

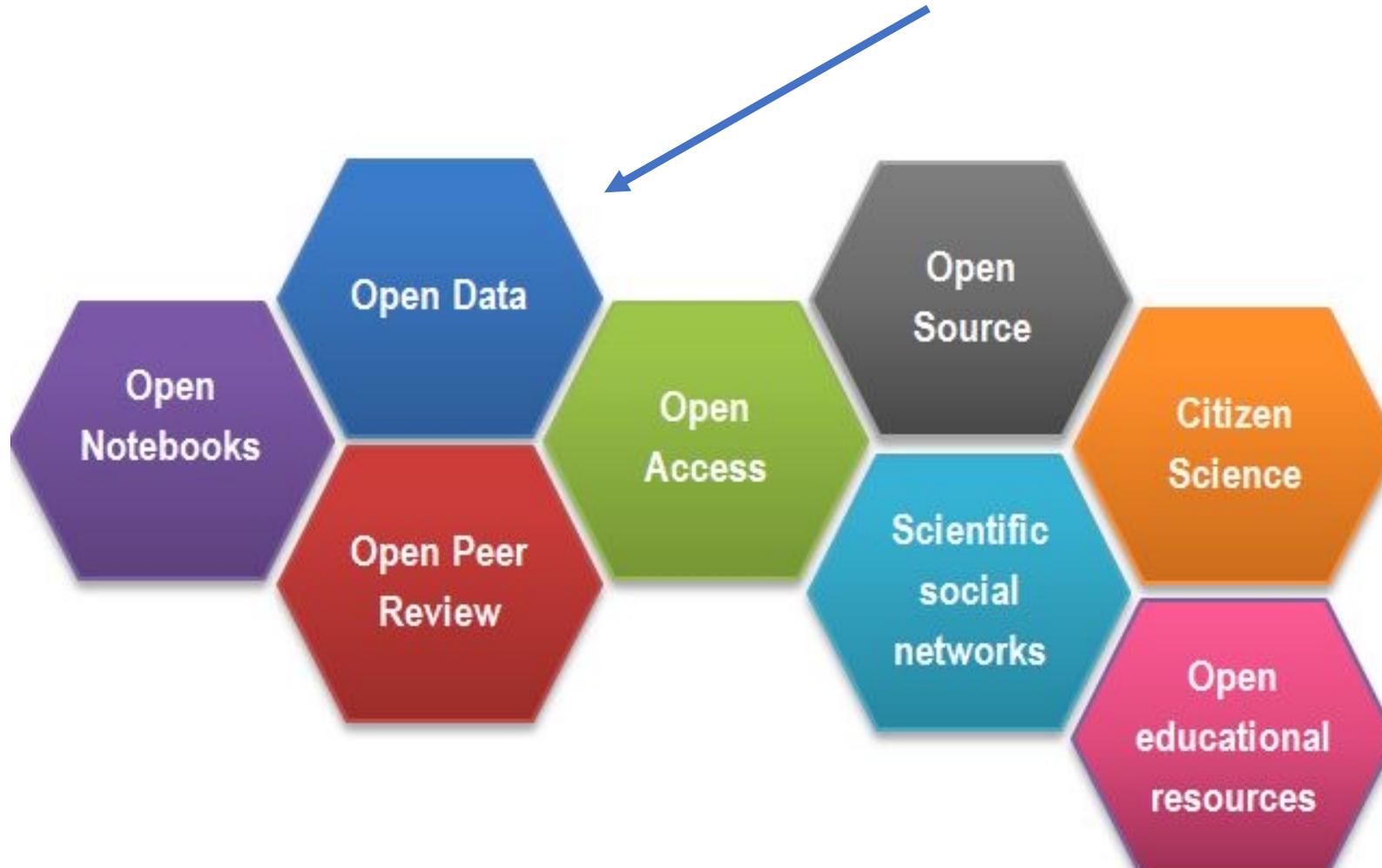
A Importância da Ciência Aberta e os Princípios FAIR

Luana Sales
PPGCI IBICT-UFRJ
luanasales@ibict.br



Para que Ciência Aberta?

A Ciência aberta não é só publicações e dados



**NOS ESTAMOS NOS AFOGANDO EM DADOS, MAS SEDENTOS DE
INFORMAÇÃO & CONHECIMENTO**



NOS ESTAMOS NOS AFOGANDO EM DADOS, MAS SEDENTOS DE
INFORMAÇÃO & CONHECIMENTO



GESTÃO/CURADORIA

GESTÃO DE DADOS DE PESQUISA

COMPARTILHAMENTO

REUSO

PRESERVAÇÃO

CONTEXTUALIZAÇÃO

PROVENIÊNCIA

AUTENTICIDADE

DADOS

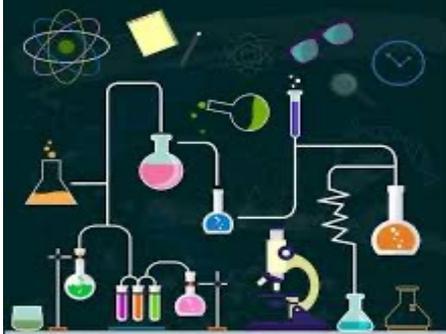


CONHECIMENTO

INFORMAÇÃO

Por que fazer Gestão de Dados?

