



CURADORIA DE DADOS DE PESQUISA

EM REPOSITÓRIOS

Luís Fernando Sayão
CNEN/CIN – UNIRIO/PPGB – CASARUI/PPGMA
Luana Farias Sales
IBICT/COEP

CURADORIA: A PRÓXIMA FRONTEIRA



CURADORIA

O COLETIVO
CARANDAI 25
É UMA CURADORIA
DE MARCAS DE MODA,
ARTE, GASTRONOMIA
E DECORAÇÃO.

CRIATIVIDADE,

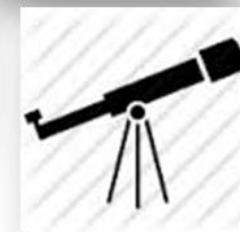
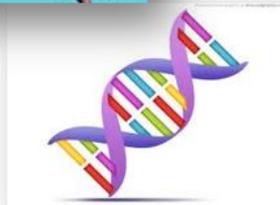
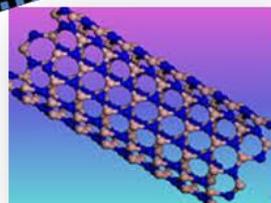
CURADORIA

CURADOR

ALGUÉM QUE MANTÉM ALGUMA
COISA PARA O BEM PÚBLICO,
CUJO VALOR FREQUENTEMENTE
NECESSITA SER REVELADO.

“

Há uma parcela dos produtos de
pesquisa que necessita de
infraestruturas
INFORMACIONAIS
TECNOLÓGICAS
POLÍTICAS
GERENCIAIS



Para se tornarem
visíveis para as comunidades
acadêmicas, Instituições de pesquisa,
agências de fomento e para o cidadão comum.

CURADORIA



Seleção, cuidado e preservação de coleções de objetos de arte estáveis, como pintura e esculturas, livros e manuscritos raros e únicos; espécimes naturais e físicos importantes, e outros artefatos considerados de valor cultural ou científico. Em termos espaciais, a curadoria ocorre em **contextos organizacionais relativamente limitados** como bibliotecas, arquivos, museus, galerias de arte, herbários e instituições similares (NRC, 2015).

Além do mais, curadoria está preocupada também em promover a disponibilidade dos objetos para audiências apropriadas (JISC).

OBJETOS FÍSICOS
OBJETOS ESTÁVEIS
OBJETOS DE CONTORNOS DEFINIDOS
AMBIENTES LIMITADOS



Necessidade da ampliação do conceito de curadoria para que compreenda infindável diversidade dos artefatos digitais e de seus usuários, e os ambientes em mutação onde eles precisam ser gerenciados.

OBJETOS DIGITAIS
OBJETOS MUTÁVEIS
OBJETOS DISTRIBUÍDOS
OBJETOS LIGADOS
AMBIENTES ILIMITADOS



A **CURADORIA DIGITAL** difere, em termos de significado e amplitude conceitual, da **CURADORIA** como ela vem sendo compreendida ao longo do tempo!

Entretanto, a **curadoria digital** mostra alguma continuidade com as **práticas tradicionais de curadoria!**



Independente de uma coleção **ser constituídas de objetos físicos ou digitais** – ou seja, de átomos e moléculas ou de bits e bytes - um curador deve **avaliar seu valor e relevância para a comunidade de usuários reais e potenciais**; determinar a **necessidade de preservação**; **documentar a origem e autenticidade**; **descrever, registrar e catalogar seu conteúdo**; **providenciar armazenamento e preservação a longo prazo**; e proporcionar um **meio de acesso e uso para os conteúdos** (NRC, 2015).



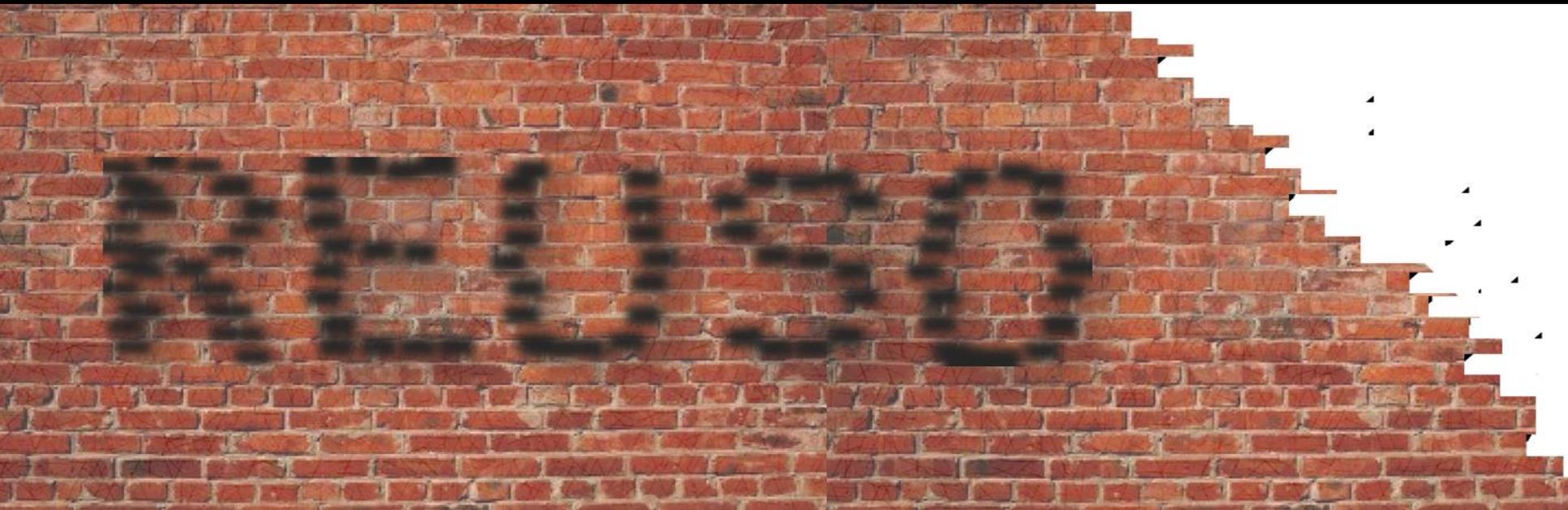
MUSEU DE HISTÓRIA NATURAL



Antes das práticas acadêmicas se deslocarem para o reino digital ou para o paradigma do *big data*, os museus de história natural já tinham ampliado o seu conceito de curadoria antecipando a demanda por gestão e aprimoramento dos dados digitais (PALMER et al., 2013, p. 2).

TEORIA DOS OBJETOS DIGITAIS







REUSO

AGREGAÇÕES

**ESPAÇOS
COLABORATIVOS**

**CURADORIA
ON-LINE**

EDUCAÇÃO

**PESQUISA
CIENTÍFICA**

**APLICATIVOS
COPUTACIONAIS**

REUSO

EM OUTROS CONTEXTOS

LIMITES TEMPORAIS

diários de bordos de navios do século XVII digitalizados e depois analisadas por climatologista do século XX

LIMITES SETORIAIS

epidemiologistas examinam dados comerciais sobre consumo em busca de remédios para a gripe

LIMITES DISCIPLINARES

pesquisadores em bioinformática combinam coleções de dados originados no domínio da biologia, genética e engenharia

REUSO DE DADOS DE PESQUISA

A probabilidade de uma coleção de dado ser reusada no futuro por outras audiências, estabelece o **critério mais simples de valor para a coleção**. Embora não seja algo simples, a partir daí pode-se estimar se vale pena arquivá-la por longo prazo



Tornar um conteúdo que foi criado para uma audiência útil para outra é um problema complexo, porque cada disciplina tem seu próprio **vocabulário, estrutura de dados e práticas de pesquisa** e formulam **questões** de forma distintas usando sua própria **terminologia**. Isto coloca um desafio importante para os **serviços de curadoria** que é criar descrições e representações, ferramentas e serviços que tornem viável o compartilhamento entre diferentes audiências (BORGMAN,2007).

AFINAL,
O QUE É DADO
DE PESQUISA?

AFINAL, O QUE É DADO DE PESQUISA?



Informação é um conceito complexo com centenas de definições [...]. Dado [por sua vez] é um conceito simples com poucas definições, porém sujeito a muitas e diferentes interpretações



CRISTINE BORGMAN (2007, P.9)

O que dificulta atribuir uma definição consensual ao dado de pesquisa é o fato idiossincrático que ele pode ser muitas coisas diferentes para pessoas e circunstâncias diferentes. Isto acontece porque dado de pesquisa é dependente de interpretação

AFINAL, O QUE É DADO DE PESQUISA?



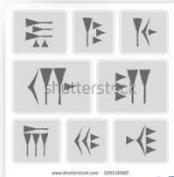
uma **sequencia de bits** proveniente de um sensor sísmico é dado de pesquisa para os sismólogos;



amostras de rochas são dados de pesquisa para um geomorfologista;



conversas gravadas são dados de pesquisa para sociólogos;

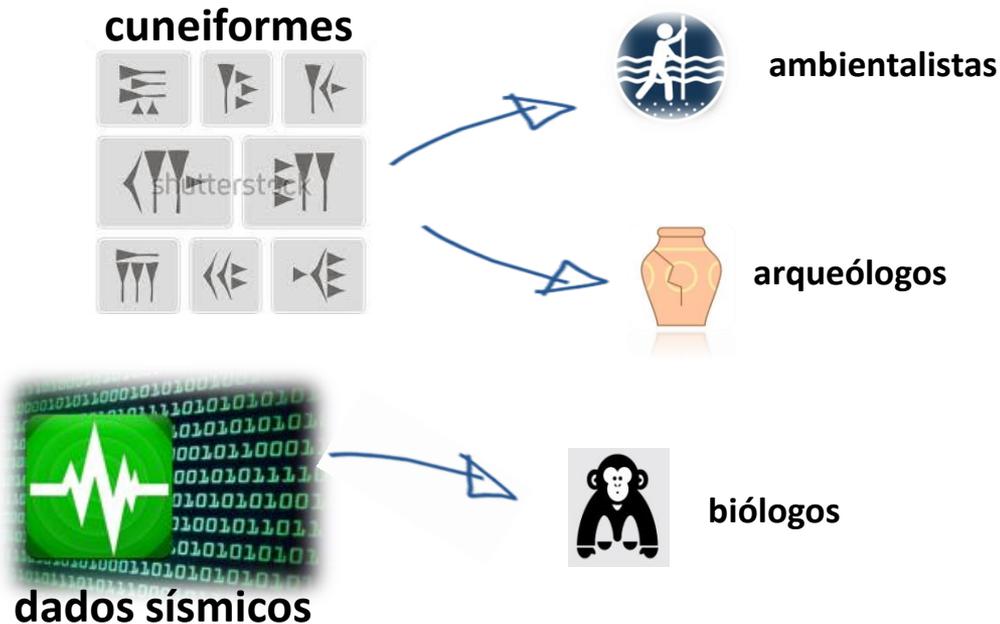


e inscrições em cuneiformes são dados de pesquisa para quem estuda linguagens do Oriente Próximo.

**AFINAL,
O QUE É DADO
DE PESQUISA?**



**... mas podem ser
reinterpretados em outros
contextos**



Porém, os **cuneiformes** podem ser também dados para o **arqueólogo** ou para o **ambientalista** que buscam **padrões climáticos históricos**; de forma similar, os **dados sísmicos** podem ser úteis para **biólogos** que estudam **comportamento animal**.

Borgman (2007, p.119)

“Dados são sempre registrados tomando como base de algum interesse, perspectiva, tecnologia e prática que determinam seus significados e utilidades em diferentes contextos”

Nielsen e Hjørland (2014, p.225)

AFINAL, O QUE É DADO DE PESQUISA?

“ DADOS DE PESQUISA SÃO GERADOS PARA DIFERENTES PROPÓSITOS, POR DIFERENTES COMUNIDADES CIENTÍFICAS E POR MEIO DE DIFERENTES PROCESSOS



AFINAL, O QUE É DADO DE PESQUISA?

REGISTRO DE FATOS USADOS
COMO FONTES PRIMÁRIAS NA
INVESTIGAÇÃO CIENTÍFICA E
QUE GERALMENTE SÃO
ACEITOS NA COMUNIDADE
CIENTÍFICA COMO
NECESSÁRIOS PARA A
VALIDAÇÃO DOS RESULTADOS
DE PESQUISA
(OCDE, 2007).

"Data are facts, observations or experiences on which an argument or theory is constructed or tested. Data may be numerical, descriptive, aural or visual. **Data may be raw, abstracted or analysed, experimental or observational.** Data include but are not limited to: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses. Research collections may include slides; artefacts; specimens; samples."

Algumas definições

"Research data is defined as **recorded factual material** commonly retained by and accepted in the scientific community as **necessary to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created."

"The data, records, files or other evidence, **irrespective of their content or form** (e.g. in print, digital, physical or other forms), that comprise research observations, findings or outcomes, including primary materials and analysed data."

DADO DE PESQUISA É TODO E QUALQUER TIPO DE REGISTRO COLETADO, OBSERVADO, GERADO OU USADO PELA PESQUISA CIENTÍFICA, TRATADO E ACEITO PELA COMUNIDADE CIENTÍFICA COMO NECESSÁRIO PARA VALIDAR OS RESULTADOS DE PESQUISA

ORIGENS DOS DADOS

DADOS OBSERVACIONAIS são obtidos de observações diretas, tais como erupção de um vulcão numa data específica, a atitude dos eleitores ou fotografia de uma supernova – que constituem enfim registros históricos que não podem ser coletados uma segunda vez e, portanto, devem ser arquivados para sempre

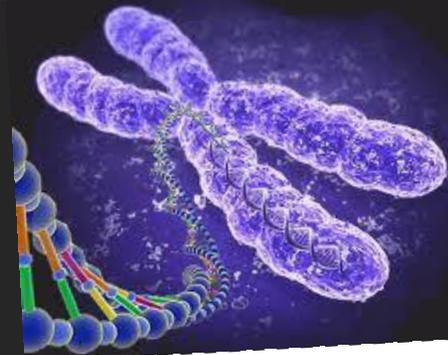


CRÍTICOS



DADOS EXPERIMENTAIS são provenientes de situações controladas em bancadas de laboratórios. Em tese, dados experimentais provenientes de experimentos que podem ser precisamente reproduzidos e não precisam ser armazenados indefinidamente; entretanto, nem sempre é possível reproduzir precisamente todas as condições experimentais.

DADOS COMPUTACIONAIS – resultados da execução de modelos computacionais ou de simulações; devem ser submetidos a uma abordagem distinta que pressupõe o arquivamento de um grande número de informações, expressos por um conjunto robusto de metadados, que incluem descrição de hardware, software e dados de entrada





DADOS GOVERNAMENTAIS

Dados provenientes de **recenseamento**, **seguro social**, **levantamentos de agências governamentais** etc. são críticos para as pesquisas nas áreas de saúde, ciências sociais e humanidades.



REGISTROS

Registros médicos críticos para as pesquisas nas áreas de saúde e ciências sociais. **Registros arquivísticos, históricos, jornalísticos, administrativos, etc**



DADOS ACUMULADOS POR REDES SOCIAIS, MÁQUINAS DE BUSCA, ETC.

