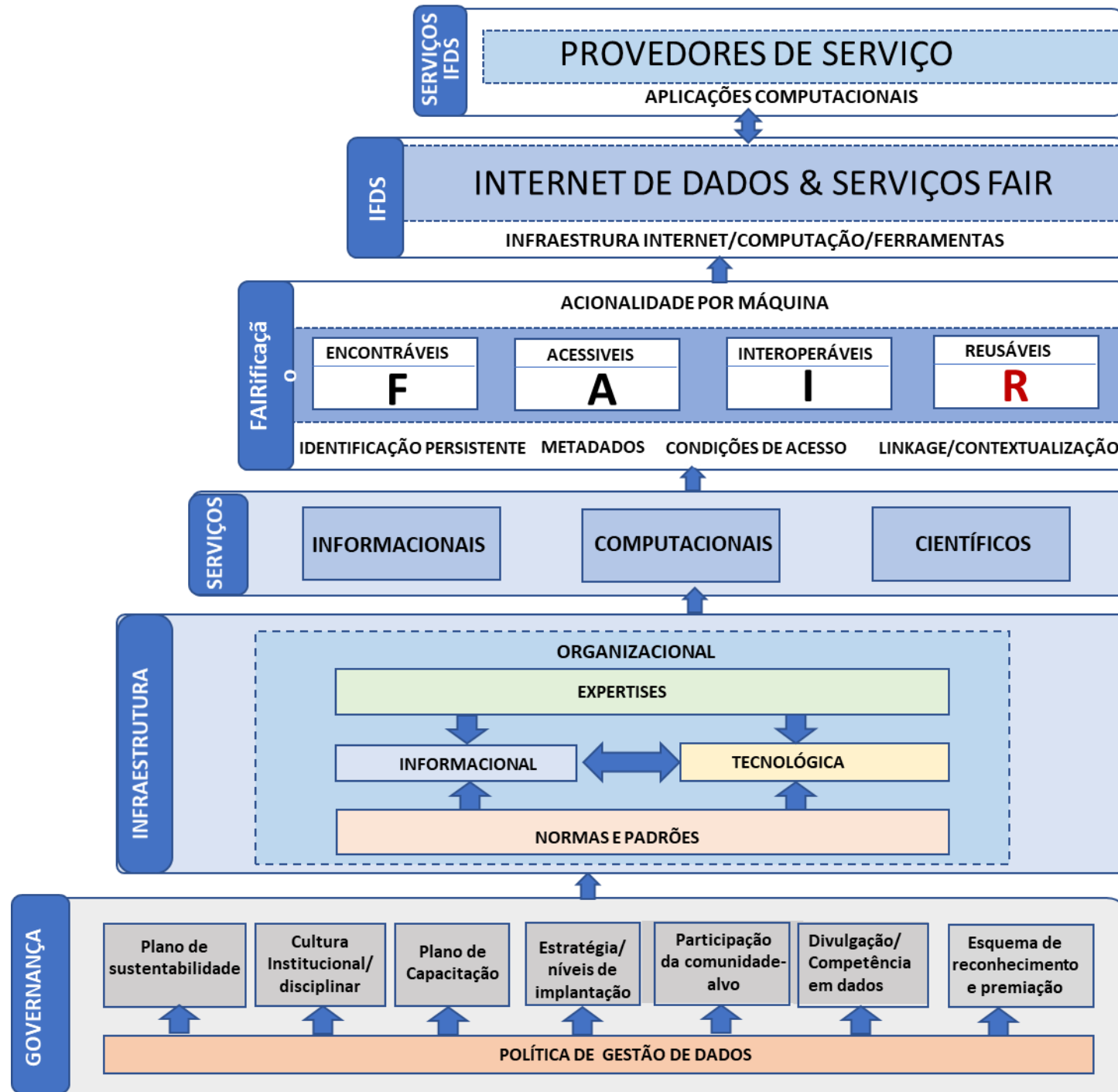
Dados com restrições de acesso

E

Dados disponíveis sob condições bem definidas para reuso.

F

Internet FAIR data: número de aplicações e serviços podem linkar e processar dados FAIR



