



**WIDaT
2026**

Da FAIRificação à IA:

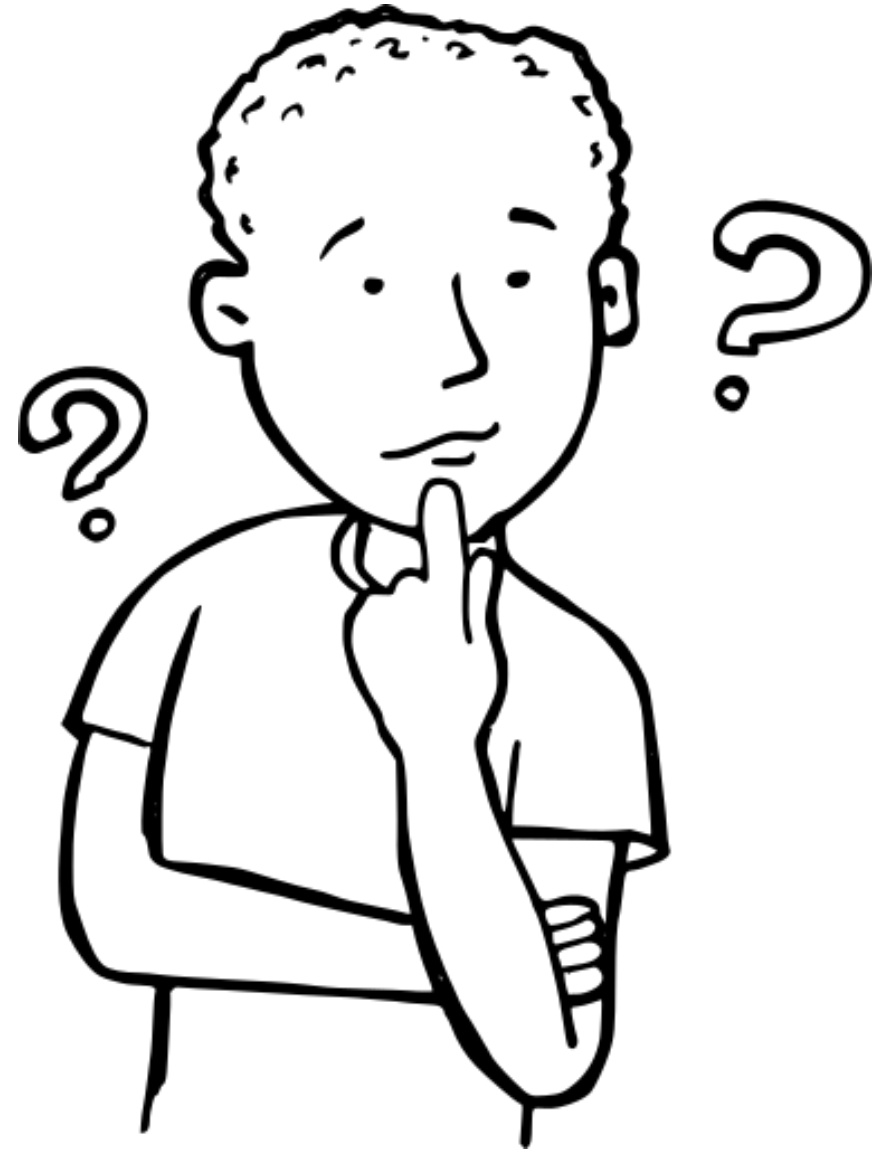
**gestão de dados de pesquisa como caminho para
inovação científica**

Prof.Luana Sales PPGCI - Ibict /
PPGB - UNIRIO - luanasales@ibict.br

Prof.Luís Sayão PPGCI - Ibict / PPGB
- UNIRIO / PPGMA - CRB -
luisfernandosayao@gmail.com

A Inteligência Artificial depende menos de algoritmos e mais de dados de qualidade

- Como a IA pode produzir conhecimento confiável?
- O que acontece quando os dados não são encontráveis, acessíveis ou reutilizáveis?
- Estamos preparando nossos dados para serem utilizados por humanos e máquinas?