Big data

Comércio Transacional baseado em dados de negócios.
Humanidades digitais

FLUXO DOS DADOS

A MAIOR PARTE DOS DADOS NÃO É DIRETAMENTE ÚTIL NO MOMENTO EM QUE COLETADA

INSTRUMENTO CIENTÍFICO



dreamstime.com

DADOS BRUTOS



PROCESSAMENTO DOS DADOS



Selecionar subset
Mesclar múltiplos datasets
Conversão
Normalização
Limpeza dos dados

COMPUTAÇÃO EM NUVEM
COMPUTAÇÃO EM GRADE

ANÁLISE DOS DADOS

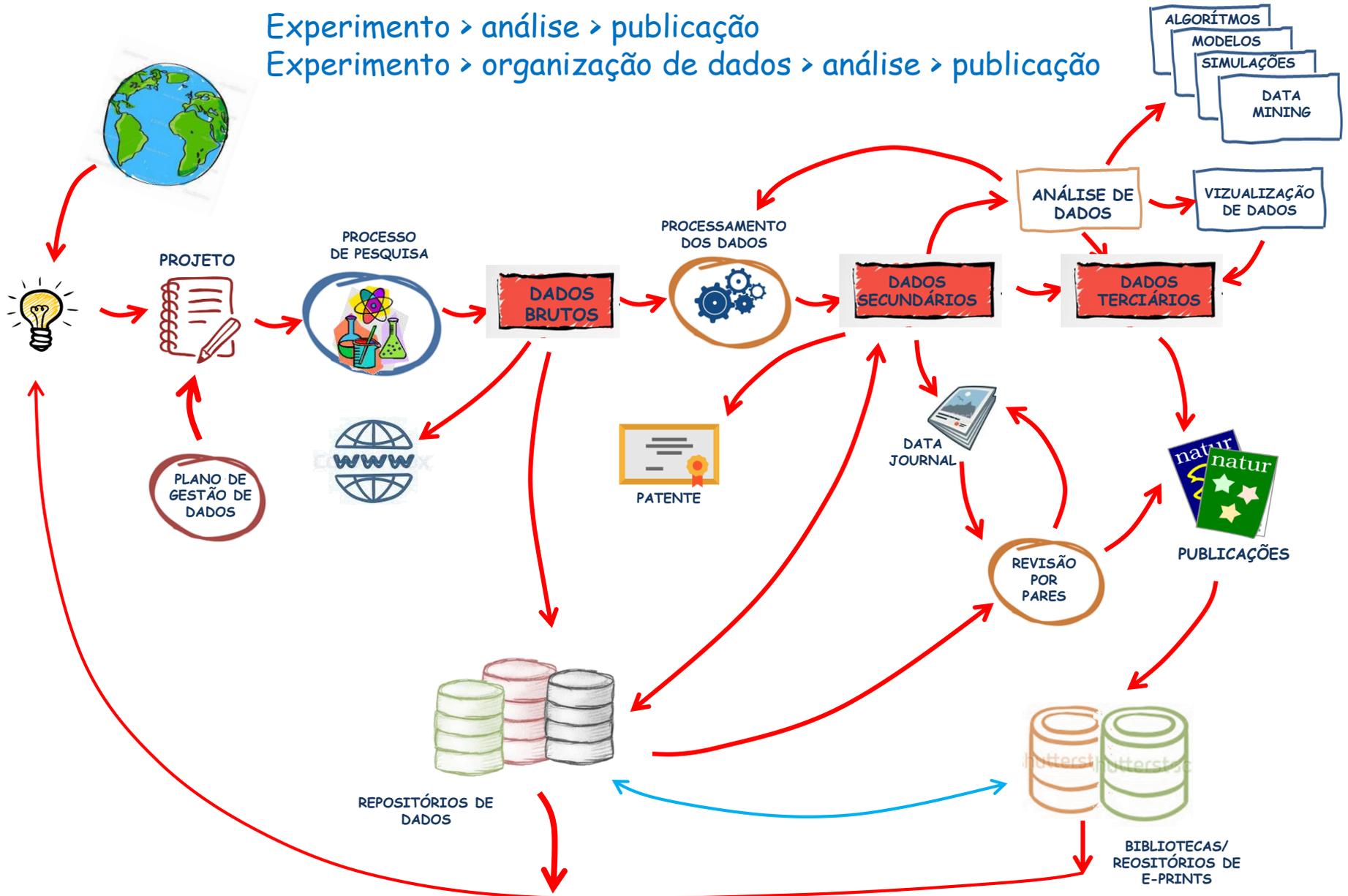


Estatísticas
Simulação
Plotagem
Visualização
Modelos
Algoritmos
Publicações



Experimento > análise > publicação

Experimento > organização de dados > análise > publicação



PROJETO DE PESQUISA

CAPTURA DE DADOS

LIMPEZA DOS DADOS

ANÁLISES E RESULTADOS

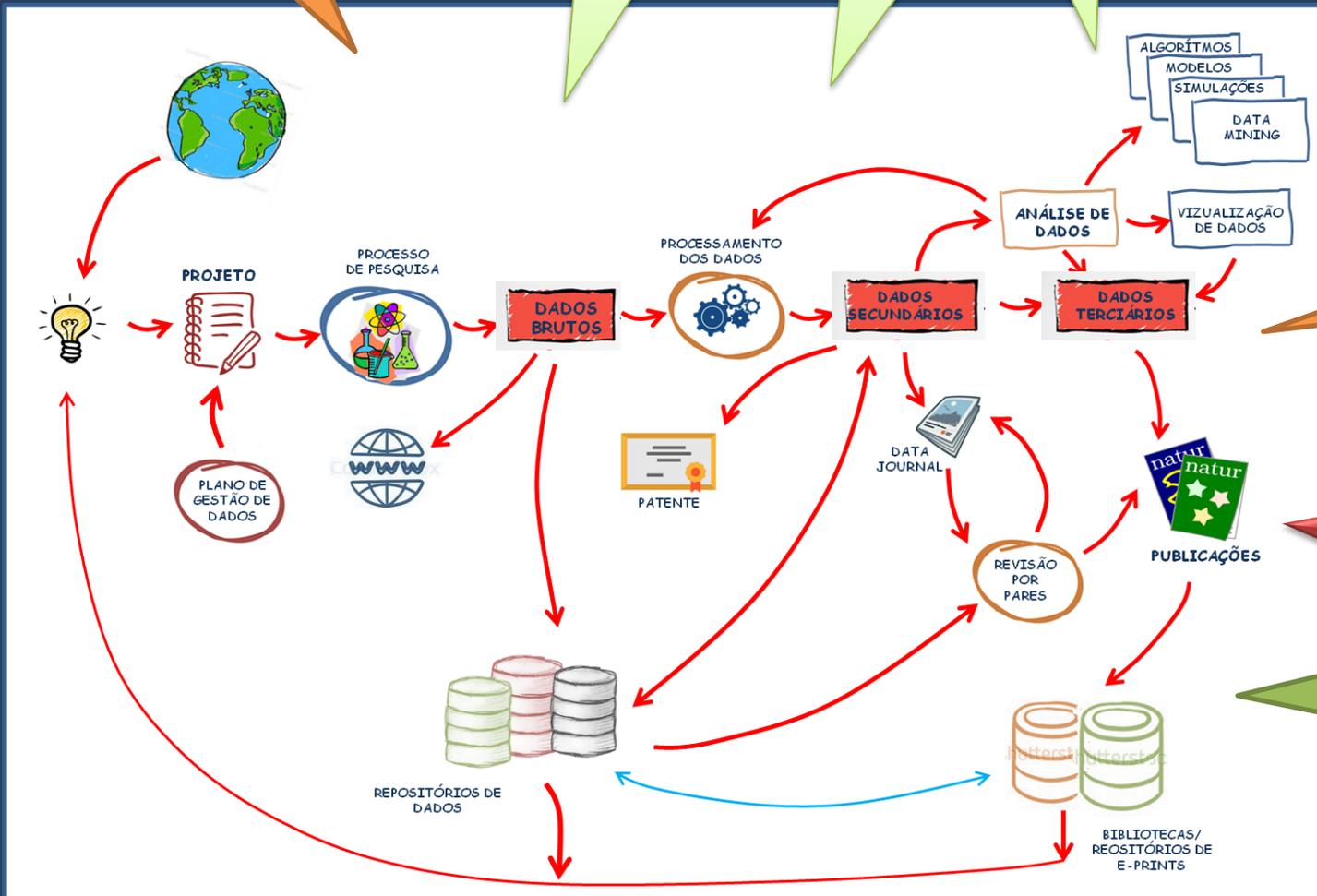
FORMULAÇÃO DE HIPÓTESES

PROCESSOS COMPUTACIONAIS

CATALOGAÇÃO

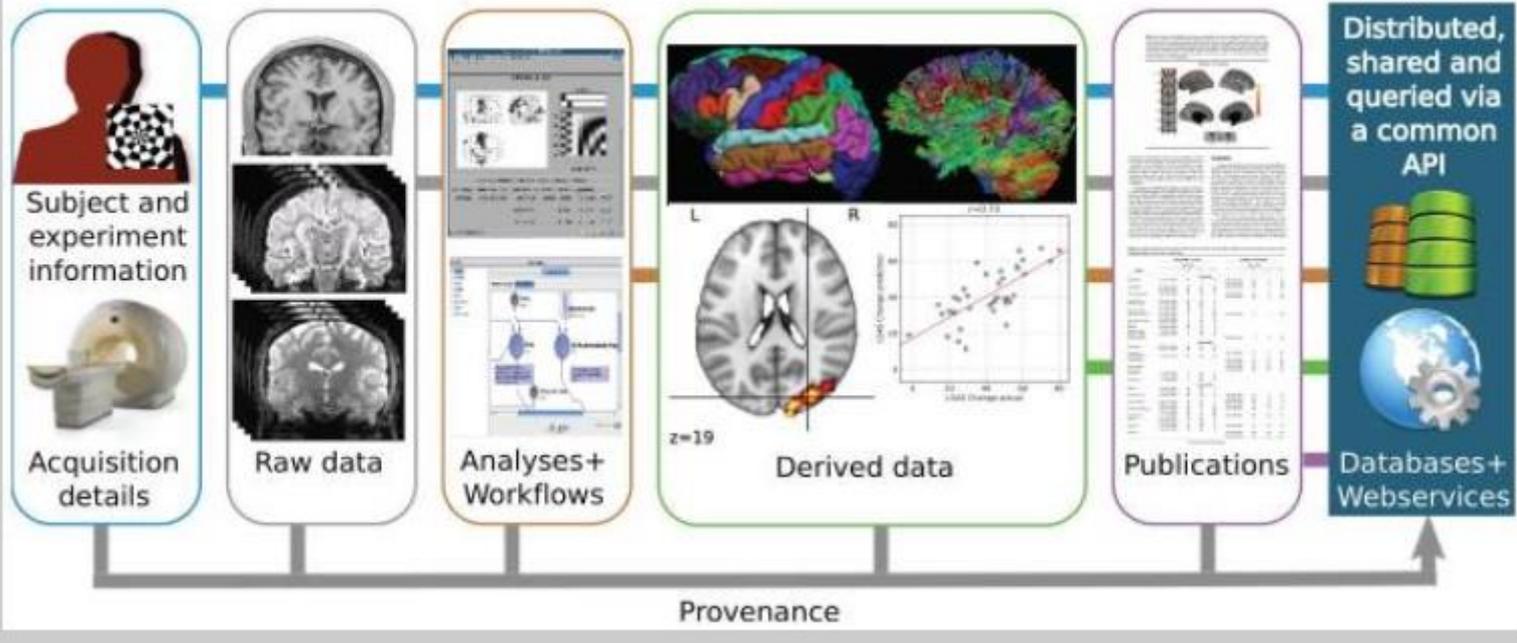
INTEGRAÇÃO COM O SISTEMA DE PUBLICAÇÃO

CURADORIA ARQUIVAMENTO PRESERVAÇÃO

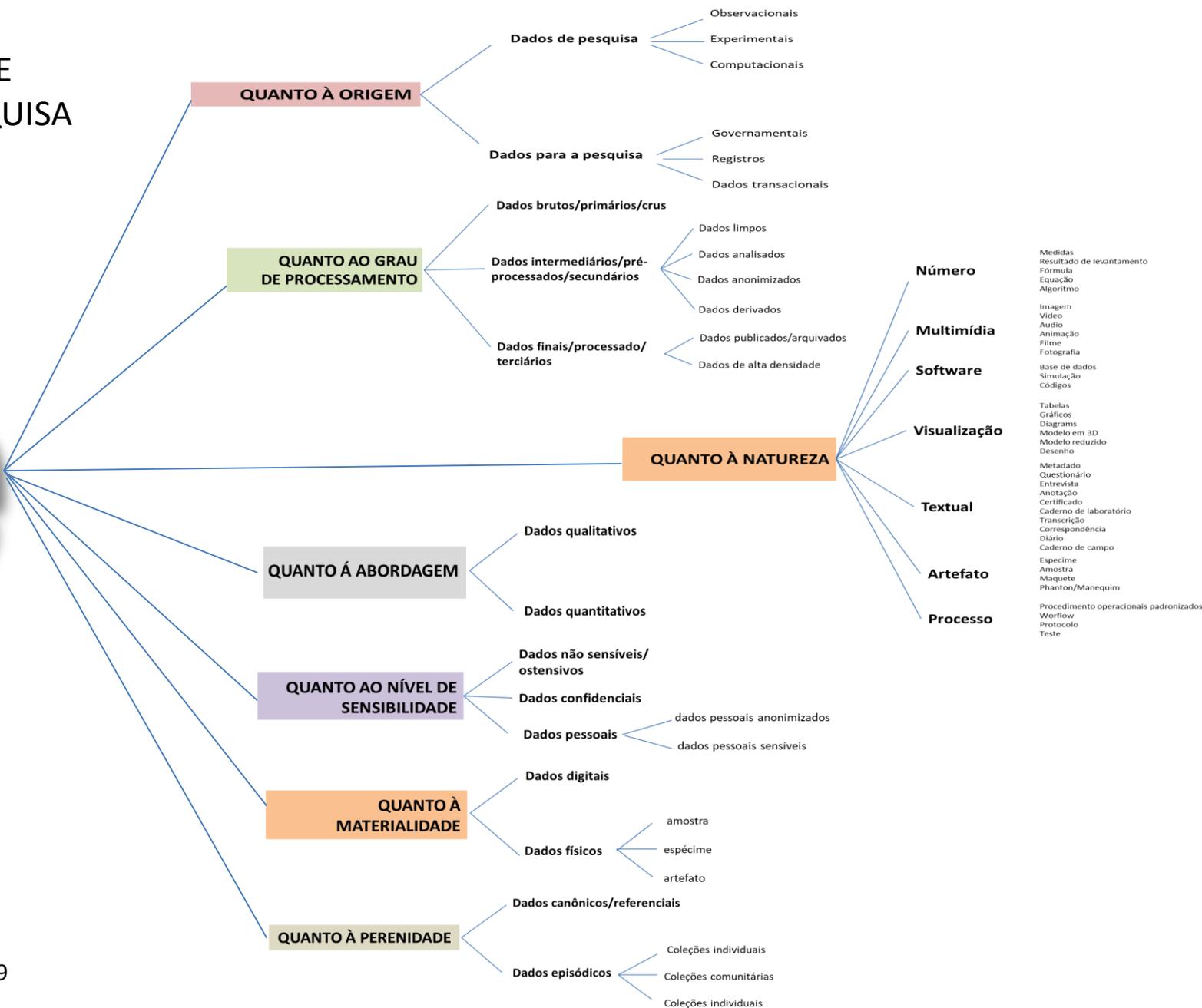
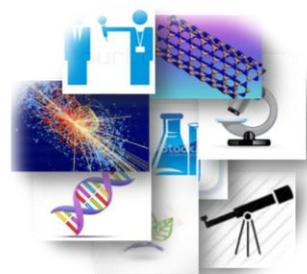