SÃO SIGNIFICANTES OS BENEFÍCIOS POTENCIAIS PARA O MUNDO ACADÊMICO DA GESTÃO DE DADOS DE PESQUISA



MELHORES DADOS >> MELHORES PESQUISAS



Aumento do potencial de **compartilhamento/reuso** de dados de pesquisa, minimizando a necessidade de **duplicar trabalho no laboratório, no campo ou na biblioteca.**



Assegura que os dados obtidos por meio de altos custos não sejam **perdidos** ou inadvertidamente **destruídos** e mantenham sua **integridade, autenticidade e confiabilidade.**



Permitir que a **recuperação, comparação e meta análise dos dados** proveniente de múltiplas fontes possa levar a **novos importantes conhecimentos.**



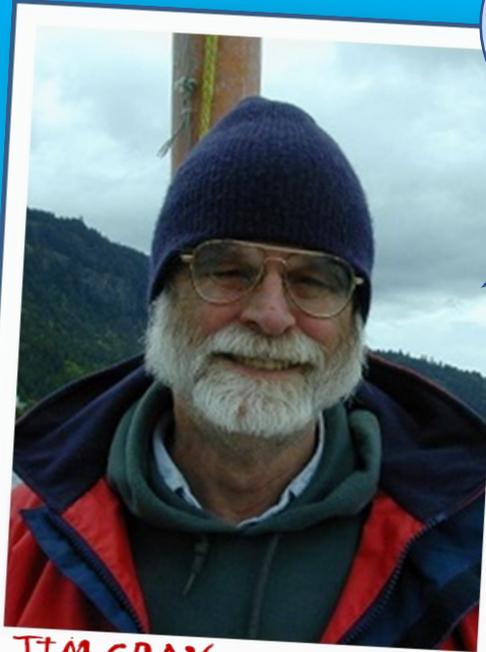
Possibilita a **verificação ou repetição de experimentos, validação de trabalhos acadêmicos** e a **verificação de descobertas**, particularmente importante em meio a preocupação nacional e internacional sobre integridade da pesquisa.



Novos temas de pesquisa – em particular temas **interdisciplinares** – podem emergir a partir **da reanálise de dados existentes ou comparações** com novos dados: crescentemente dados podem se tornar **ponto de partida para novas pesquisas**, bem como um produto para as pesquisas correntes.

DADO DE PESQUISA NÃO FALA POR SI PRÓPRIO

Dados de pesquisa são **incompreensíveis e portanto inúteis** a menos que haja uma **descrição detalhada e clara** de como e quando eles foram obtidos e de como os **dados derivados** foram produzidos !!!



JIM GRAY
Cientista da Computação
Desaparecido em 2007

Para entender os dados os usuários futuros necessitam de metadados, caso contrário eles não saberão os detalhes de como os dados foram **obtidos e preparados** : 1) como os **instrumentos foram projetados e construídos**; 2) **quando, onde e como** os dados foram coletados; e 3) e não terão **uma descrição dos processos que levaram aos dados derivados**, que são tipicamente usados para análises científicas de dados.

Gray, 2002





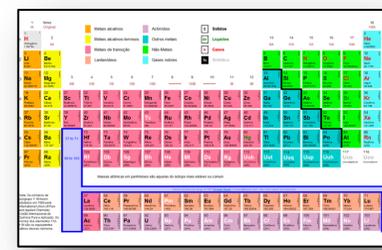
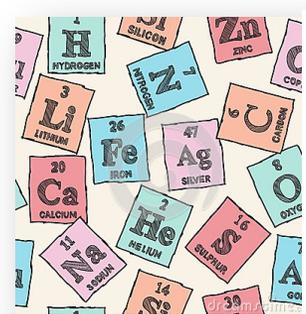
SIGNIFICADO
ESTRUTURA
IDENTIFICAÇÃO
CONTEXTO
PROVENIÊNCIA

METADADOS

METADADOS QUE DOCUMENTAM OS ELEMENTOS SEMÂNTICOS, AS PARTES DOS OBJETOS E SUAS RELAÇÕES, AS DEPENDÊNCIAS TÉCNICAS, A PROVENIÊNCIA, A IDENTIFICAÇÃO PERSISTENTE, AS RESTRIÇÕES E DIREITOS ASSOCIADOS AOS DADOS, AS POSSÍVEIS INTERVENÇÕES SOFRIDAS E SEUS EFEITOS. OU SEJA, OS METADADOS DEVEM REGISTRAR IDEALMENTE TUDO QUE DEVE SER DE INTERESSE DO PESQUISADOR, **INCLUINDO MODELOS DE DADOS, EQUIPAMENTOS ESPECIAIS, ESPECIFICAÇÃO DA INSTRUMENTAÇÃO, LINHAGEM DOS DADOS E MUITO MAIS.**

POR QUE? QUEM? O QUE? COMO? QUANDO? ONDE?

