

GESTÃO DE DADOS DE PESQUISA

Luis Fernando Sayão
CNEN/CIN

Luana Farias Sales
CNEN/IEN
IBICT



AGENDA

eScience e Ciência Aberta

Dados de Pesquisa

Gestão e Curadoria

Repositório de Dados

Política Nacional

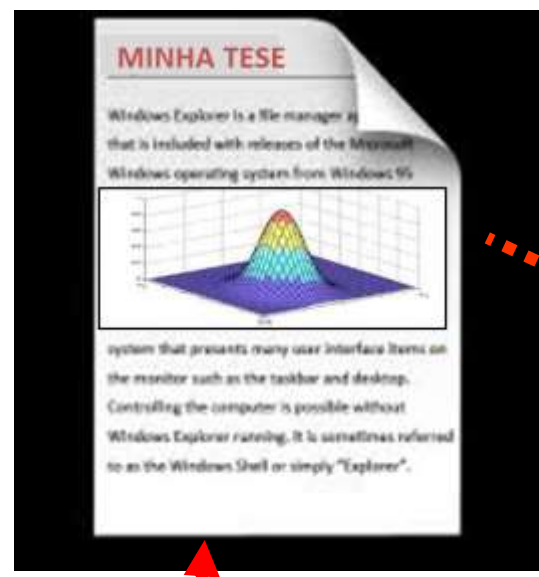


DADOS DE PESQUISA SÃO MUITO SUSCETÍVEIS A PERDAS

PROJETO DE PESQUISA

PUBLICAÇÃO

DADOS de PESQUISA



- BIBLIOTECA CONVENCIONAL
- BIBLIOTECA DIGITAL
- REPOSITÓRIO DIGITAL



A collage of research data. It includes a graph titled "ERBS SOLAR MONITOR" showing solar irradiance from 1982 to 2002. Next to it is a table with numerical data. Below these are a bar chart and a 3D molecular model of a protein structure.

Year	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Annual Total	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Jan	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Feb	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Mar	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Apr	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
May	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Jun	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Jul	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Aug	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Sep	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Oct	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Nov	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032
Dec	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032	1032

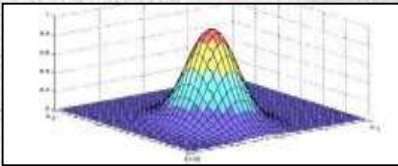


VISÍVEL INVISÍVEL

O TEXTO ACADÊMICO APRESENTA APENAS OS DADOS DE PESQUISA DE FORMA CONDENSADA

MINHA TESE

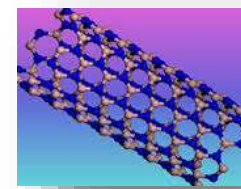
Windows Explorer is a file manager that is included with releases of the Microsoft Windows operating system from Windows 95



system that presents many user interface items on the monitor such as the taskbar and desktop. Controlling the computer is possible without Windows Explorer running. It is sometimes referred to as the Windows Shell or simply "Explorer".

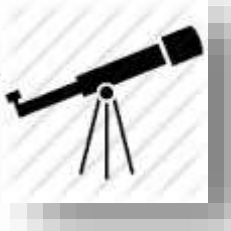
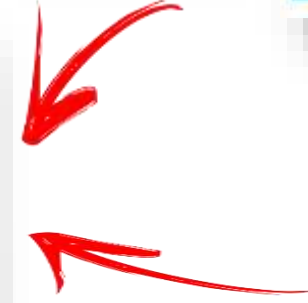
UMA VISÃO DOS DADOS 

**[revisão por pares]
[validação da pesquisa]**



ATV-26 Test Results
(Automated Assessment)

QUESTION	ANSWER	SCORE	STATUS	FEEDBACK
1. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
2. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
3. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
4. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
5. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
6. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
7. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
8. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
9. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!
10. Which of the following is a file manager?	Windows Explorer	100%	Correct	Good job!



OS DADOS SUBJACENTES ÀS PUBLICAÇÕES APOIAM A:

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. Wouters-Winshouwer 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 the 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

	0	1	2	3	4	5
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
3	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
4	0.0005	-0.0004	0.0006	0.0009	0.0009	-0.0009
5	0.0007	0.0026	-0.0003	0.0006	0.0041	0.0029
6	0.0014	0.0014	0.0004	0.0001	0.0008	0.0018
7	0.0032	-0.0006	0.0018	-0.0018	0.0002	-0.0002
8	0.0027	-0.0029	0.0009	0.0019	0.0004	-0.0003
9	0.0028	0.0009	-0.0009	0.0022	0.0012	0.0001
10	0.0012	0.0045	0.0027	0.0001	0.0025	-0.0003
11	0.0016	0.0001	-0.0019	0.0006	0.0003	0.0004
12	0.0001	-0.0001	-0.0001	0.0004	0.0009	0.0001
13	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
14	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
15	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
16	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
17	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
18	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001



REUSO

DOS DADOS DE PESQUISA

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-Windhauer 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 the 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 initiative 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

	0	1	2	3	4
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0001	0.0001	0.0001	0.0001	0.0001
2	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
3	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
4	0.0005	-0.0004	0.0006	0.0009	0.0009
5	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0001	0.0001	0.0001	0.0001	0.0001
7	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
8	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0001	0.0001	0.0001	0.0001	0.0001
10	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
11	0.0000	0.0000	0.0000	0.0000	0.0000
12	0.0001	0.0001	0.0001	0.0001	0.0001
13	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
14	0.0000	0.0000	0.0000	0.0000	0.0000
15	0.0001	0.0001	0.0001	0.0001	0.0001
16	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
17	0.0000	0.0000	0.0000	0.0000	0.0000