POLÍTICAS – SUSTENTABILIDADE – CONFORMIDADE LEGAL E ÉTICA

Stages of Electronic Data Capture



TAXONOMIA DE DADO DE PESQUISA VOLTADA PARA CURADORIA

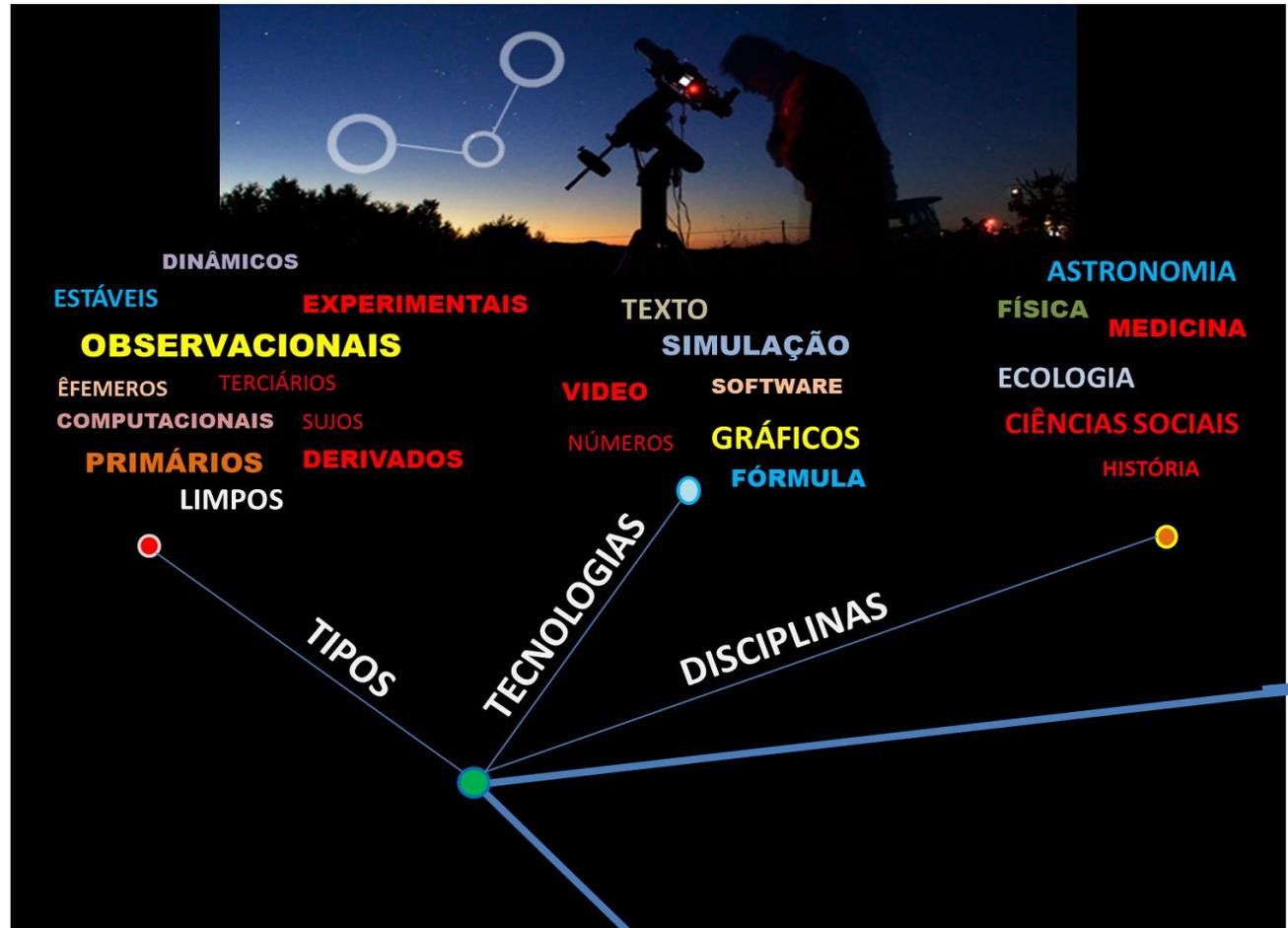


**POR QUE A GESTÃO
DE DADOS DE
PESQUISA
NECESSITA DE
NOVAS
ABORDAGENS?**

DADOS DE PESQUISA DEMANDAM UM GESTÃO MAIS COMPLEXA

Em comparação com a gestão de artigos, livros e teses as funções, as descrições, os padrões e os controles são mais **numerosos e complexos**.

Essa complexidade, no entanto, varia de acordo com os ambientes disciplinares, os tipos de objetos e as tecnologias subjacentes considerados e com a política adotada pela instituição.



GERENCIANDO O INVISÍVEL



Pela primeira vez em 3.500 anos de atividade de gestão de documentos, produzimos registros que não existem para o olho humano. Completamente diferentes das placas de argila da Babilônia, dos papiros egípcios, dos pergaminhos romanos, do papel moderno e mesmo do microfilme

E pela primeira não estamos produzindo, gerenciando e guardando artefatos físicos, mas tentando entender e preservar padrões virtuais que dão a informação digital seu conteúdo, estrutura, contexto e assim o seu significado, que são completamente controlados por software

ORIENTADA POR DADOS

**BIG
SCIENCE**

GRANDES INSTRUMENTOS
ALTOS CUSTOS
LONGA DURAÇÃO
MUITOS COLABORADORES
PESQUISA DISTRIBUÍDA

**SMALL
SCIENCE**

PEQUENOS INSTRUMENTOS
BAIXOS CUSTOS
PEQUENA DURAÇÃO
EQUIPES PEQUENAS
PESQUISA LOCAL

ORIENTADA POR HIPÓTESES

PIRÂMIDE DE GESTÃO DE DADOS





A CAUDA LONGA DA CIÊNCIA

A MAIORIA DAS COLEÇÕES DE DADOS PRODUZIDAS PELA PESQUISA CIENTÍFICA É GERADO/COLETADO POR PEQUENOS LABORATÓRIOS E PESQUISADORES INDIVIDUALMENTE NAS UNIVERSIDADES E INSTITUTOS DE PESQUISA, QUE DESENVOLVEM UM GRANDE NÚMERO DE PROJETOS CIENTÍFICOS

DOMÍNIOS ESPECÍFICOS

ASTRONOMIA
FÍSICA NUCLEAR
GENOMA
PROTEÍNA
SENSORIAMENTO REMOTO



Dados da Grande Ciência são fáceis de manipular, compreender e arquivar;
A Pequena Ciência é terrivelmente heterogênea e muito mais vasta e gera 2-3 vezes mais dados do que a Big Science (MacColl, 2010)

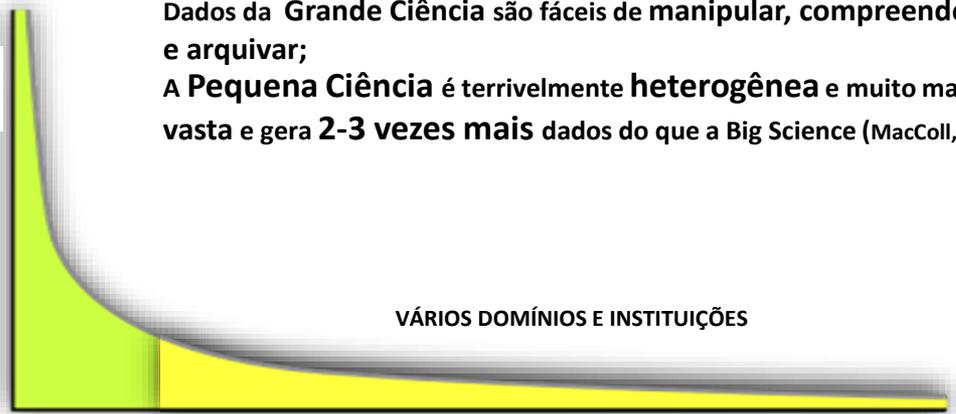
Volume dos dados



Número de datasets

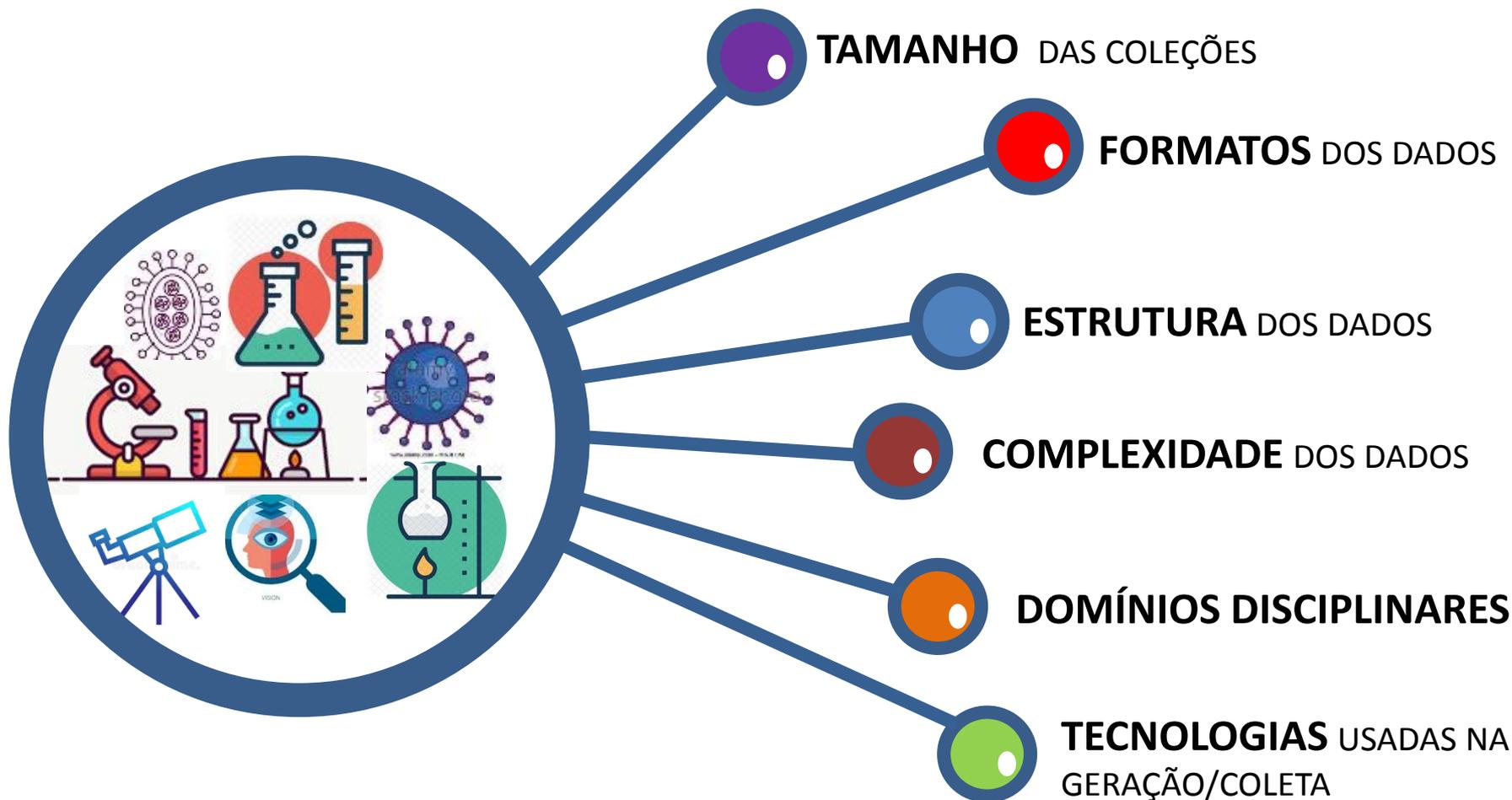


VÁRIOS DOMÍNIOS E INSTITUIÇÕES



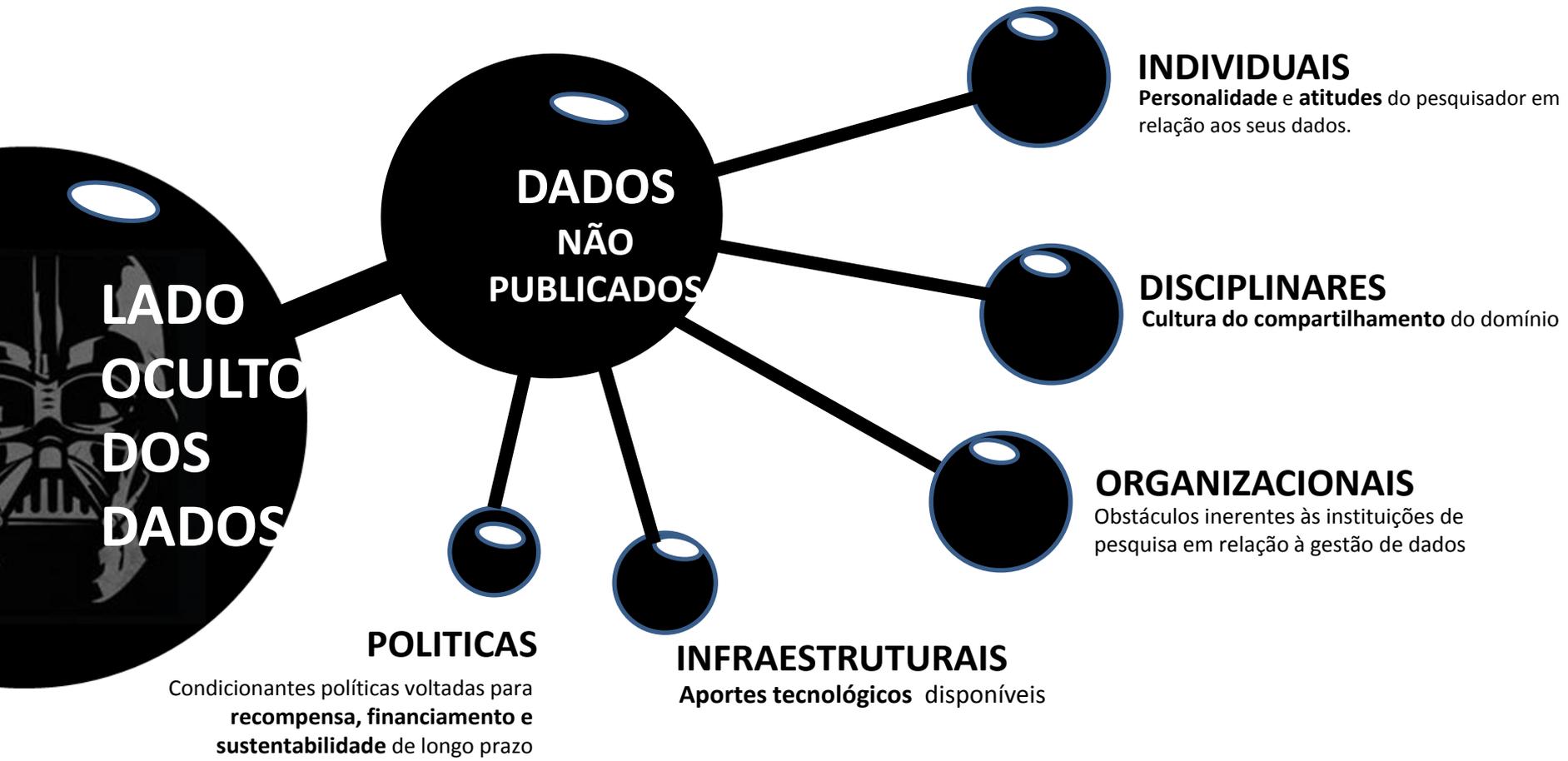
PEQUENOS LABORATÓRIOS, EQUIPES E PESQUISADORES INDIVIDUAIS

DADOS DA CAUDA LONGA: HETEROGENEIDADE EM VÁRIAS DIMENSÕES



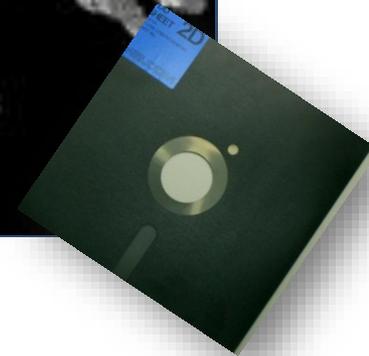
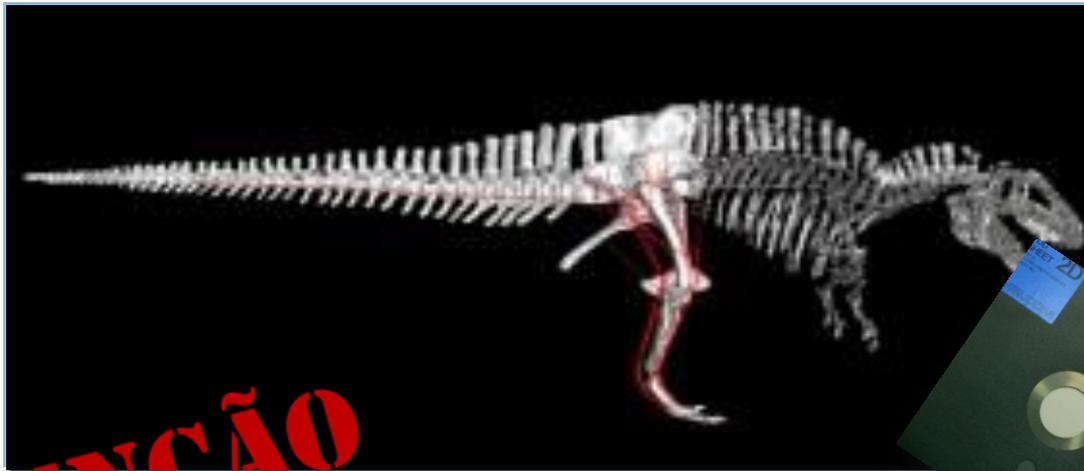
+50%