SIGNIFICADO
ESTRUTURA
IDENTIFICAÇÃO
CONTEXTO



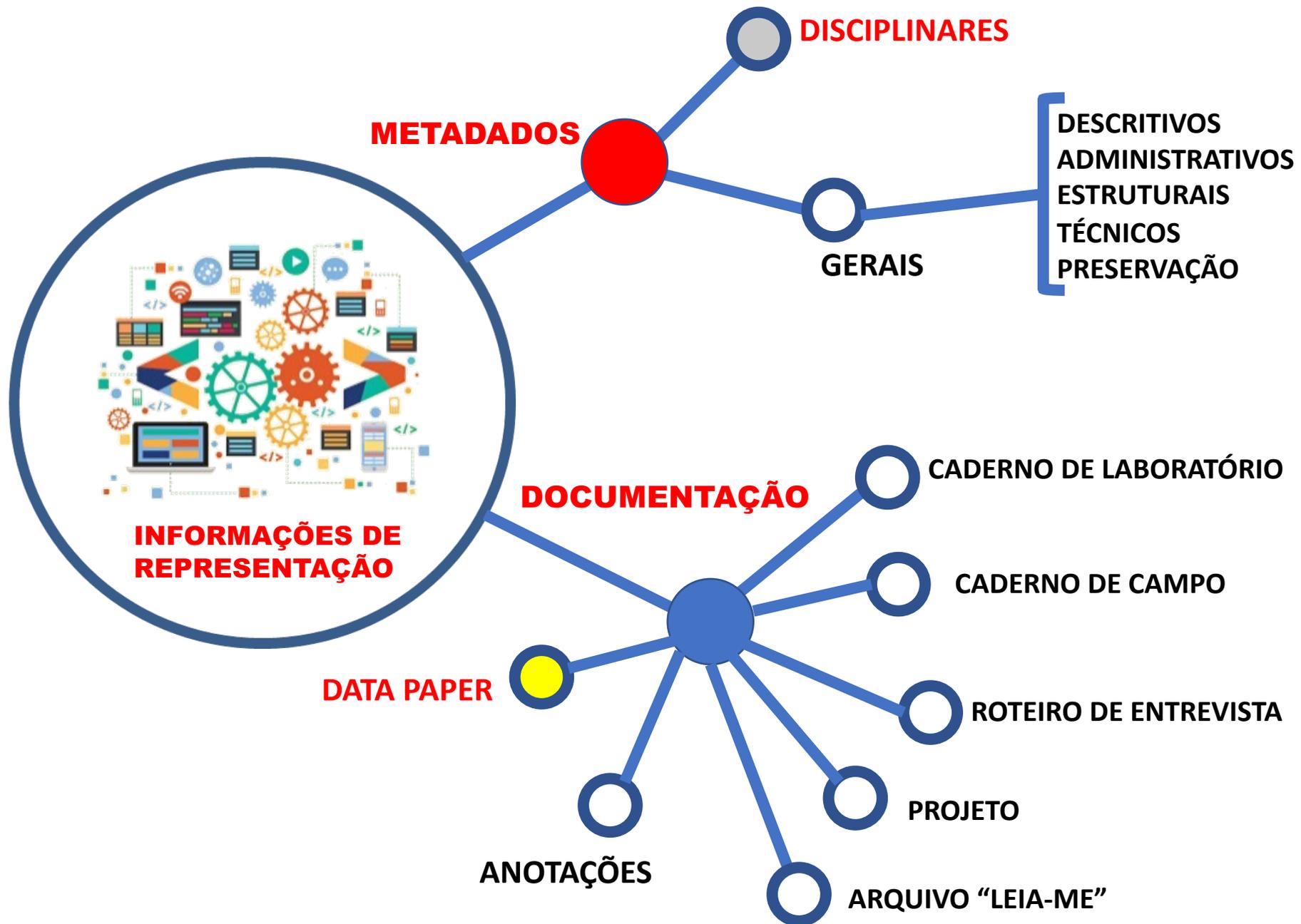
DADOS

+

INFORMAÇÃO DE REPRESENTAÇÃO

=

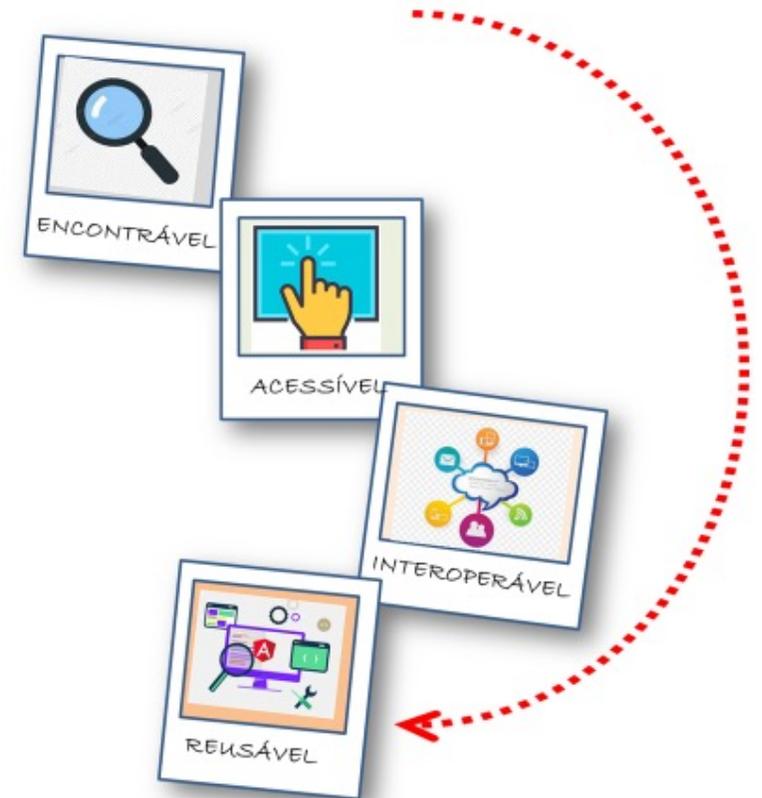
INFORMAÇÃO



PARA QUE DADOS DE PESQUISA SEJAM REUSADOS, ELES PRECISAM PASSAR POR UMA GESTÃO ADEQUADA QUE OS TRANSFORME EM 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.
- Dados FAIR são para serem encontrados, interoperados, processados e reusados por máquinas