Acesso de dados de pesquisa e o
elo entre Ciência Aberta
Qualificada e Inteligência
Artificial

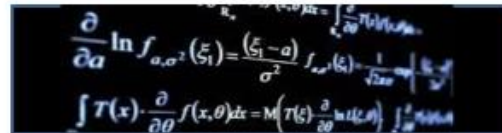
A transformação da pesquisa

OS PARADIGMAS CIENTÍFICOS [segundo a eScience]



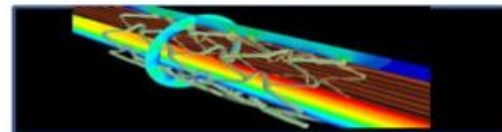
1º PARADIGMA:

Ciência experimental ou **empírica** estuda a relação entre fenômenos por meio de experimentos; descreve os fenômenos naturais



2º PARADIGMA:

Ciência teórica ou **descritiva** formula **modelos** para descrição, explicação e generalização dos fenômenos naturais



3º PARADIGMA:

Ciência computacional baseada em **simulação** de fenômenos complexos por meio de uso de *software*, gera grande quantidade de **dados**



4º PARADIGMA:

Ciência exploratória baseado no uso intensivo de dados, exploração estatística e mineração de dados
Unifica teorias, experimentos e simulação

- Dados capturados por instrumentos ou gerados por simulação;
- Processados por software
- Informação/conhecimento armazenadas em computadores
- Análise de base de dados por estatística e mineração



Pré-Renascença
Experimentos

1600

Pré-computação
Leis da mecânica clássica,
Eletromagnetismo,
Gravitação universal

1950

Pré-big data
Simulação da natureza,
Método Monte Carlo

2010

Agora
Detecção de padrões,
relações e anomalias.
Big Data
Inteligência Artificial

PROTAGONISMO DE DADOS NA CIÊNCIA



eScience

BIG DATA CIENTÍFICO
Grandes projetos
Observatórios
Instalações complexas
Dados distribuídos
Simulação por computador

Ciência aberta

DADOS ABERTOS
Metodologias
Equipamentos
Software
Cadernos de laboratório
Roteiro de entrevistas
Resultados negativos

Cauda longa

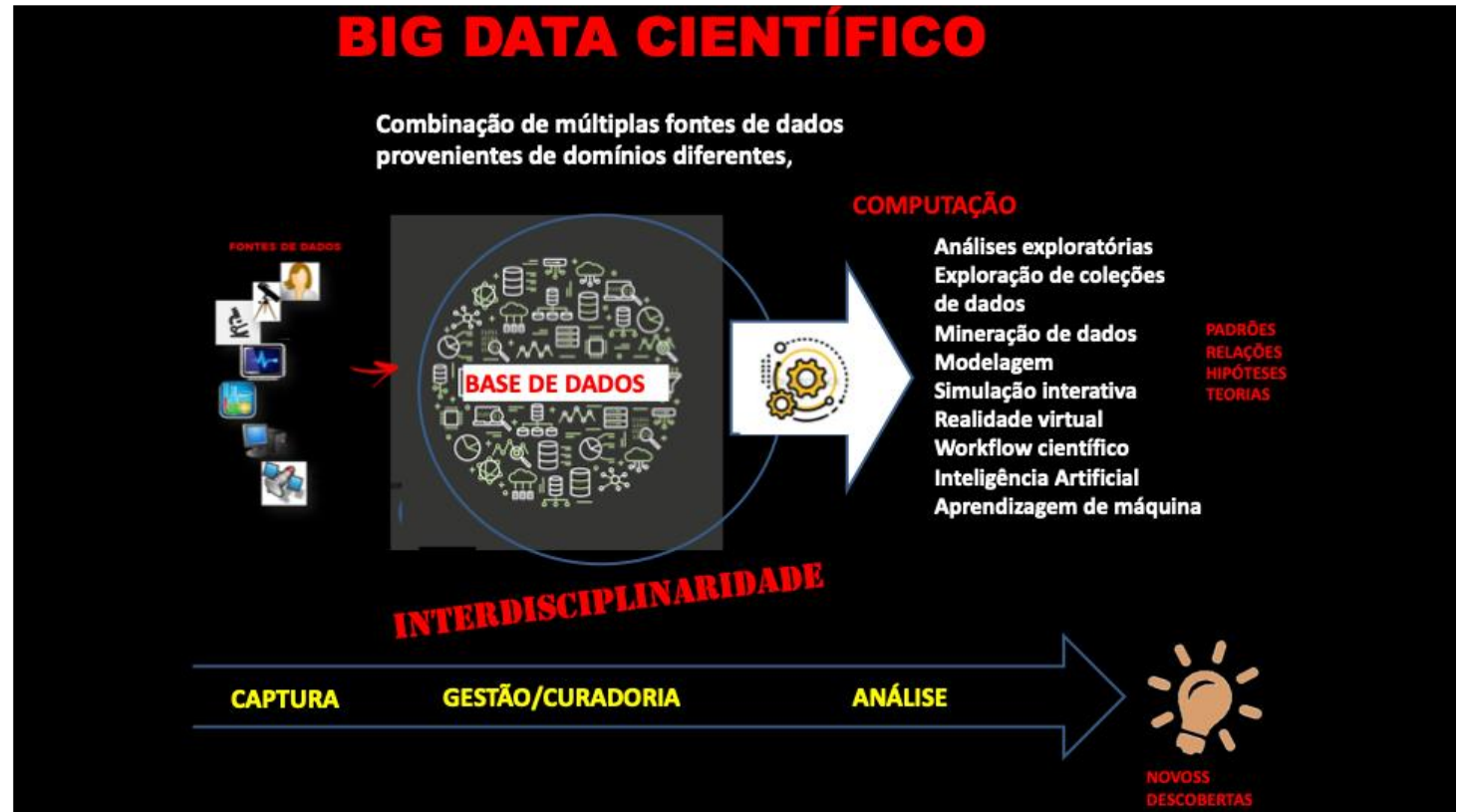
**DADOS DOS DO GRANDE
NÚMERO DE PEQUENOS
LABORATÓRIOS**
Heterogêneos
Não tratados
Invisíveis
**Coletivamente é o maior
volume**