DOS ACHADOS NÃO FORAM PUBLICADOS



O COMPRTILHAMENTO PODE REVELAR VALORES IMPORTANTES OCULTOS NESSES DADOS

A INFORMAÇÃO DIGITAL NÃO SOBREVIVE INERCIALMENTE



INTENÇÃO

[A mesma tecnologia que muda a pesquisa científica coloca os dados gerados em risco e nos impõe o desafio estratégico, gerencial e político de criar, arquivar, preservar e tornar disponível esses dados

04

27

56

01

16

44

02

01

17

POR QUE?

QUEM?

O QUE?

COMO?

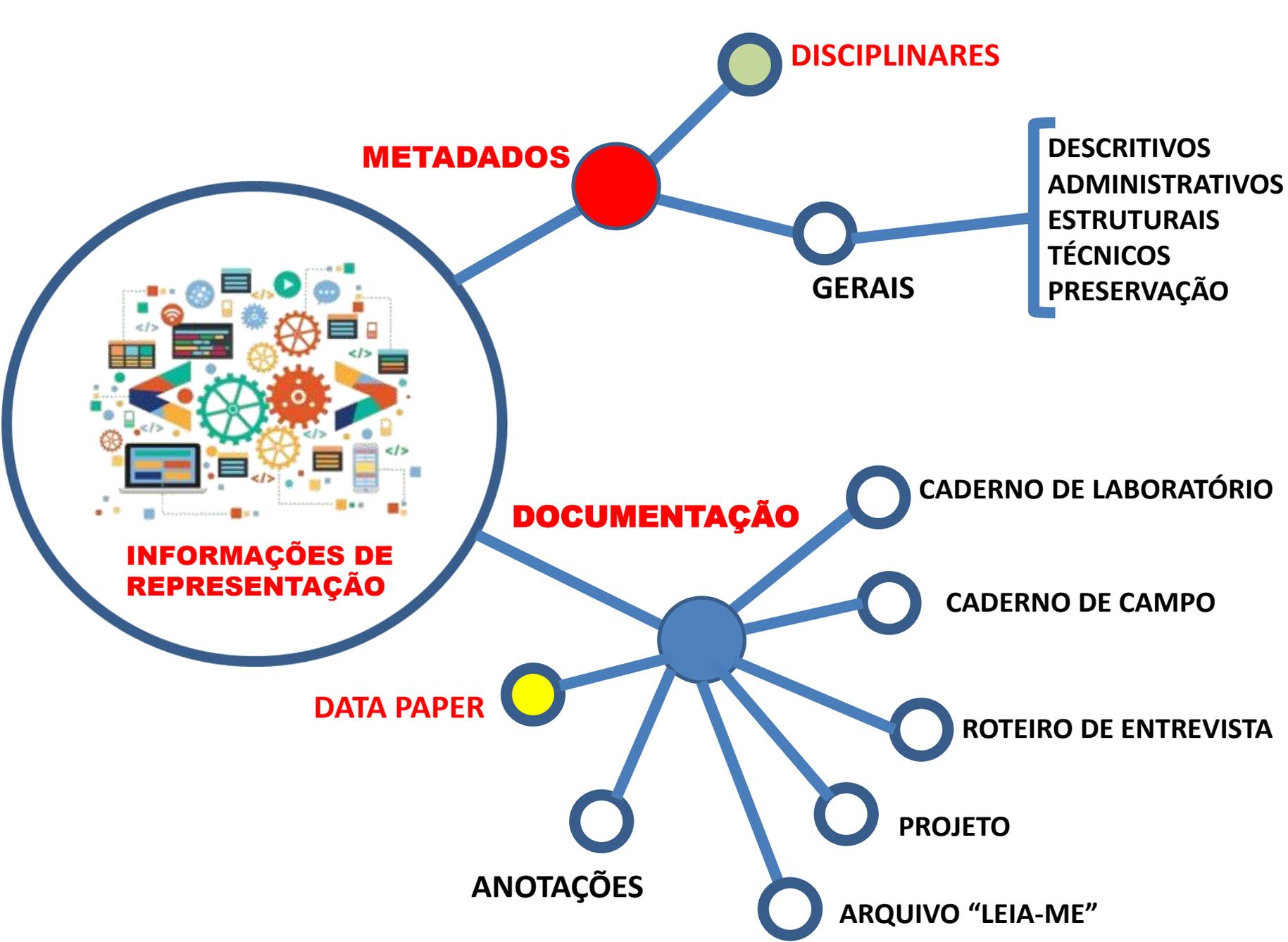
QUANDO?

ONDE?



**SIGNIFICADO
ESTRUTURA
IDENTIFICAÇÃO
CONTEXTO**

PROVENIÊNCIA





REFERÊNCIA

A capacidade das coleções de dados e suas versões hospedadas nos repositórios de serem **IDENTIFICADAS** permanentemente torna-se essencial para o **acesso, preservação e citação**; é um fator importante também nos processos de **interoperabilidade** e de **linking** com outros recursos via, por exemplo, *linked data*.

IDENTIFICADORES PERSISTENTES

DOI

URN

HANDLES

Específicos

CONTROLE DE VERSÕES

UFG – UNIVERSAL FINGERPRINT

TIMESTAMPING

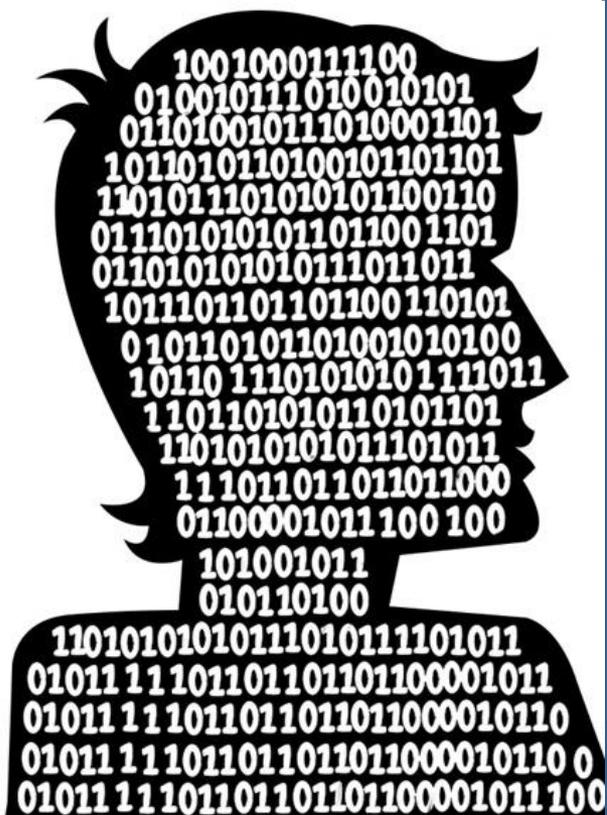
CITAÇÃO PADRONIZADA

FERRAMENTAS DE APOIO À CITAÇÃO

EXPORTAÇÃO EM FORMATOS DIVERSOS/COMPARTILHAMENTO

O controle de versões é um processo importante para o fundamento da reprodutibilidade da pesquisa, para a integridade da referência às coleções de dados e para proveniência dos seus conteúdos. Isto por que as coleções de dados podem evoluir no tempo por vários motivos

Questões éticas e de consentimento



Coletar, usar e compartilhar dados no âmbito de pesquisas que envolvam **peças exige que obrigações éticas e legais sejam respeitadas.**

Quando a pesquisa envolve obter dados de pessoas, o que se espera do pesquisador é que ele mantenha um comportamento pautado por um rigoroso código de ética, que seja condizente com os padrões e protocolos recomendados pelas entidades profissionais, instituições de pesquisa e organizações financiadoras de pesquisa e, sobretudo, com a legislação do país concernente a esse aspecto.

Este comportamento deve permear todo o ciclo de pesquisa, incluindo especialmente a fase de compartilhamento dos dados.



Nesse contexto a compreensão de três tipos de dados se torna essencial:

DADOS PESSOAIS

São dados relacionados a indivíduos vivos, que podem ser identificados a partir desses dados ou a partir desses dados combinados com outras informações.

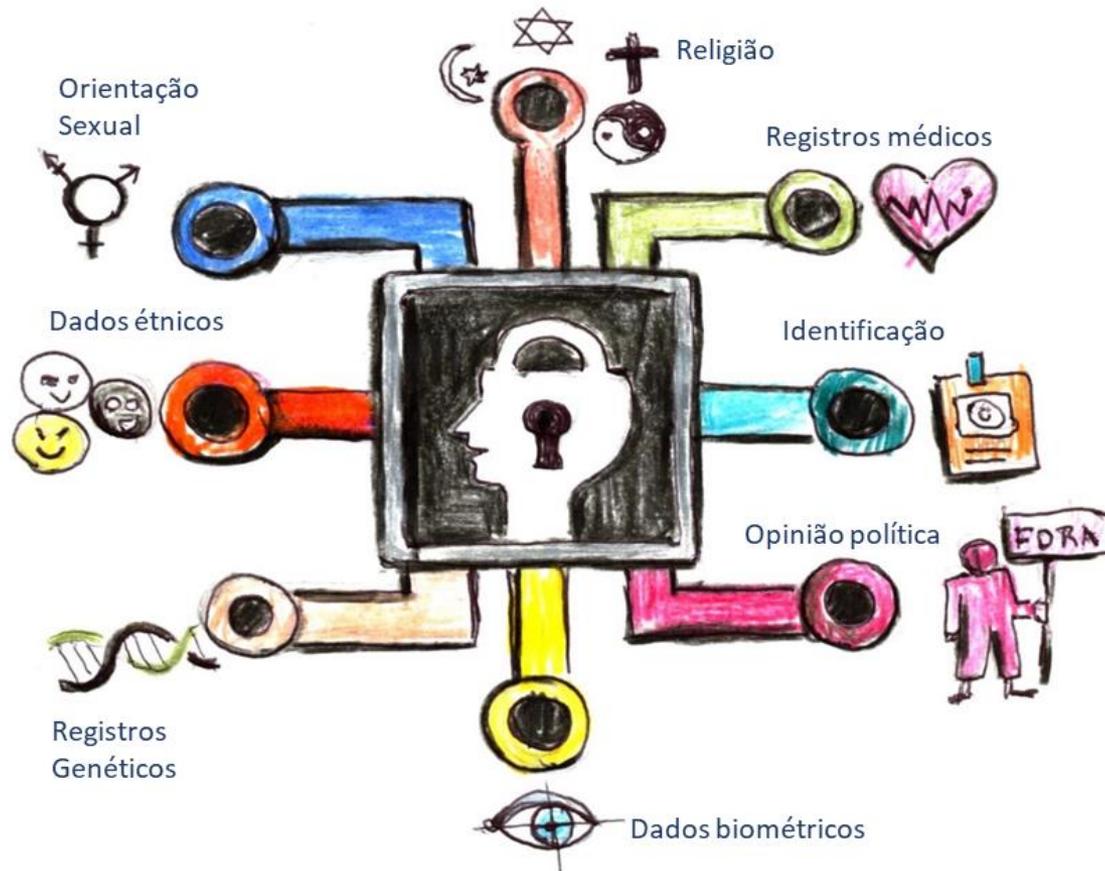
DADOS CONFIDENCIAIS

São dados que não estão em domínio público tais como informações sobre negócios, lucros, saúde, detalhes médicos e opiniões políticas, entregues em confiança ou que duas partes concordam em mantê-los confidenciais, isto é, secretos.

DADOS PESSOAIS SENSÍVEIS

São dados sobre raça, origem étnica, opinião política, religião ou crenças similares, filiação sindical, doença física ou mental, vida sexual, etc.