PRINCÍPIOS
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 Organizers

- Scott Lusher, NLeSC Amsterdam
- Barend Mons, Leiden UMC

Topics

- Towards a Modular Blueprint "Floor-plan" of a Safe and Fair Data Stewardship, Trading 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 endorse a concise and

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

Scientific Data 3, Article number: 160018 (2016) | Download Citation &

An Addendum to this article was published on 19 March 2019

Download PDF

781 Citations | 1337 Altmetric | Article metrics >

Associated Content

Collection: Metadata Quality

Sections | **References**

Abstract

Comment

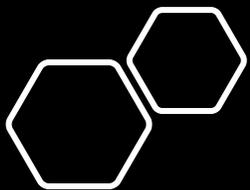
Additional Information

References

Acknowledgements

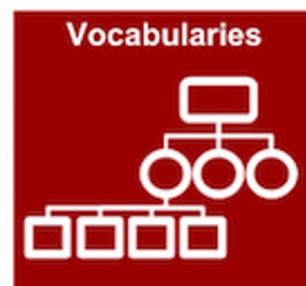
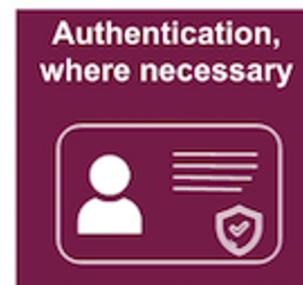
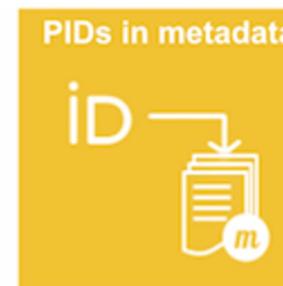
The Lorentz Center is an international center in the sciences. Its aim is to organize workshops for scientists in an atmosphere that fosters collaborative work, discussions and presentations. For registration, see: www.lorentzcenter.nl

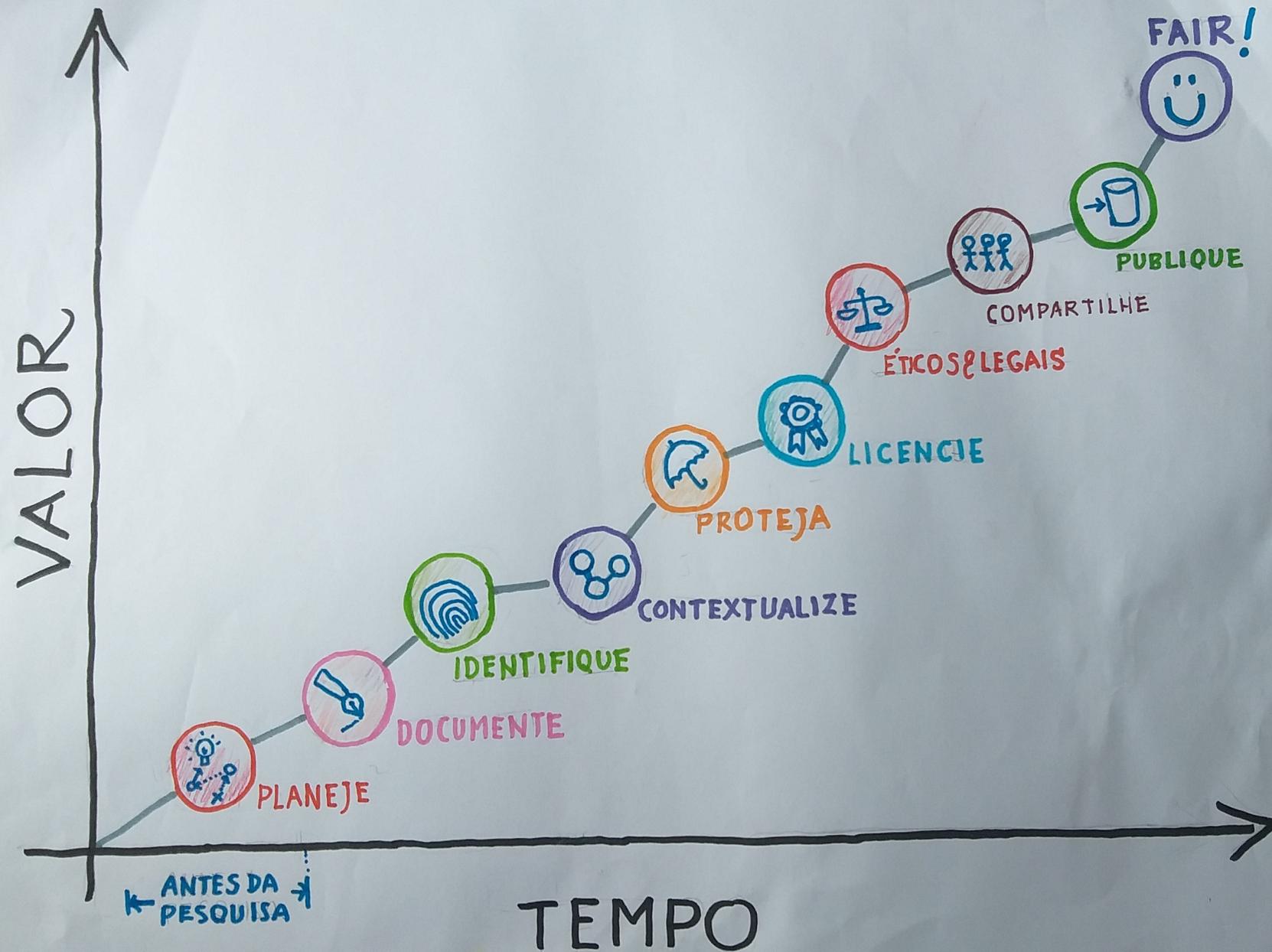
W w w . l o r e n t z c e n t e r . n l



15 PRINCÍPIOS

Criados para nortear boas práticas no âmbito da Gestão de Dados científicos, os princípios FAIR foram difundidos globalmente como forma de tornar DADOS de pesquisa, dados que podiam ser reusados por homens e por máquinas sem restrições