Humanidades
Digitais

**TECNOLOGIA
COMPUTACIONAL
APLICADAS A ESTUDOS EM
HUMANIDADES.**
Humanidades estudando
Tecnologias digitais
(Boble)

O novo contexto da ciência

- Crescimento exponencial de dados científicos
- Necessidade de colaboração global
- Automação da descoberta científica



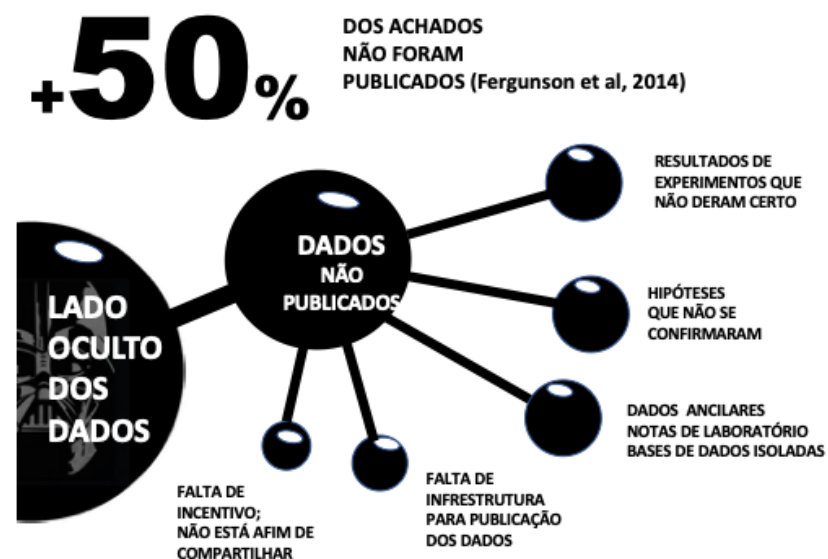
O problema dos dados de pesquisa

DADOS PRECISAM ESTAR VISÍVEIS

- Nem todo dado é reutilizável

- Dados dispersos
- Falta de documentação
- Metadados insuficientes
- Formatos proprietários
- Ausência de proveniência