DADOS PESSOAIS SENSÍVEIS



CONTEÚDO

ESTRUTURA

ESTRUTURA + CONTEÚDO



**LICENÇAS APROPRIADAS AO
COMPARTILHAMENTO E REUSO**

**CURADORIA
DE DADOS
DE PESQUISA**

CURADORIA DIGITAL DE DADOS DE PESQUISA

“

Manutenção, preservação e agregação de valor a dados de pesquisa **durante o seu ciclo de vida.** (DCC, 2003)

“

Todas as atividades envolvidas na **gestão de dados**, desde o planejamento de sua criação – quando os sistemas são projetados – passando pelas boas práticas de digitalização, na seleção dos formatos e da documentação, e na garantia de estarem sempre adequados para serem **descobertos e reusados agora e no futuro.** (ABBOTT, 2008)

**GESTÃO DINÂMICA E
APRIMORAMENTO DOS ATIVOS DE
INFORMAÇÃO DIGITAIS
PARA USO ATUAL E FUTURO**

ANOTAÇÃO, PRESERVAÇÃO E DESCRIÇÃO DE COLEÇÕES DE DADOS POR ESPECIALISTA

GERENCIAR E PROMOVER O USO DE DADOS A PARTIR DO SEU PONTO DE CRIAÇÃO PARA ASSEGURAR QUE ELE ESTA APTO PARA PROPÓSITOS CORRENTES E **DISPONÍVEL PARA DESCOBERTA E REUSO.**

PROCESSOS NECESSÁRIOS PARA UMA **BOA CRIAÇÃO E GESTÃO DE DADOS** E A CAPACIDADE DE **ADICIONAR VALOR** AO DADO PARA **GERAR NOVAS FORMAS DE INFORMAÇÃO E CONHECIMENTO**

ESTÁ PREOCUPADA COM A **GESTÃO DE LONGO PRAZO DOS DADOS** DA **SUA COLETA INICIAL AO ARQUIVAMENTO NUM REPOSITÓRIO**

MANTER E ADICIONAR VALOR A UM CORPO FIDEDIGNO DE INFORMAÇÃO DIGITAL PARA **USO CORRENTE E FUTURO.**

MANTER A INFORMAÇÃO DIGITAL QUE É PRODUZIDA NO DECORRER DA PESQUISA DE UMA MANEIRA QUE PRESERVE SEU **SIGNIFICADO E UTILIDADE** COMO **UM INSUMO POTENCIAL PARA OUTRAS PESQUISAS**

ENVOLVE **MANTER LINKS COM ANOTAÇÕES E COM MATERIAIS PUBLICADOS**

CONTÍNUO ENRIQUECIMENTO E ATUALIZAÇÃO PARA MANTER OS DADOS ADEQUADOS AOS PROPÓSITOS DE TRANSMITIR CONHECIMENTO

DESCONSTRUINDO O CONCEITO DE CURADORIA DE DADOS DE PESQUISA

**AÇÕES
GERENCIAIS,
TECNOLÓGICAS E
POLÍTICAS**

**NECESSÁRIAS PARA
MANTER OS DADOS
POR TODO O SEU CICLO
DE VIDA – DESDE A SUA
CRIAÇÃO -VISANDO O
USO CORRENTE E
FUTURO**

**QUE PRESSUPÕE
ADICIONAR VALOR**

ORGANIZAÇÕES EM COLEÇÕES
DOCUMENTAÇÃO
ATRIBUIÇÃO DE METADADOS
IDENTIFICAÇÃO
ARQUIVAMENTO
PRESERVAÇÃO
SEGURANÇA FÍSICA
AVALIAÇÃO (AUTENTICAÇÃO E
VERIFICAÇÃO)
CONTROLE DE QUALIDADE
ANOTAÇÃO
CONTEXTUALIZAÇÃO

PARA GARANTIR

INTELIGENTEMENTE
ABERTOS
COMPREENSÍVEIS
LONGEVOS
DISPONÍVEIS
RECUPERÁVEIS
ACESSÍVEIS
AVALIÁVEIS
(PROVENIÊNCIA/
INTEGRIDADE/
QUALIDADE)
CONFORMIDADE LEGAL
E ÉTICA
PADRONIZADOS
INTEROPERÁVEIS

**COM OBJETIVO
FINAL**

USABILIDADE/REUSO
REPRODUTIBILIDADE
INTERDISCIPLINARIDADE
INPUT PARA NOVAS
PESQUISAS
ENSINO DAS CIÊNCIAS
MEMÓRIA ACADÊMICA
VALIDAÇÃO DA PESQUISA

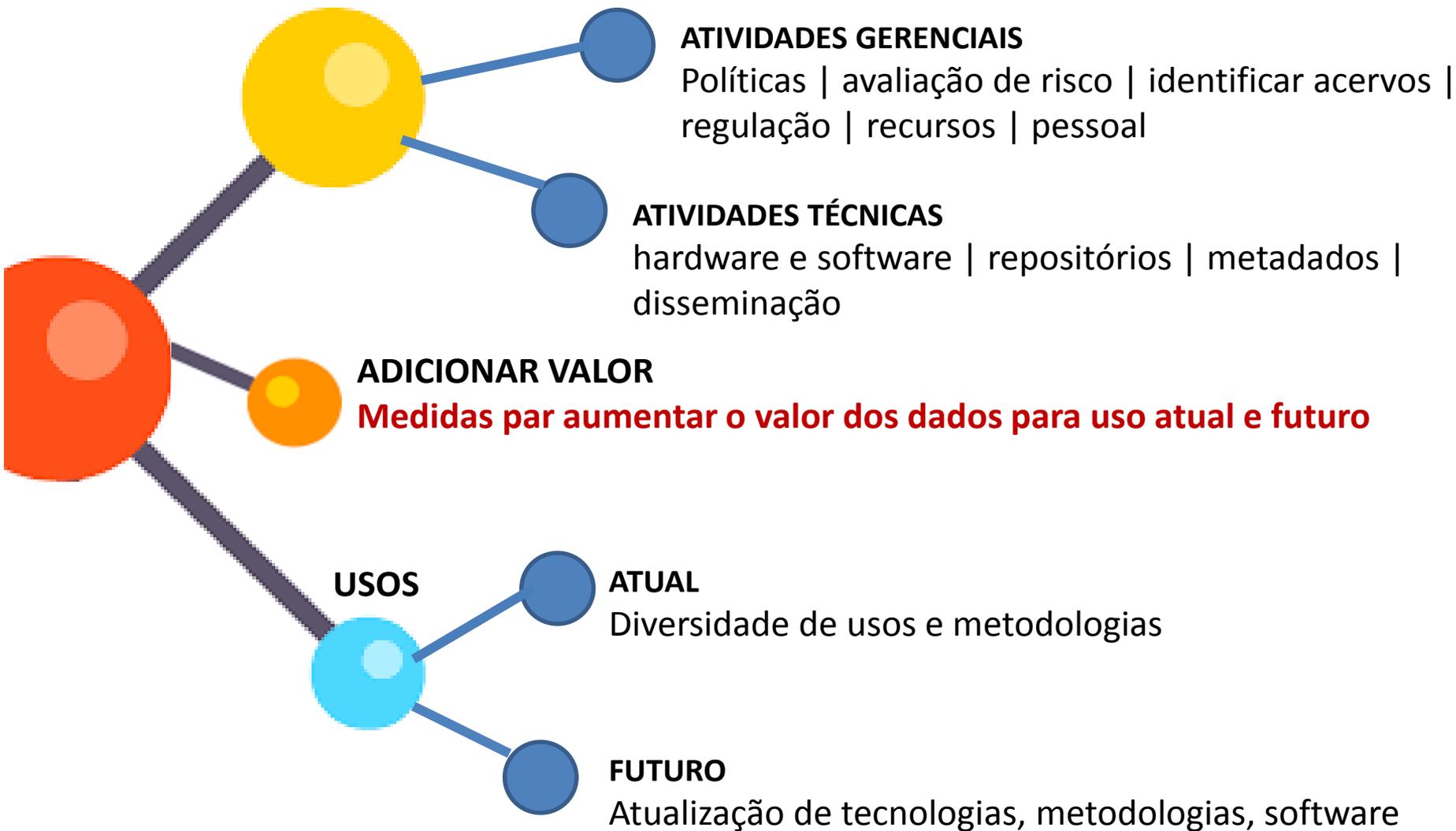
**VOLTADO PARA
UM PÚBLICO-ALVO**

**ALINHADO COM O
FLUXO DE PESQUISA**



GESTÃO ATIVA

Distingue curadoria de simplesmente coletar e armazenar dados



GESTÃO DINÂMICA E APRIMORAMENTO DOS ATIVOS **DE INFORMAÇÃO DIGITAIS** PARA USO ATUAL E FUTURO

GESTÃO DINÂMICA DE INFORMAÇÃO DIGITAL

ATIVIDADES GERENCIAIS

- Desenvolvimento de **políticas de curadoria**
- **Avaliação de riscos:** tecnologia atual e práticas de curadoria;
- **Identificação dos ativos** informacionais;
- **Avaliação dos sistemas e processos de curadoria;**
- Monitorar a conformidade legal e com as melhores práticas;
- Mobilizar recursos financeiros e técnicos
- Recrutar e treinar pessoal qualificado

ATIVIDADES TÉCNICAS

- Trabalhar diretamente com as **tecnologias** que apoiam a gestão da informação tais como **desenvolver e operar repositórios** para gestão de longo prazo de informação digital;
- **Organizar e catalogar** os ativos de informação;
- **Criar ou aprimorar metadados** associados com os objetos e datasets;
- **Disseminar os datasets;**
- **Gerenciar o acesso ao repositório e seu conteúdo.**

GESTÃO DINÂMICA E APRIMORAMENTO DOS ATIVOS DE INFORMAÇÃO DIGITAIS PARA USO **ATUAL E FUTURO**

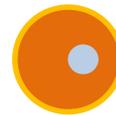
USOS ATUAL

Requer a curadoria da
informação digital para a
DIVERSIDADE de USOS e
METODOLOGIAS



USO FUTURO

Exige demandas adicionais da
curadoria para garantir a
preservação e o acesso aos dados
digitais ao longo do tempo. Implica
em atualização das tecnologias,
software, hardware e metadados



**USO
ATUAL E
FUTURO**



DESCONSTRUINDO O CONCEITO DE CURADORIA DE DADOS DE PESQUISA

CURADORIA DIGITAL ATIVIDADE DE GESTÃO E DE PROMOÇÃO DE USO DE DADO DE PESQUISA DESDE A CRIAÇÃO/SELEÇÃO, ASSEGURANDO A SUA ADEQUAÇÃO PARA PROPÓSITOS ATUAIS E FUTUROS E TORNANDO-OS **DISPONÍVEIS PARA DESCOBERTA E REUSO.**

GESTÃO DE MUDANÇAS AO LONGO DO TEMPO E DO ESPAÇO



ARQUIVAMENTO

ATIVIDADE DE CURADORIA QUE ASSEGURA QUE O DADO ESTÁ SELECIONADO, ARMAZENADO E PODE SER ACESSADO E QUE A SUA

INTEGRIDADE FÍSICA E LÓGICA ESTÁ MANTIDA AO LONGO DO TEMPO

PRESERVAÇÃO DIGITAL

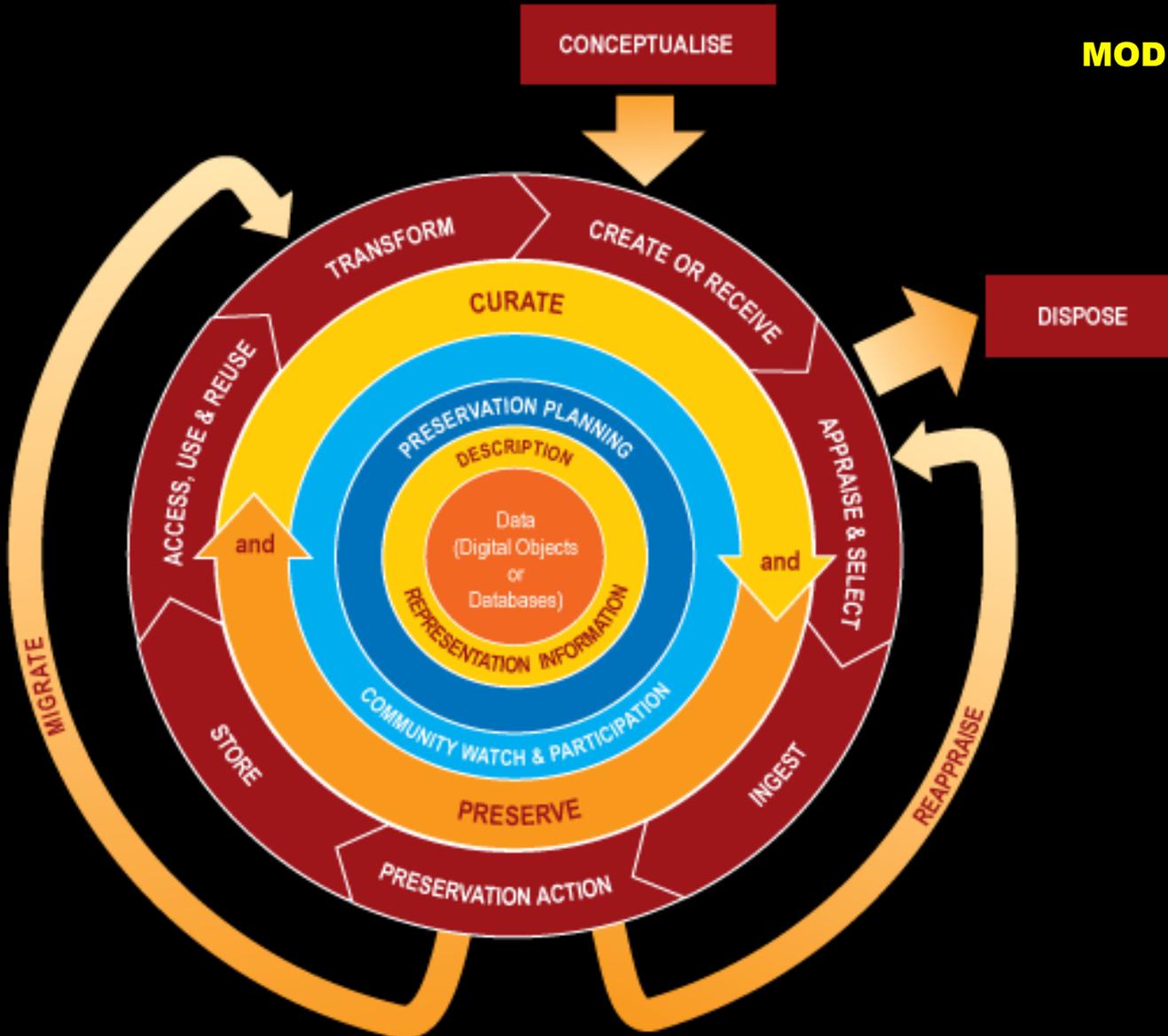
ATIVIDADE INCLUÍDA NA CURADORIA ATRAVÉS DA QUAL ITENS ESPECÍFICOS DE DADOS SÃO MANTIDOS AO LONGO DO TEMPO DE FORMA QUE ELES POSSAM SER ACESSADOS E COMPREENDIDOS NO FUTURO A DESPEITO DAS MUDANÇAS TECNOLÓGICAS



D | C | C

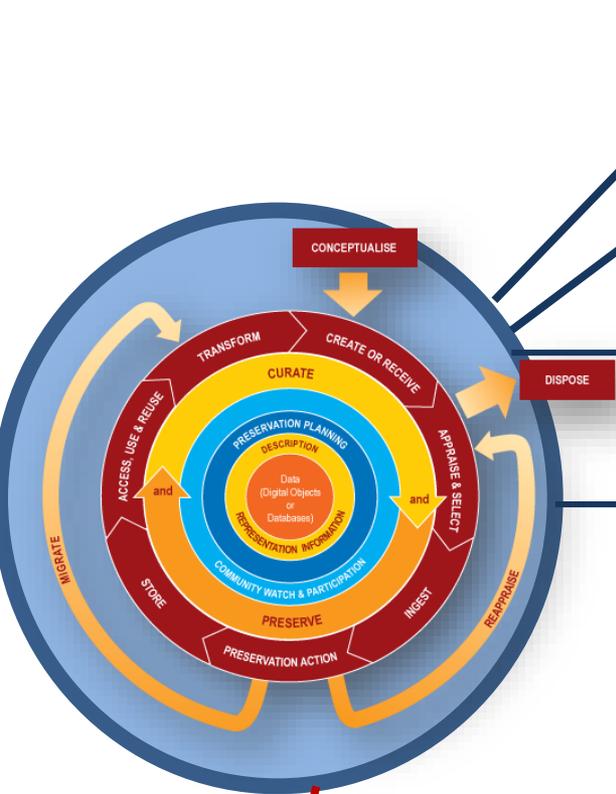
because good research needs good data

MODELO DO CICLO DE VIDA DA CURADORIA



O Modelo de Ciclo de Vida da Curadoria fornece uma visão geral /gráfica de alto nível das etapas necessárias para o sucesso da **curadoria e preservação de dados**

O modelo pode ser usado para planejar as atividades dentro da organização para garantir que todos os **passos necessários do ciclo de vida de curadoria são cobertos**



Modelo é **indicativo** e não exaustivo.

Adaptável a **diferentes domínios**

Extensível para permitir que atividades de curadoria e preservação possam ser planejadas em diferentes níveis de **granularidade**.

PODE SER USADO PARA:

- definir **papéis e responsabilidades**;
- construir **infraestruturas tecnológicas e de e padrões**;
- e assegurar que os **processos e as políticas sejam devidamente documentadas**.