Curadoria – atividade de gestão de dados que compreende os diversos processos de agregação de valor aos dados por todo o seu ciclo vida que se inicia antes de sua criação/coleta; tem como objetivo de assegurar que eles estejam adequados aos propósitos atuais e potenciais e disponíveis para descoberta e reuso agora e no futuro. Para *datasets* dinâmicos isto significa um enriquecimento ou atualização contínua com o objetivo de mantê-los adequados e prontos para possíveis repropósitos.



ALGUMAS VERDADES INCONVENIENTES

- PRINCÍPIOS FAIR NÃO FORAM FEITOS PARA A CIÊNCIA ABERTA
- DADOS FAIR NÃO SÃO DADOS ABERTOS
- É PRECISO MAIS QUE ABRIR DADOS E PUBLICAÇÕES PARA ATINGIRMOS O OBJETIVO DA CIÊNCIA ABERTA

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.

**LIMITES
DA
ABERTURA
DOS
DADOS
DE
PESQUISA**

1

INTERESSE COMERCIAL

VALORIZAÇÃO COMERCIAL DOS DADOS
PROTEGER A INOVAÇÃO
PATENTES

2

PRIVACIDADE

INFORMAÇÕES PESSOAIS: CRÍTICO PARA A PESQUISA NAS ÁREAS DE SAÚDE E CIÊNCIAS SOCIAIS;
GOVERNANÇA X PROTEÇÃO DA CONFIDENCIALIDADE
ANONIMIZAÇÃO NÃO GARANTE

3

AMEAÇA A SEGURANÇA

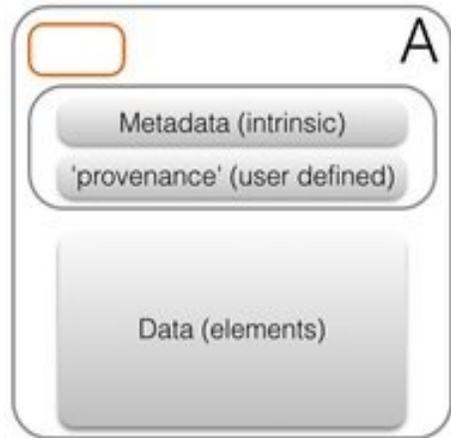
SEGURANÇA E SAÚDE PÚBLICA
PROTEGIDO CONTRA INCIDENTES NÃO INTENCIONAIS
PROTEGIDO CONTRA ATAQUES DELIBERADOS
“SÓ METADE DOS DADOS QUE DEVEM SER PROTEGIDOS ESTÁ PROTEGIDO” (IDC, 2011)

Mas...

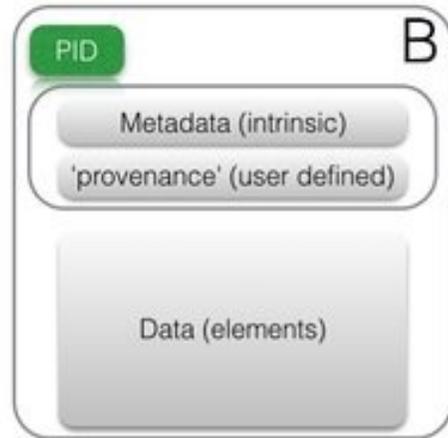
A CIÊNCIA ABERTA
PRECISA DOS PRINCÍPIOS
FAIR PARA SER
OPERACIONALIZADA

NIVEIS DE FAIRNESS X ABERTURA DOS DADOS

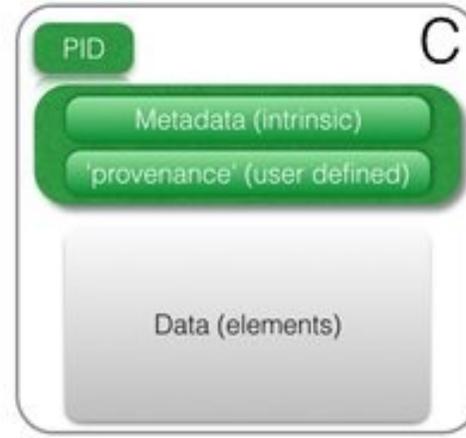
Re-useless data (80%)