Dados
invisíveis
Ou Dark Data



O COMPARTILHAMENTO PODE REVELAR VALORES IMPORTANTES OCULTOS NESSES DADOS

C I Ê N C I A I N V I S Í V E L

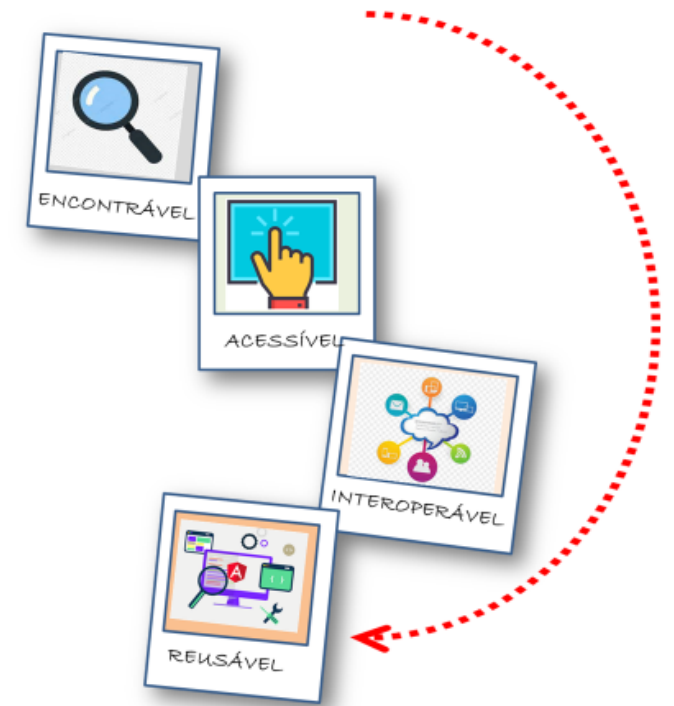
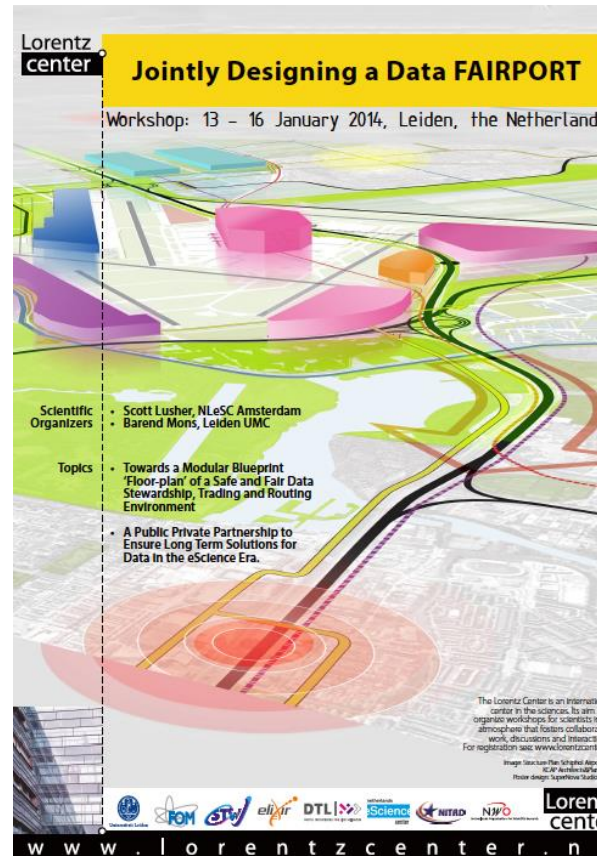
- Resultado: Muitos dados são produzidos, mas poucos podem ser efetivamente reutilizados.

O surgimento d

- Em 2014, no *Jointly designing a data FAIRPORT Conference*, um grupo de especialistas de diversas áreas do conhecimento que atuavam no contexto da E-Science e estavam interessados no reuso de dados estabelecem os princípios FAIR no contexto da e-Science.

PRINCÍPIOS

FAIR



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



TORNAR OS DADOS FAIR



ENCONTRÁVEL

FINDABLE

Tornar seu dados fáceis de encontrar por você mesmo ou por seus parceiros e colaboradores, otimizando o tempo dedicada a pesquisa e a inovação, mas significa também que seus dados sejam capazes de serem descobertos por sistemas computacionais.

ACESSÍVEL

ACCESSIBLE

Indica que seus dados estão depositados em serviços apropriados como repositório, **mas não significa que os dados estejam disponíveis abertamente para todos**, entretanto as condições de acesso definidas por você devem estar explícitas. A descrição dos dados - os metadados - devem também estar sempre disponíveis

INTEROPERÁVEL

INTEROPERABLE

Tornar o dado de pesquisa **interoperável**—via padrões e ontologias—permite integrar fontes heterogêneas, reaproveitar ferramentas e automatizar análises, acelerando descobertas e aumentando a confiabilidade dos resultados.