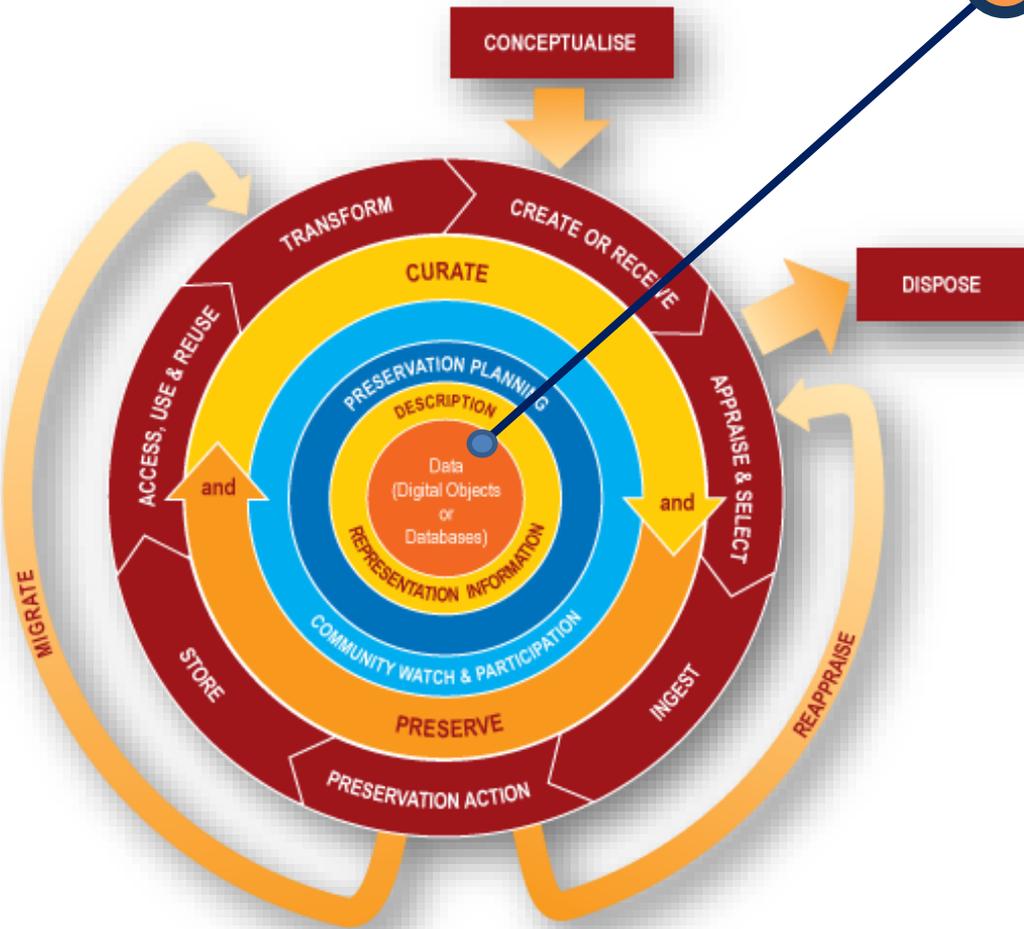
IDENTIFICA AÇÕES QUE:

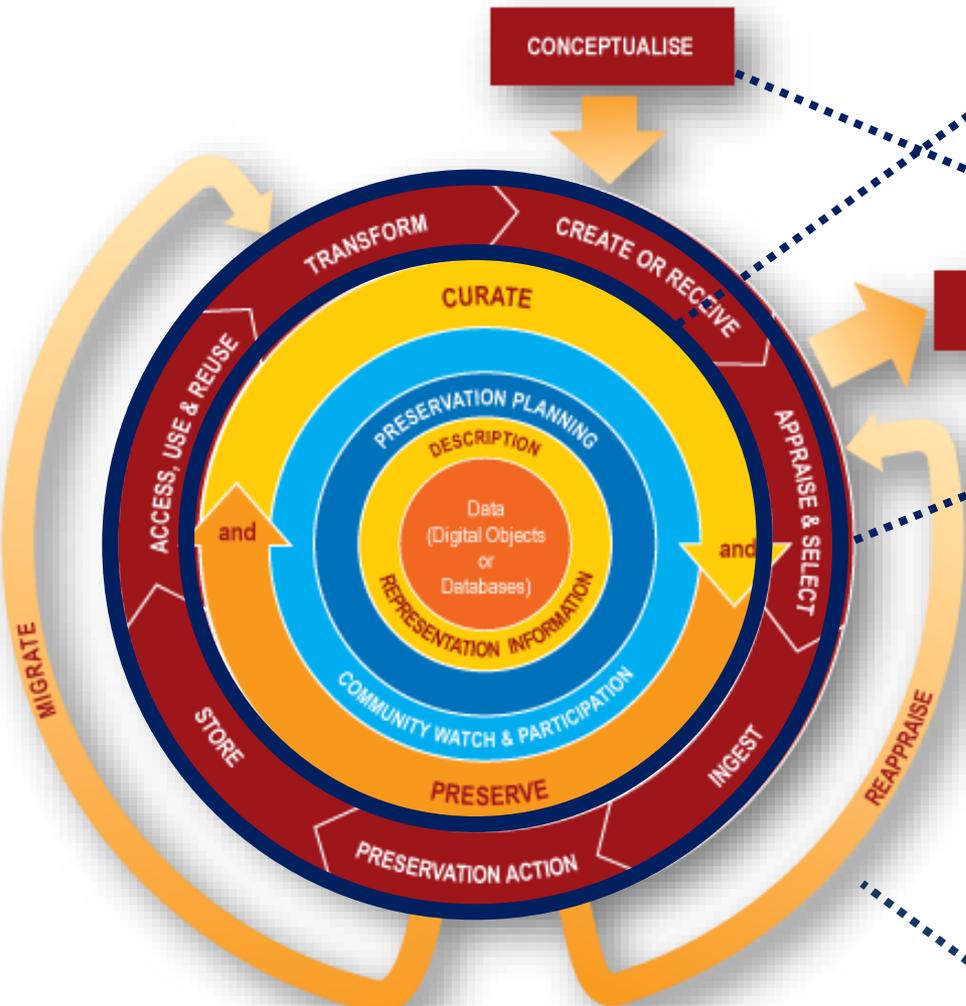
- são aplicáveis ao **longo de todo o ciclo de vida** digital
- precisam ser **realizadas sequencialmente** para a curadoria ser bem sucedida
- devem ser realizadas **ocasionalmente**, como as circunstâncias o exigirem.

DADO

Objetos digitais { simples
complexos

Bases de dados





AÇÕES PARA TODO O CICLO DE VIDA DA CURADORIA

DESCRIÇÃO e REPRESENTAÇÃO
 PLANEJAMENTO DA PRESERVAÇÃO
 PARTICIPAÇÃO e MONITORAMENTO
 CURADORIA e PRESERVAÇÃO

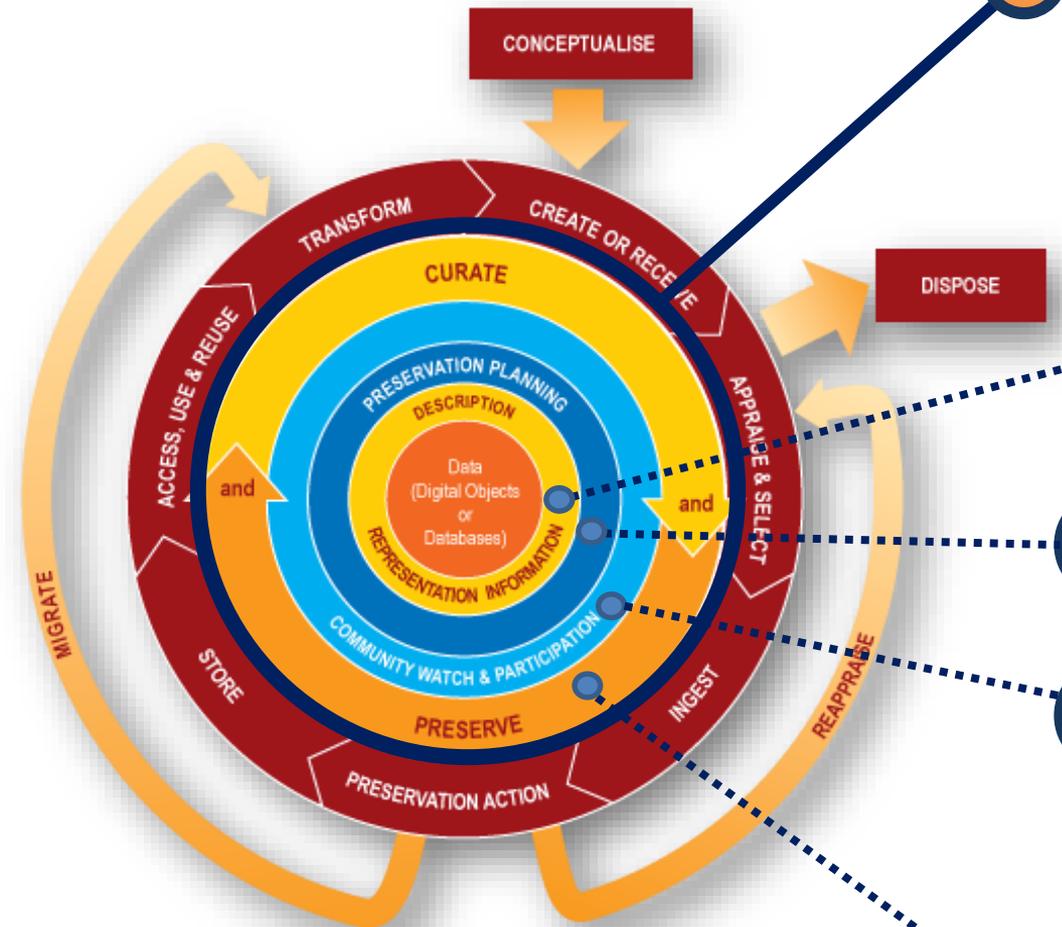
AÇÕES SEQUENCIAIS

CONCEITUALIZAÇÃO
 CRIAÇÃO OU RECEBIMENTO
 AVALIAÇÃO E SELEÇÃO
 ARQUIVAMENTO
 AÇÕES DE PRESERVAÇÃO
 ARMAZENAMENTO
 ACESSO, USO E REUSO
 TRANSFORMAÇÃO

AÇÕES OCASIONAIS

DESCARTE
 REAVALIAÇÃO
 MIGRAÇÃO

AÇÕES PARA TODO O CICLO DE VIDA DA CURADORIA



DESCRIÇÃO E (GESTÃO) DA INFORMAÇÃO DE REPRESENTAÇÃO

Criação, coleção, preservação e manutenção de **metadados** suficientes para permitir que o dado seja usado e reusado pelo tempo que seu valor justifique a curadoria continuada.

1

PLANO DE PRESERVAÇÃO

Estratégias, políticas e procedimentos para todas as atividades de curadoria

2

MONITORAMENTO E PARTICIPAÇÃO DA COMUNIDADE

Observação da comunidade-alvo em relação às suas mudanças de exigências sobre os dados; participação no desenvolvimento de padrões, ferramentas e softwares relevantes para os dados.

3

CURADORIA E PRESERVAÇÃO

Empreender continuamente ações administrativas e gerenciais em prol da preservação e da gestão durante todo o ciclo de vida da curadoria.

4

AÇÕES SEQUENCIAIS

Não são ações exclusivamente concernentes à **curadoria**, mas representam estágios do ciclo de vida do dados que devem ter um componente de curadoria



AÇÕES SEQUENCIAIS (CONT.)

Não são ações exclusivamente concernentes à **curadoria**, mas representam estágios do ciclo de vida do dados que devem ter um componente de curadoria

TRANSFORMAÇÃO

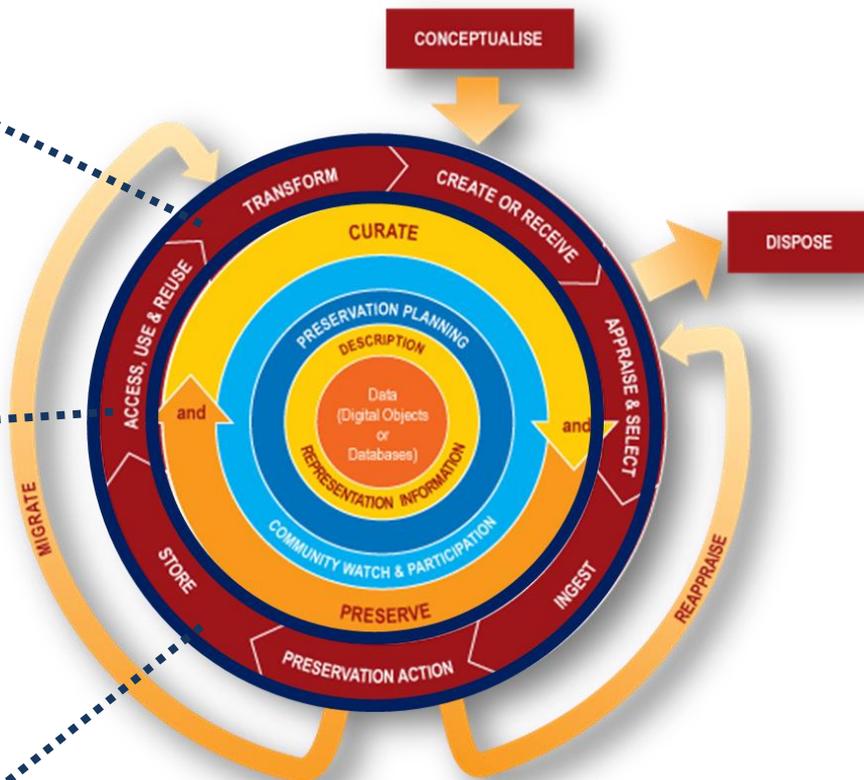
criar novos dados a partir do original como, por exemplo, a migração para um formato diferente ou a geração de resultados derivados por seleção ou consulta.