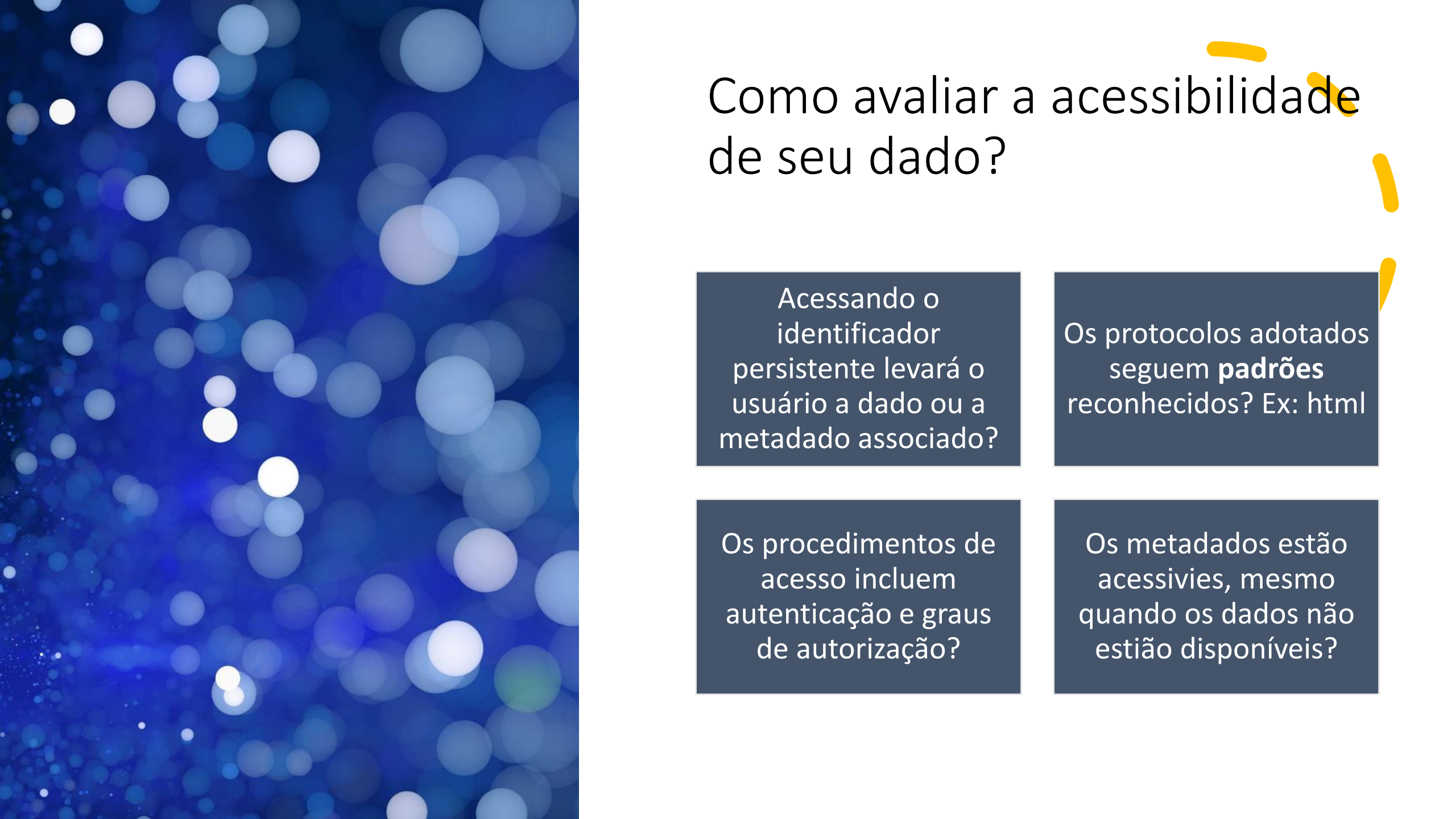
Como avaliar a encontrabilidade de seus dados?

Os seus dados estão associados a um **identificador persistente**?

Existem metadados ricos descrevendo os seus dados?

Os metadados estão acessíveis online em uma ferramenta de busca? Ex: catálogo ou repositório de dados

O registro do metadado especifica o identificador persistente?



Como avaliar a acessibilidade de seu dado?

Acessando o identificador persistente levará o usuário a dado ou a metadado associado?

Os protocolos adotados seguem **padrões** reconhecidos? Ex: html

Os procedimentos de acesso incluem autenticação e graus de autorização?

Os metadados estão acessíveis, mesmo quando os dados não estão disponíveis?

Como avaliar o nível de interoperabilidade de seus dados?



Os dados estão em formatos claramente compreendidos e de preferência abertos?

Os metadados seguem **normas** relevantes?

Os vocabulários controlados, palavras-chave, tesouros ou ontologias são utilizados sempre que possível?

Referências e links qualificados são fornecidos para uso por outros?

Como avaliar a
nível de
possibilidade
de reuso dos
seus dados?

Os dados são precisos e bem descritos com muitos atributos relevantes?

Os dados possuem uma licença de uso de dados clara e acessível?

Está claro como, por que e por quem os dados foram criados e processados?

Os dados e os metadados atendem a padrões e domínio relevante?

DICA DE
FERRAMENTA
PARA
AVALIAÇÃO



<https://wrco.ufpb.br/fair/index.html>

- CONVITE: <https://www.go-fair-brasil.org/>