REUSÁVEL

REUSABLE

Tornar o dado reutilizável—com metadados ricos, formatos abertos e licenças claras—acelera a análise, reduz erros e viabiliza a replicação, economizando tempo, evitando retrabalho e ampliando o impacto das pesquisas.

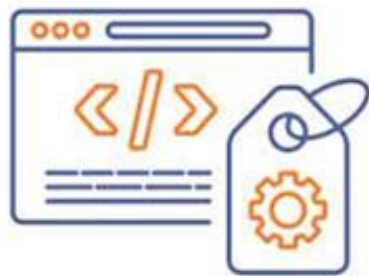
FAIR não significa:

- 
- Aberto
 - Gratuito
 - Sem restrições

FAIR é sobre...

Dados

Bem descritos
Bem documentados
Legível por
humanos e
máquinas



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 acionalidade 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

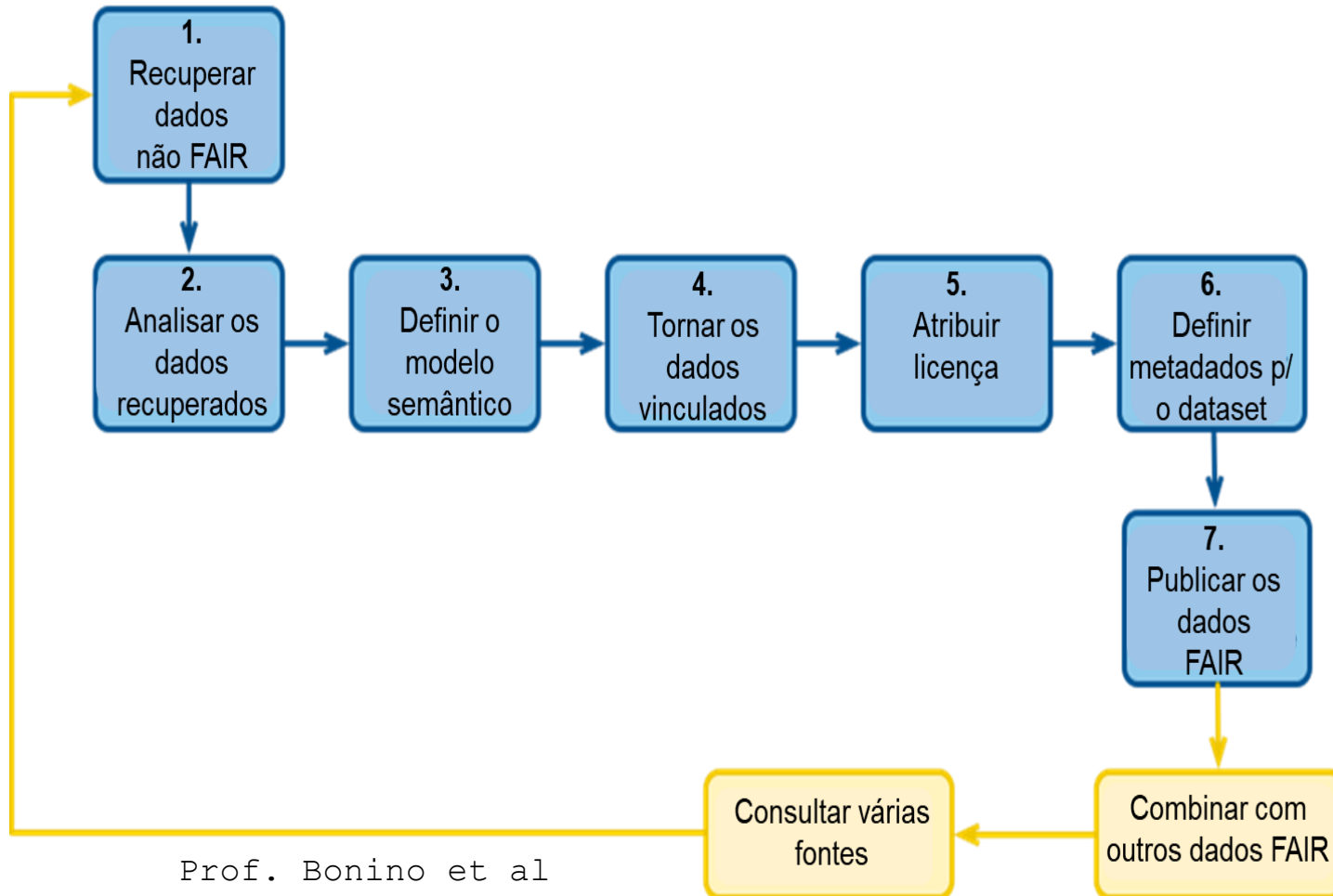
- **FAIR não é igual a aberto**;
- Há razões legítimas para blindar os dados e serviços gerados com fundos públicos que requerem medidas adicionais de autorização e autenticação tanto para humanos quanto para máquinas;
- FAIR não endereça questões e morais sobre a abertura dos dados: critério do custodiante.