ACESSO, USO E REUSO

implica em garantir que o dado possa ser acessado tanto pela sua comunidade-alvo quanto pelos demais usuários interessados no reuso dos dados

ARMAZENAMENTO

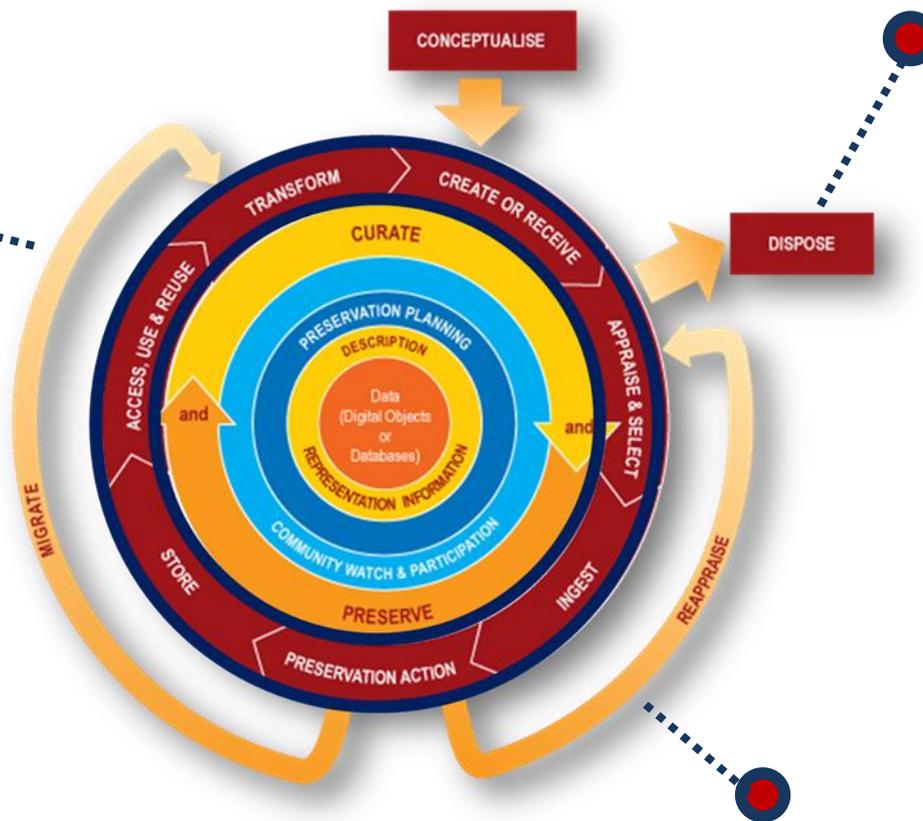
dos dados de forma segura e aderente aos padrões relevantes,



AÇÕES OCASIONAIS

Ações aplicadas eventualmente, quando necessário

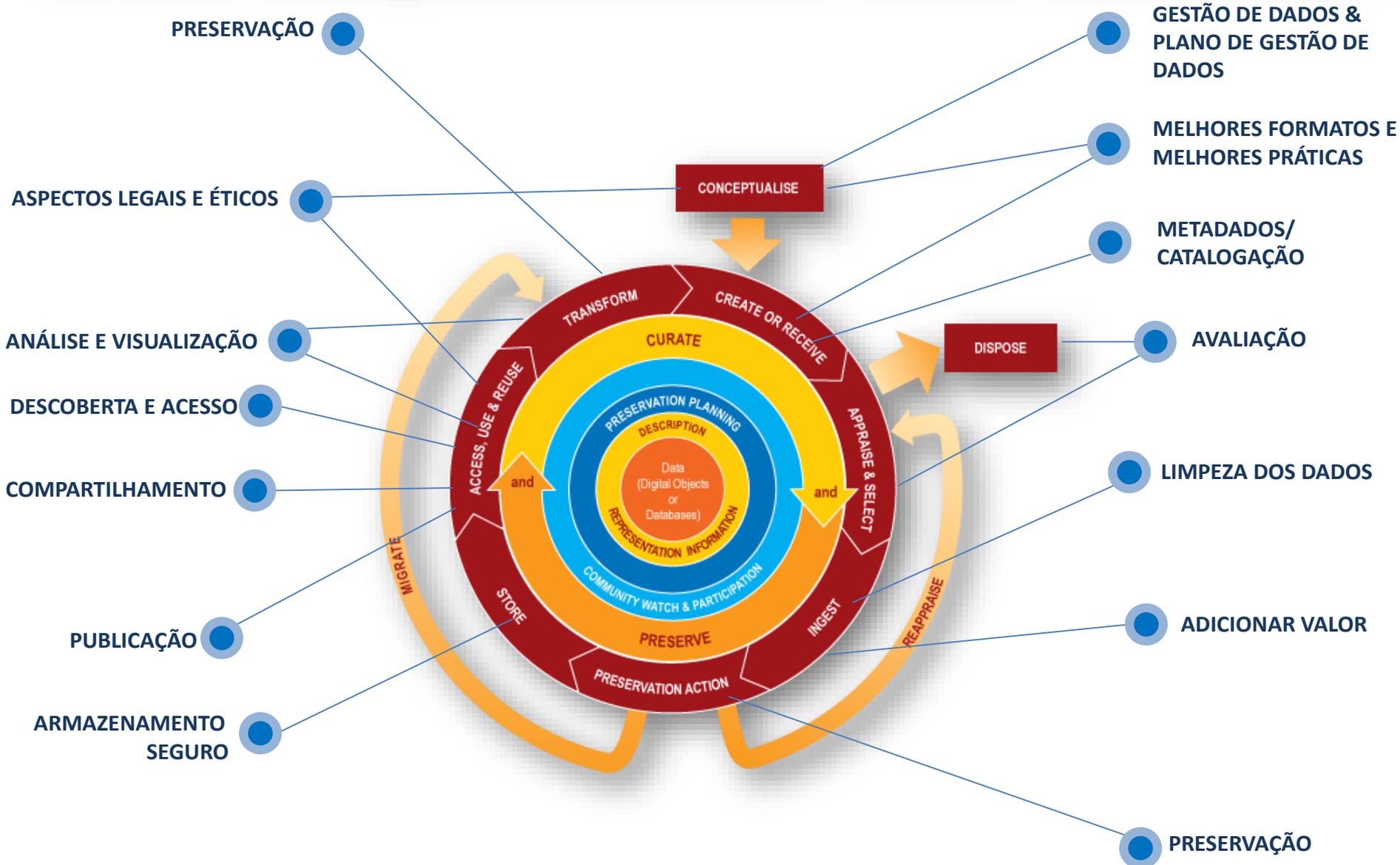
MIGRAÇÃO
do dado para um formato diferente para compatibilizá-lo com o ambiente de armazenamento, ou para garantir a imunidade à obsolescência de *hardware* e *software*.



DESCARTE
dos dados que não foram selecionados para curadoria e preservação. Os dados podem ser transferidos para outro arquivo, depósito, *data center* ou outro centro de custódia. A natureza do dado pode exigir destruição segura.

REAVLIAÇÃO
dos dados retornados por falhas nos procedimentos de validação para nova avaliação e possível seleção.

MAPEAMENTO SERVIÇOS DE GESTÃO DE DADOS X CICLO DE VIDA DA CURADORIA



MAPEAMENTO CICLO DE CURADORIA DO DCC BIBLIOTECA



PESQUISA EM PROGRESSO

geração/coleta dos dados ativa

Gestão de curto prazo

Análise de dados

Processamento dos dados

Versionamento

Armazenamento

Backups



A gestão
acontece
em
dois
momentos



CURADORIA

PESQUISA FINALIZADA

Publicação dos dados

Preservação de longo prazo

Contextualização

Ambientes confiáveis

Acesso/Reuso

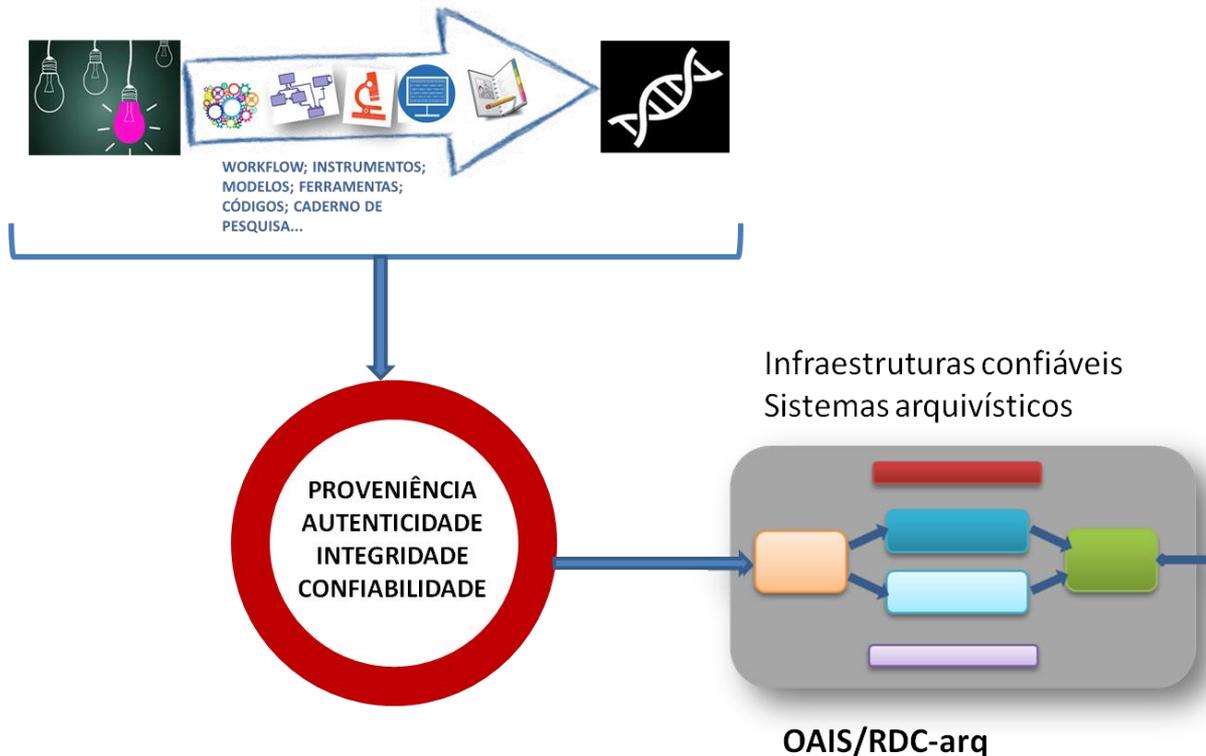
Metadados





IMPACTO NAS BIBLIOTECAS DE PESQUISA
pós-publicação → pré-publicação

COLEÇÃO DE DADOS COMO EVIDÊNCIA DA ATIVIDADE DE PESQUISA



“

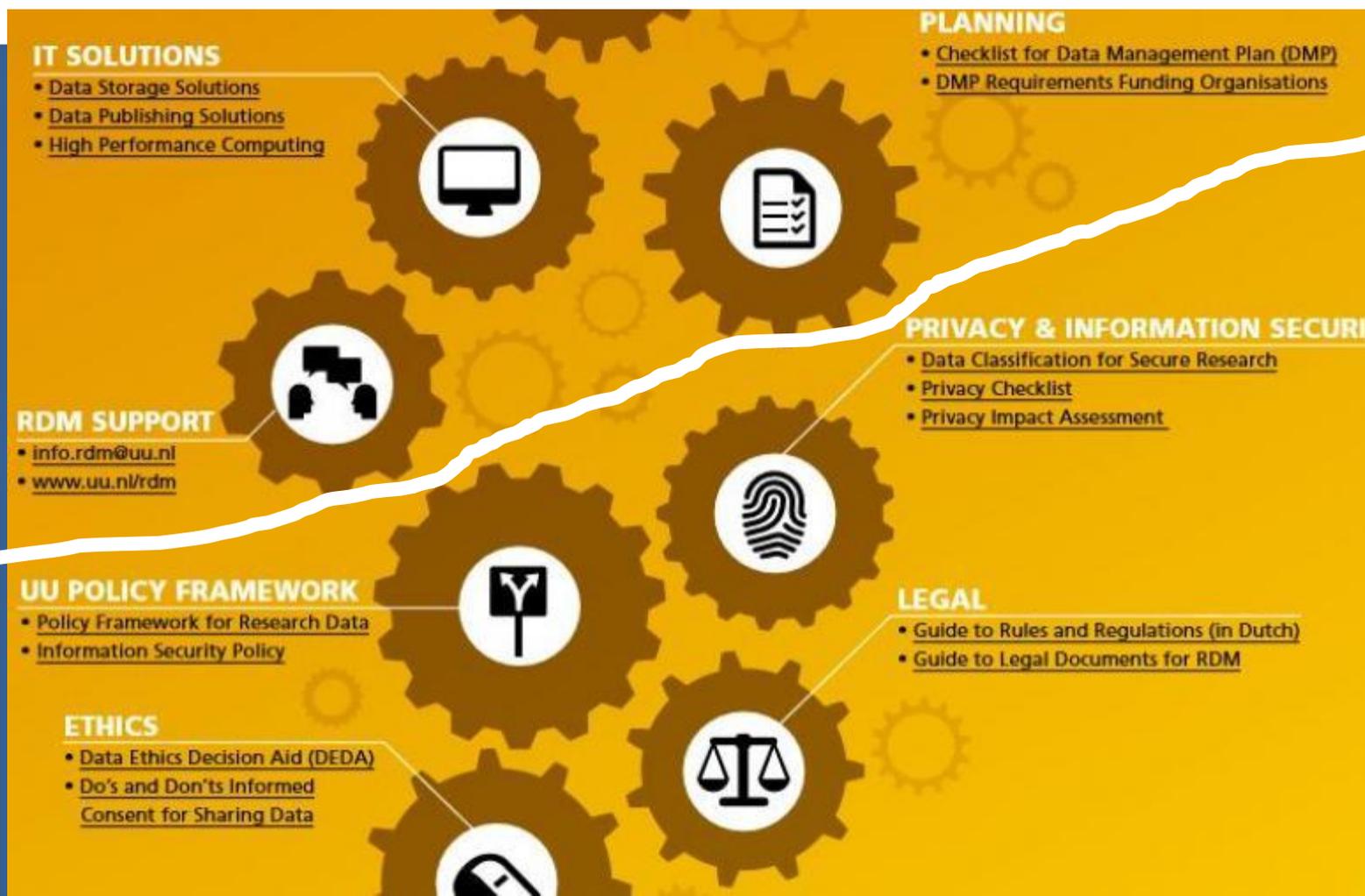
O reuso confiável de dados de pesquisa só é possível se eles forem gerenciados de forma que sua **autenticidade** e **integridade** sejam mantidas ao longo do tempo, isto porque **um pesquisador confia nos dados coletados ou gerado por outro pesquisador** para dar prosseguimento ao seu trabalho. Isto coloca como em primeiro plano a questão da **proveniência** e de como ela é endereçada pelos sistemas que cuidam da **curadoria de dados**, da **preservação digital** e do **arquivamento confiável**.

EXISTEM DUAS ÁREAS DE REQUISITOS NA GESTÃO DADOS CIENTÍFICOS:

a primeira está relacionada com as infraestruturas – sistemas, normas e protocolos - necessárias para assegurar a coleta, preservação e acesso, e ainda a disponibilidade de serviços de amplo espectro;
A SEGUNDA CONSIDERA OS ASPECTOS POLÍTICOS, LEGAIS E ÉTICOS DECORRENTES DO ACESSO E REUSO DOS DADOS ALÉM DO CONTEXTO INICIAL PARA QUE FORAM GERADOS.

INFRAESTRUTURAS:
SISTEMAS, NORMAS
E PROTOCOLOS -
NECESSÁRIAS PARA
ASSEGURAR A
COLETA,
PRESERVAÇÃO E
ACESSO, E AINDA A
DISPONIBILIDADE DE
SERVIÇOS

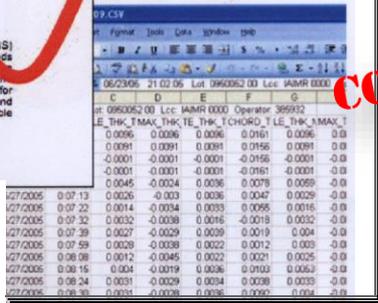
MARCOS
POLÍTICOS
ÉTICOS
LEGAIS E
DE PRIVACIDADE



REUSO

demonstrate their results and help their peers to verify these results. It also makes other researchers aware of the availability of these resources, which may lead to their reuse, saving other researchers the work of e.g. recollecting research data. They also enable creating indirect links between different publications that are possibly related. The Internet provides an infrastructure to publish text with visualizations, animations, research data, etc. Woutersen-Windouwer and Grandama (2008) indicated several initiatives for publishing enhanced publications on the web, but showed that these initiatives are not easily applicable: they don't fit into existing repository systems, there is little scientific awarding for the additional efforts required for this type of publication and archives do not know how to ingest this material. More generic solutions are needed to overcome these issues.