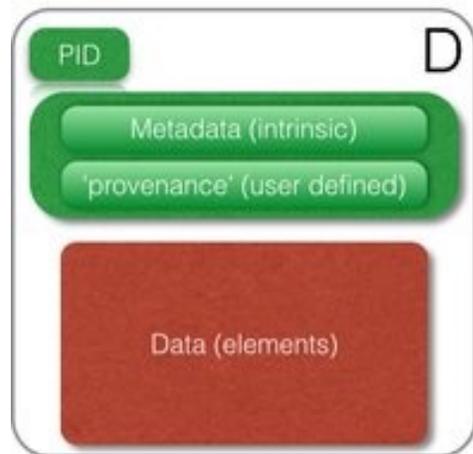
Findable



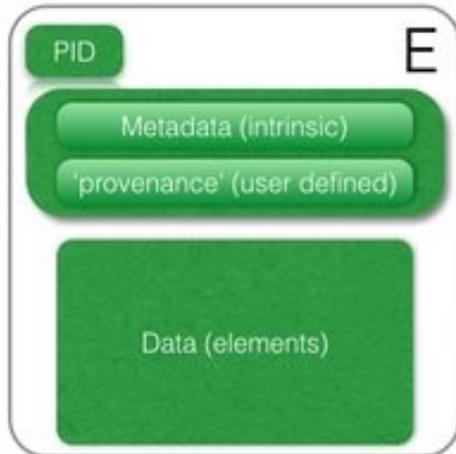
FAIR metadata



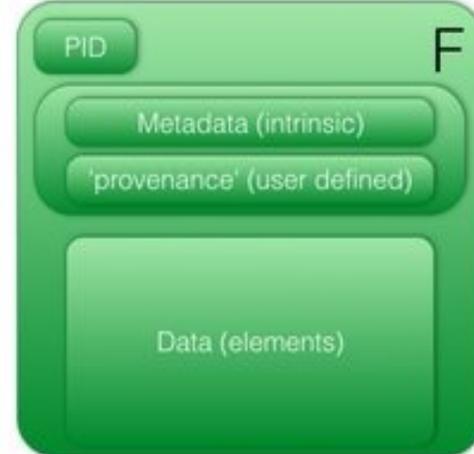
FAIR data-
restricted access



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

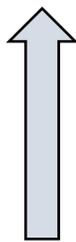
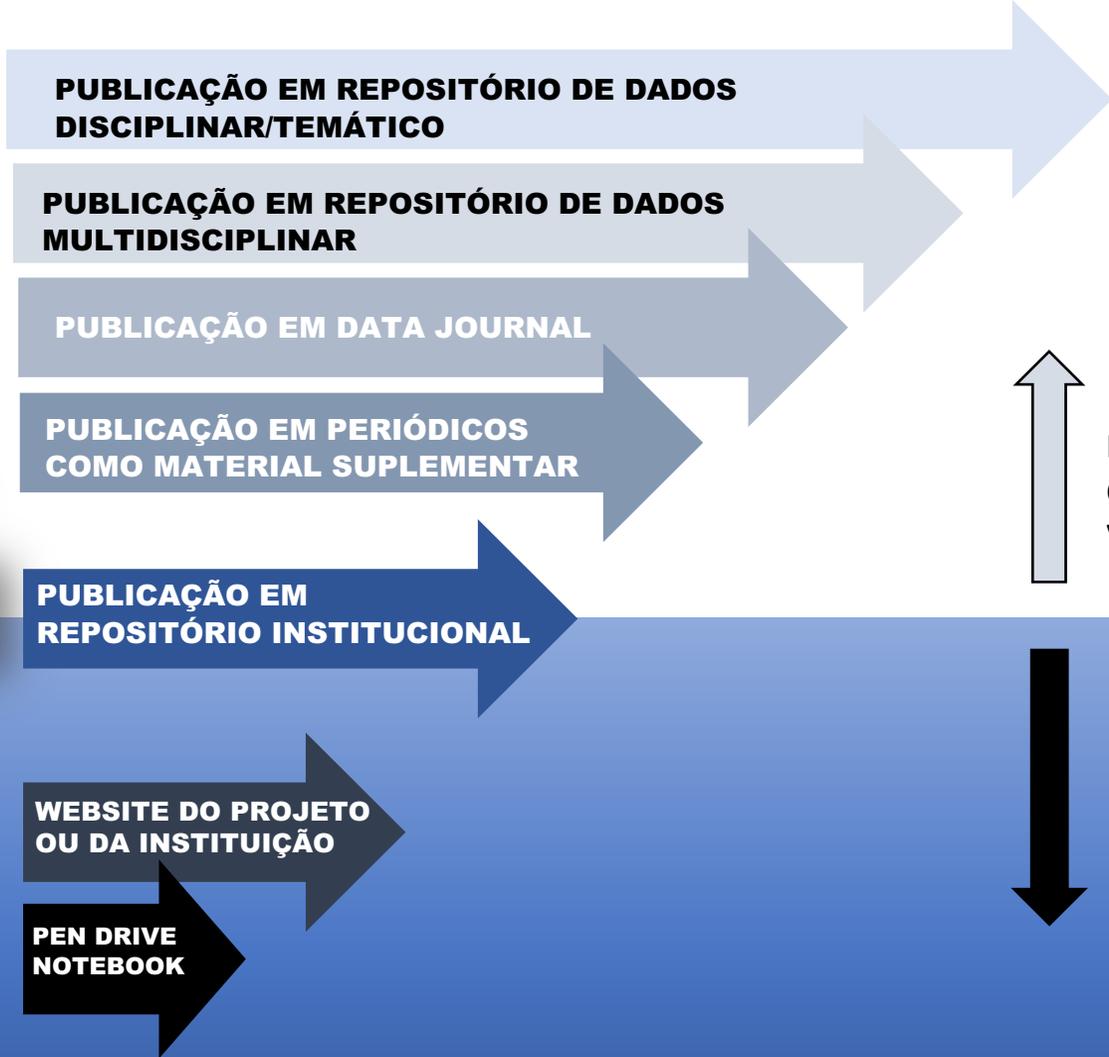
Dados com restrições de acesso

E

Dados disponíveis sob condições bem definidas para reuso.