Da publicação do dado à FAIRificação: o que é FAIRificação?

- Definição: Processo de transformar dados em recursos preparados para descoberta, integração e reutilização.
 - Metadados ricos
 - Identificadores persistentes
 - Vocabulários controlados
 - Ontologias
 - Proveniência
 - Licenciamento

A FAIRificação na prática

- Componentes da FAIRificação



Dados Brutos
↓
Curadoria
↓
Metadados
↓
PIDs
↓
Vocabulários
↓
Repositórios
↓
Dados FAIR

© 2006 The Authors

PLATAFORMA DE GESTÃO DE DADOS
Enfoque disciplinar/comunitário

Exigência das agências de fomento

Gestão como parte do projeto de pesquisa

DESCRIÇÃO
METADADOS
POLÍTICA DE COMPARTILHAMENTO
GESTÃO DE ARQUIVOS
PRAZOS DE DIGITALIZAÇÃO
PESQUISA
ORÇAMENTO
ENVIO
DE

[illegible]

Além da rotulagem e da categorização, os metadados podem ser considerados de forma mais universal **como uma linguagem de valor agregado** que serve como uma camada integrativa em um sistema de informação”.

Da FAIRificação à Machine Actionability

- Os princípios FAIR foram concebidos para: "Machines and humans"
- Conceitos:
 - Machine-readable -/ legível por máquina = a máquina consegue ler
 - Machine-actionable / acionável por máquina = a máquina consegue interpretar e executar ações automaticamente

Para dados serem acionáveis por máquina, eles precisam ser descritos por metadados ricos e estruturados, vocabulários controlados (ontologias), grafos do conhecimento (relações semânticas explícitas) identificadores persistentes e APIs

Por que a IA precisa de dados FAIR?

- A IA é tão boa quanto os dados que utiliza!!!
- Problemas da IA sem FAIR:
 - Alucinações
 - Viés
 - Falta de rastreabilidade
 - Baixa reprodutibilidade

IT'S NOT FAIR!



FAIR é um pré-requisito para IA científica confiável

- Benefícios da IA com FAIR:
 - Dados confiáveis
 - Proveniência explícita
 - Contexto preservado
 - Reuso automatizado



FAIR Digital Objects e Ecossistemas Inteligentes

- **A convergência**

- FAIR Digital Objects (FDOs)
- PIDs
- Metadados acionáveis
- Infraestruturas FAIR