Data Archiving and Networked Services (DANS) is an institute of both the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organization for Scientific Research (NWO). DANS is responsible for archiving research data from the humanities and social sciences, keeping these data accessible and



	C	D	E	F	G	
06/23/06	21.02.56	Lat	0963063.00	Loc	IAMR 0000	
08/00/02.00	Lcc	IAMR 0000	Operator	369932		
LE_THK_TMAX_THK_TE_THK_TCHORO_TLE_THK_1MAX_1						
0.0006	0.0006	0.0006	0.0161	0.0006	0.0	
0.0091	0.0091	0.0091	0.0156	0.0091	0.0	
-0.0001	-0.0001	-0.0001	-0.0156	-0.0001	0.0	
-0.0001	-0.0001	-0.0001	-0.0161	-0.0001	0.0	
0.0045	-0.0024	0.0036	0.0078	0.0069	0.0	
0.0026	-0.0023	0.0036	0.0047	0.0029	0.0	
0.0014	-0.0034	0.0033	0.0065	0.0016	0.0	
0.0032	-0.0038	0.0016	-0.0019	0.0032	0.0	
0.0027	-0.0029	0.0039	0.0019	0.004	0.0	
0.0009	-0.0038	0.0022	0.0012	0.003	0.0	
0.0005	0.0012	-0.0045	0.0022	0.0021	0.0025	0.0
0.0015	0.004	-0.0019	0.0036	0.0103	0.0063	0.0
0.0024	0.0031	-0.0029	0.0034	0.0038	0.0033	0.0
0.0016	0.0031	-0.0038	0.0036	0.0003	0.004	0.0

COMPARTILHAMENTO



REPRODUTIBILIDADE

VALIDAÇÃO

AUTOCORREÇÃO

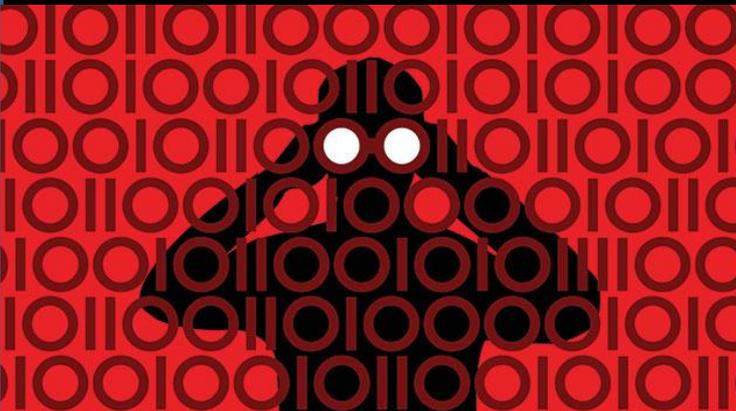
AVALIAÇÃO POR PARES

INTERPRETAÇÃO

BASE PARA NOVAS PESQUISAS

PUBLICAÇÃO

Um crescente número de novas modalidades de publicação está surgindo como resposta ao desafio de dar visibilidade e implementar estratégias de compartilhamento de dados de pesquisa. É importante observar que os mecanismos de publicação de dados tomam como solução um alinhamento ao sistema de reputação científica



As novas modalidades de publicação de dados e de suas representações descritivas demonstram com clareza que é possível de ancorar os sistemas de compartilhamento de dados às formas tradicionais de publicação, embora isso exija um alto grau de inovação e uma nova dinâmica que imponha mais velocidade nos processos de avaliação, que pode ser algo que se desenrole no tempo e se distribua no espaço de forma menos exclusiva (PAMPEL; DALLMEIR-TIESSSEN, 2015).



A publicação dos dados de pesquisa como **objeto de informação independente**, em repositórios de dados ou centros de dados.



A publicação de **documentação textual** em *data journal* sobre dados de pesquisa na forma de *data papers*



A publicação de dados de pesquisa **enriquecendo um artigo** por meio de *links* que podem ter valor semântico, nas chamadas **publicações ampliadas**



Publicação de dados de pesquisas de **experimentos que não deram certos e hipóteses não confirmadas** em periódicos voltados para essa condição

PAPÉIS NA GESTÃO DE DADOS DE PESQUISA

GESTOR – administrador de C&T que compreende a importância dos dados no âmbito institucionais, nacional e internacional, nessa direção apoia a definição de políticas, negocia recursos junto às agências de fomento, implanta e-infraestruturas e adquire ferramentas e coleções de dados.

PESQUISADOR - Autor/criador/coletor dos dados; envolvido na pesquisa que produz os dados; o autor dos dados deve assegurar que os metadados, o registro dos dados, contexto e qualidade está em conformidade com os padrões da comunidade (NSC, 2005). Elabora junto com o bibliotecário/arquivista o PGD

BIBLIOTECÁRIO DE DADOS - Profissional da área de **biblioteconomia** com formação em gestão de repositórios de dados e de curadoria, indexação e catalogação de dados e conhecedor dos fluxos das pesquisas locais. Promove cursos e apoia a elaboração do PGD

ARQUIVISTA DE DADOS – profissional de **arquivologia** responsável pelo arquivamento e preservação de longo prazo dos dados e garantia de autenticidade, integridade e confiabilidade

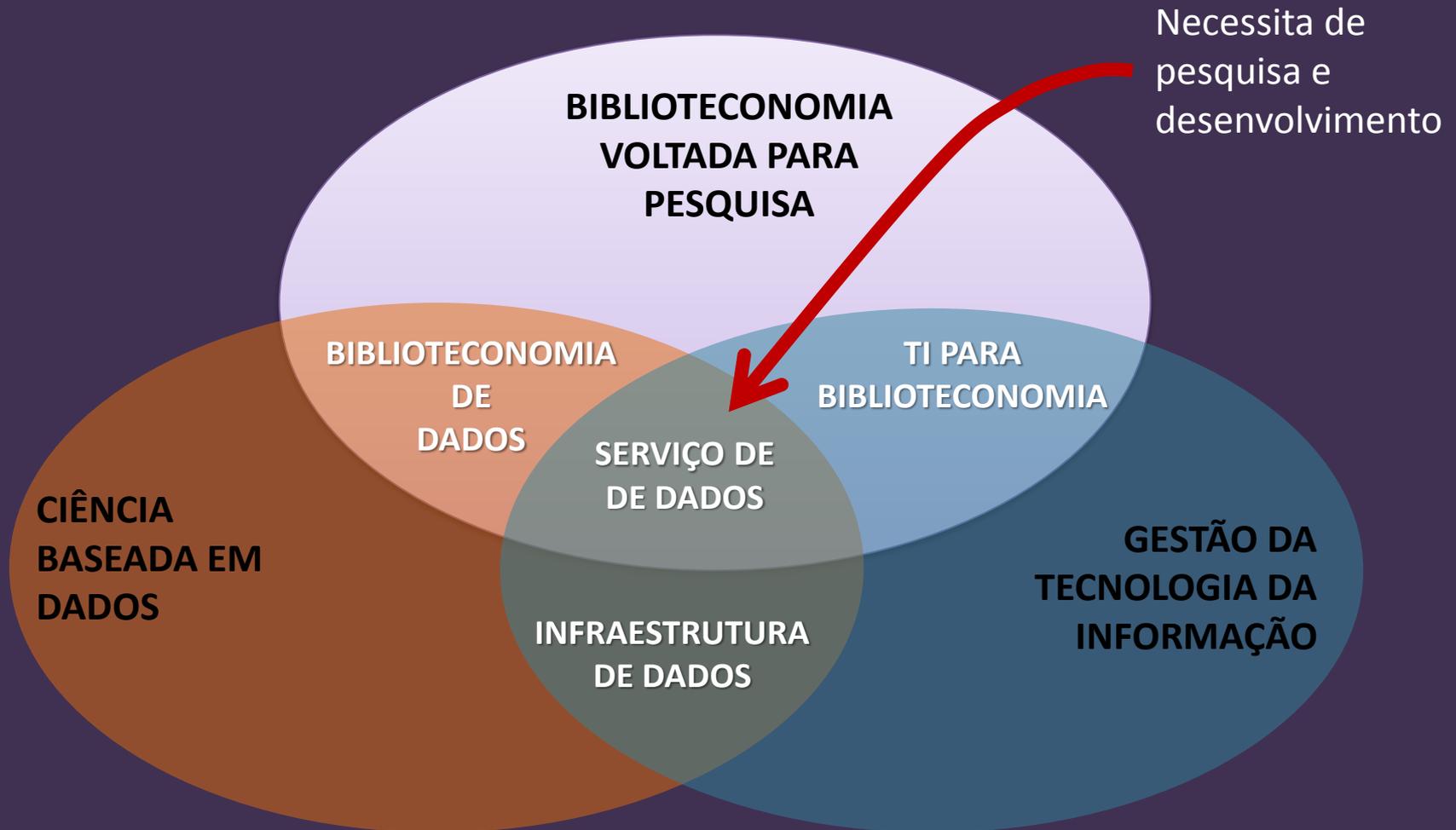
CIENTISTA DE DADOS – profissional das **áreas de computação** e/ou da área disciplinar que contribui no desenvolvimento de tecnologias de análise, manipulação, visualização, modelagem, algoritmos para as coleções de dados. Trabalha próximo aos pesquisadores

GERENTE DE DADOS – **tecnologista da informação** responsável pela manutenção e operação das bases de dados, segurança e armazenamento dos dados: backups, checagem de integridade, etc.

CURADOR DE DADOS – **pesquisador ou cientista de informação** com conhecimento disciplinar que adiciona valor aos dados por meio de documentação, integração, anotações, *mashup*, etc. Promove o compartilhamento e reuso, avalia para a preservação e cria serviços,

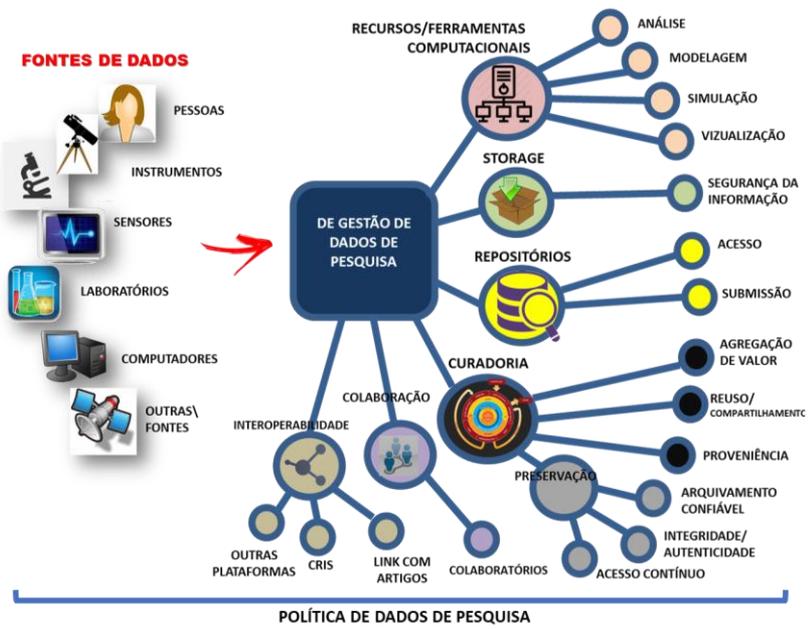
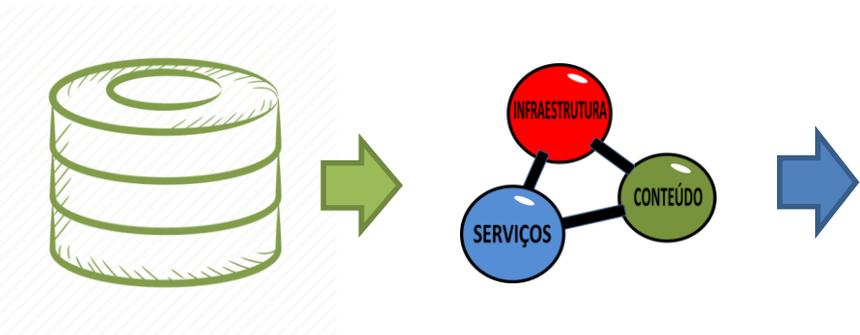


INFRAESTRUTURA DE DADOS E A BIBLIOTECA DE PESQUISA



PLATAFORMA DE GESTÃO DE DADOS

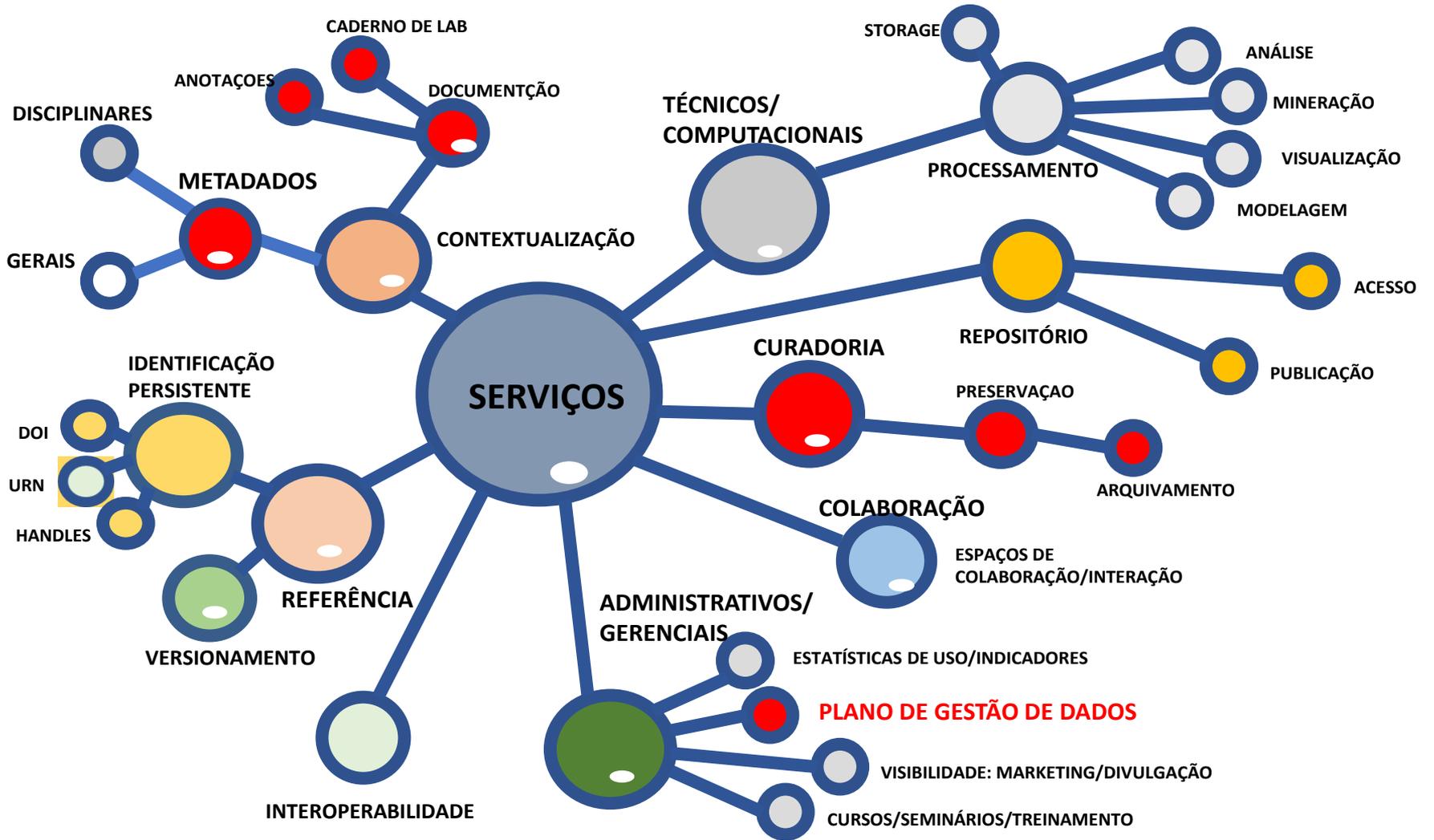
REPOSITÓRIO



COLETA & ARMAZENAMENTO

GESTÃO

CURADORIA





REPOSITÓRIO DE DADOS



CEMITÉRIO DE DADOS ?

DADOS DE PESQUISA



Quem ama cuida

Luana Sales - IBICT
Luís Sayão - CNEN

Ilustrações: Joana Marques

OBRIGADO!