F

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

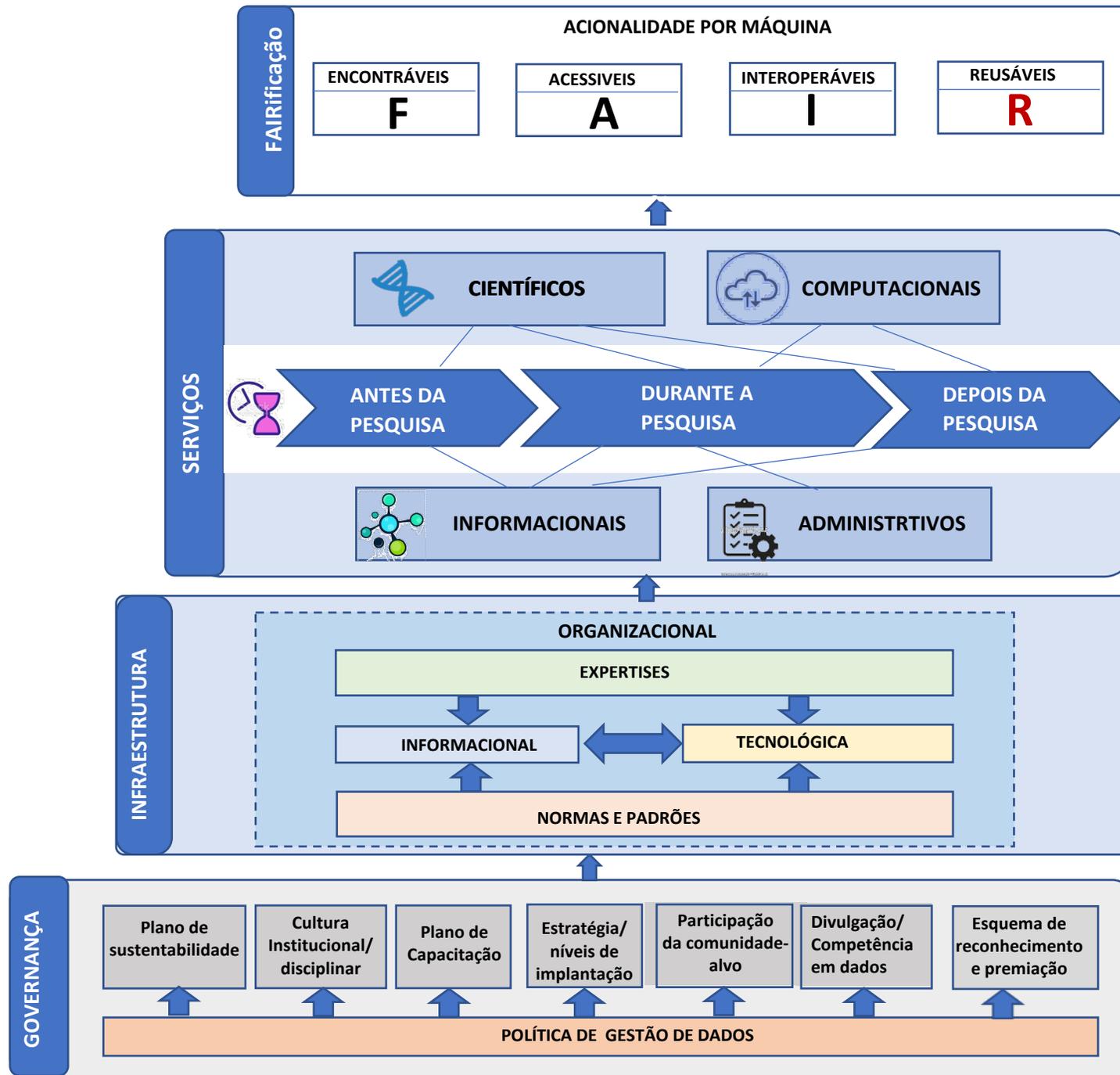


**REUSO
COMPARTILHAMENTO
VISIBILIDADE**



INVISIBILIDADE

“ É preciso criar serviços que operacionalizem a Ciência Aberta e apoiem a FAIRificação dos Dados