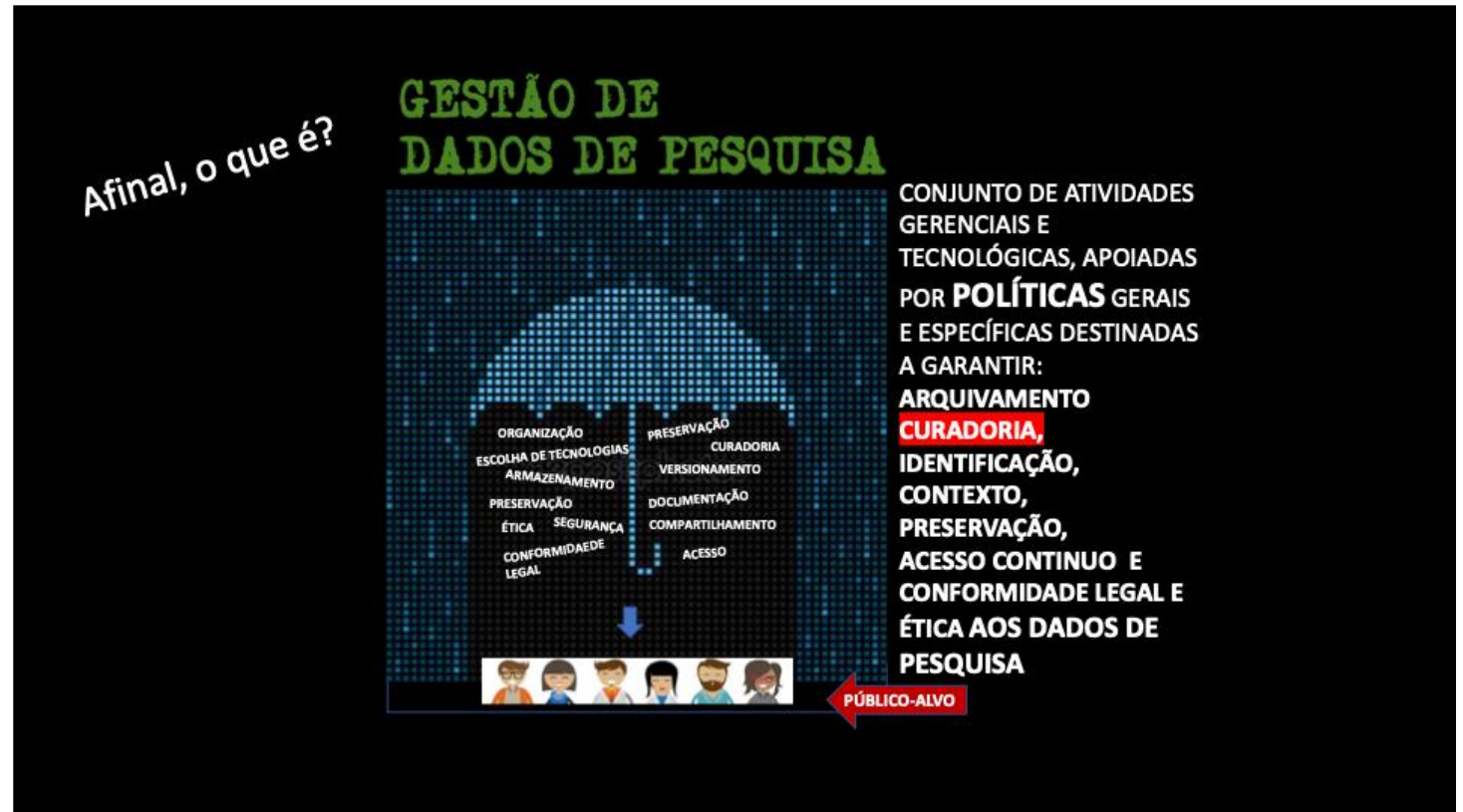
FAIR Digital Objects (FDO) e Internet FAIR de Dados e Serviços (IFDS)

- FDO - unidade digital identificada de forma única e persistente, associada a metadados ricos e padronizados, capaz de ser descoberta, acessada, interpretada e processada automaticamente por sistemas computacionais em ambientes distribuídos.
- IFDS - ecossistema digital global no qual dados, metadados, serviços, softwares e outros objetos digitais são organizados segundo os princípios FAIR e podem ser descobertos, acessados, combinados e reutilizados automaticamente por humanos e máquinas.



O papel da Ciência da Informação: contribuições da área

- Muito além da tecnologia...
 - Organização do Conhecimento
 - Metadados
 - Curadoria Digital
 - Proveniência
 - Governança de Dados
 - Preservação Digital
 - Interoperabilidade semântica
- Não existe IA científica confiável sem uma boa Gestão de Dados e Metadados



O que precisamos fazer agora?

- **Agenda para instituições e pesquisadores**
 - Planejamento da gestão de dados
 - Capacitação em gestão de dados FAIR
 - Criação de Infraestruturas FAIR
 - Estabelecimento de políticas institucionais
 - Curadoria contínua
 - Integração das bases com IA



A FAIRificação não é o destino; é a ponte...

....Da gestão de dados para:

- Ciência Aberta Qualificada
- Reprodutibilidade Científica
- Interoperabilidade
- Inteligência Artificial
- Inovação científica



À Guisa de Conclusão

A gestão de dados de pesquisa evoluiu de uma prática de organização e preservação para uma infraestrutura estratégica capaz de tornar os dados acionáveis por máquinas, constituindo a base sobre a qual a Inteligência Artificial poderá produzir ciência confiável, transparente e reutilizável.

Se os dados são o combustível da IA, a FAIRificação é o processo que transforma esse combustível em energia para novas descobertas científicas