REPOSITÓRIO

SERVIÇOS, FERRAMENTAS,
PADRÕES, INFRAESTRUTURAS



PLATAFORMA DE GESTÃO DE DADOS

Enfoque disciplinar/comunitário



FAIRIFICAÇÃO

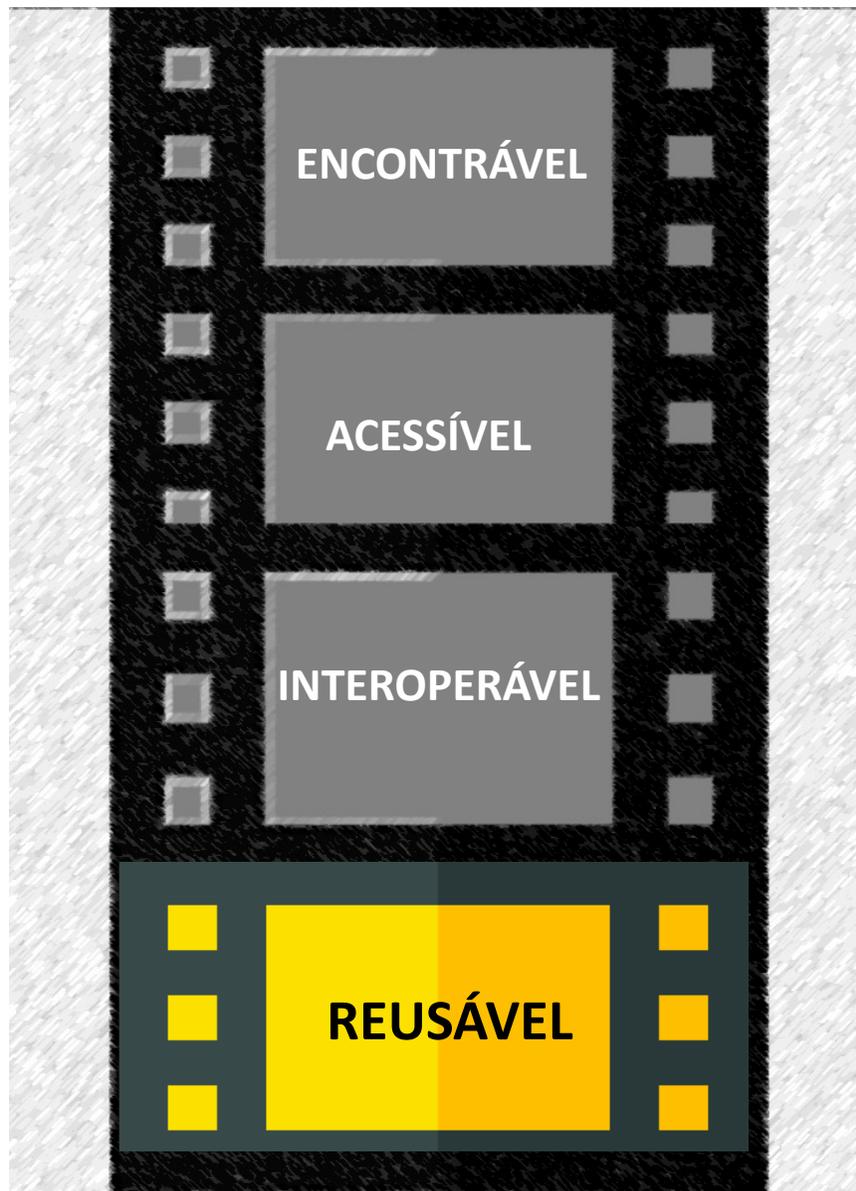
DADOS FAIR

FACILMENTE
ENCONTRÁVEL
POR **MAQUINAS** E
HU.MANOS
**ACIONÁVEL POR
MÁQUINA**

**AUTENTICAÇÃO E
VERIFICAÇÃO** É POSSIVEL
DEPOIS DO USUÁRIO
ACESSAR O (META)DADO
**PPOTEÇAO/CONDIÇÕES DE
ACESSO.**

**INTEGRAÇÃO COM
OUTROS OBJETOS** E
SISTEMAS OU WORFLOW
PARA ANÁLISE,
ARMAZENAMENTO E
PROCESSAMENTO SÃO
INTEROPERAVEIS

DADO -CONDIÇÕES DE
REUSO. NEM SEMPRE
DISPONÍVEL.
METADADO – SEMPRE
DISPONÍVEL



**AMBOS OS PRINCÍPIOS TÊM O PROPÓSITO DE
TORNAR OS DADOS REUSÁVEIS**

DADOS ABERTOS

DADOS DISPONÍVEIS
DE FORMA COMPLETA, A
BAIXO CUSTO E
PREFERENCIALMENTE
VIA INTERNET.

**NÃO HÁ BARREIRAS
PARA O ACESSO AOS
DADOS:** ECONÔMICAS,
LEGAIS, LINGUISTÍCAS

**USO, REUSO E
REDISTRIBUIÇÃO** DE
DADOS E METADADOS
POR QUEM QUEIRA.
NÃO HÁ DISCRIMINAÇÃO
EM RELAÇÃO A QUEM
VAI USAR E PARA QUE
PROPÓSITO

À Guisa de Conclusão

- Nem todo Dado Aberto é FAIR
- Nem todo Dado FAIR é Aberto
- Não basta abrir por abrir...
- Dados abertos devem ser reusados
- Para serem reusados, dados precisam ser compreendidos
- Para serem compreendidos, dados precisam de bem gerenciados
- Dados bem gerenciados são Dados FAIR

	FAIR DATA	OPEN DATA	SIMILARIDADE	DIFFERENÇA
FINDABLE	Dados devem ser encontrados por máquinas e humanos	Dados devem estar disponíveis em sua integridade e com baixo custo de reprodução	Encontrabilidade e acessibilidade no FAIR é similar a disponibilidade de dados abertos, pois prover acesso ao usuário	Disponibilidade de dados abertos refere-se à integridade dos dados e não menciona outras condições
ACESSIBLE	Autenticação e verificação devem ser feitos antes do acesso ao dado pelo usuário			Dados abertos focam na destruição de barreiras à acessibilidade dos dados, enquanto acessibilidade nos princípios FAIR ressalta a necessidade de proteção e que condições de acesso devem ser formuladas para atender circunstâncias específicas relacionadas ao dado
INTEROPERABLE	Dado deve ser integrado com outros dados e sistemas ou workflows para análise, armazenamento, processamento e interoperabilidade			A interoperabilidade é promovida pela criação de instâncias de ontologias legíveis por máquinas que representam dados, ligados aos metadados em linguagens como JSON ou RDF, amplamente usadas na Web Semântica.
REUSABLE	Metadados e dados devem ser definidos para reuso e para poderem ser replicados ou misturados em outros contextos.	Os dados devem estar disponíveis sob condições de reuso e redistribuição. qualquer pessoa deve poder utilizar, reutilizar e redistribuir os dados - não há discriminação com base na finalidade para a qual os dados serão utilizados ou nos indivíduos/grupos que desejam utilizá-los.	Ambos princípios têm o propósito de tornar dados reusáveis	Dados abertos não mencionam metadados e focam na neutralidade da redistribuição