GESTÃO & CURADORIA

**ANALISADOS
EM NOVOS E
DIFERENTES
CONTEXTOS**

Os pesquisadores começaram a creditar **toda a confiança** nos conteúdos digitais **criados por outros pesquisadores** para dar prosseguimento aos seus empreendimentos



CENÁRIOS

HIPERINFORMAÇÃO

DILÚVIO DE DADOS NA CIÊNCIA

DILÚVIO 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**



eScience

OS PARADIGMAS CIENTÍFICOS

1º PARADIGMA:

Ciência experimental ou empírica estuda a relação entre fenômenos por meio de experimentos

2º PARADIGMA:

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

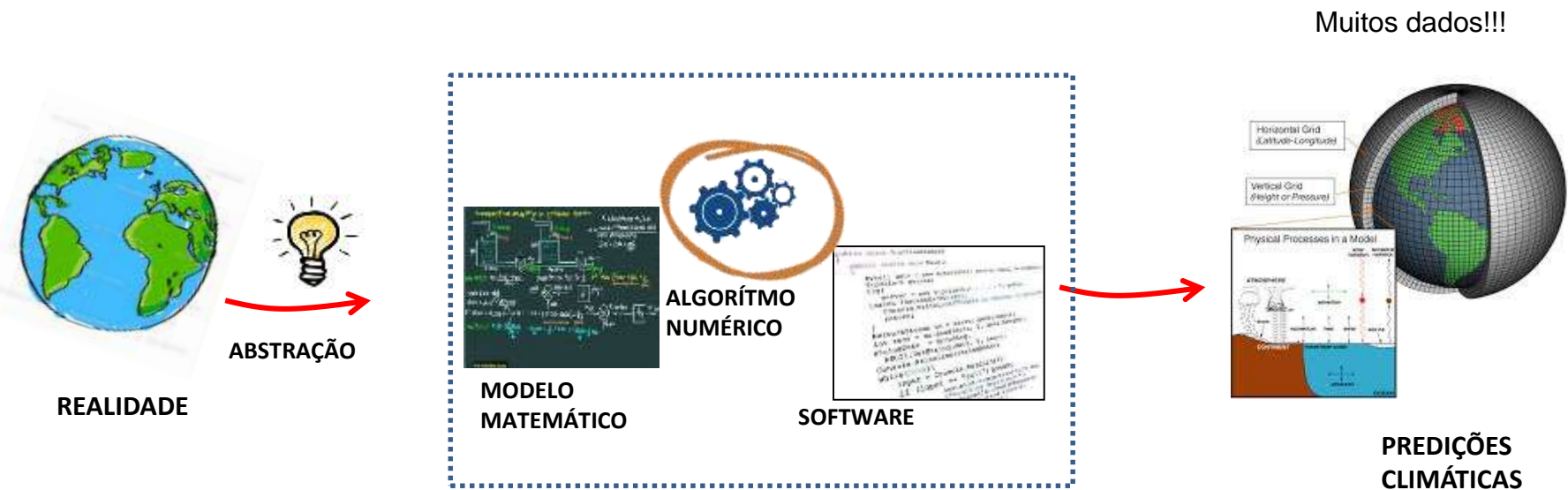
3º PARADIGMA:

Ciência baseada em **simulação** uso de *software* e grande **geração de dados**



SIMULAÇÃO POR COMPUTADOR

É análogo a um experimento físico, mas usa equações matemáticas para representar o mundo real



No eScience a simulação deixa de ser uma ferramenta que auxilia o pesquisador a fazer ciência para transformar o modo de fazer ciência e definir um novo **PARADIMA CIENTÍFICO**.

**EXEMPLO
DE
RESULTADO
DE
PESQUISA
NA ÁREA
NUCLEAR
(CNEN/IEN)**

**VIRTUAL SIMULATION OF A NUCLEAR
POWER PLANT'S CONTROL ROOM AS A
TOOL FOR ERGONOMIC EVALUATION.**



Leandro Barbosa S. Gatto ^a, Antônio Carlos A. Mól ^{a,b,c},
Isaac J.A. Luquetti dos Santos ^a,
Carlos Alexandre F. Jorge ^{a,*}, Ana Paula Legey ^c

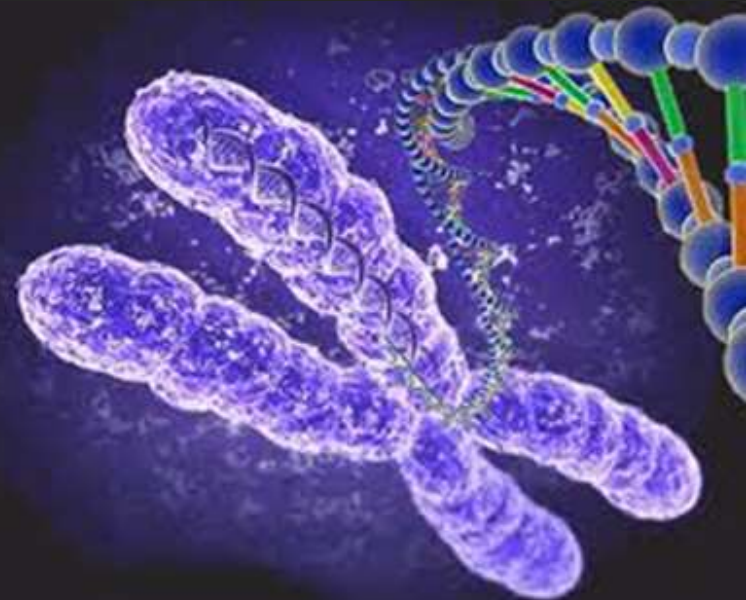
INSTITUTO DE ENGENHARIA NUCLEAR - CNEN/IEN

O DILÚVIO DE DADOS

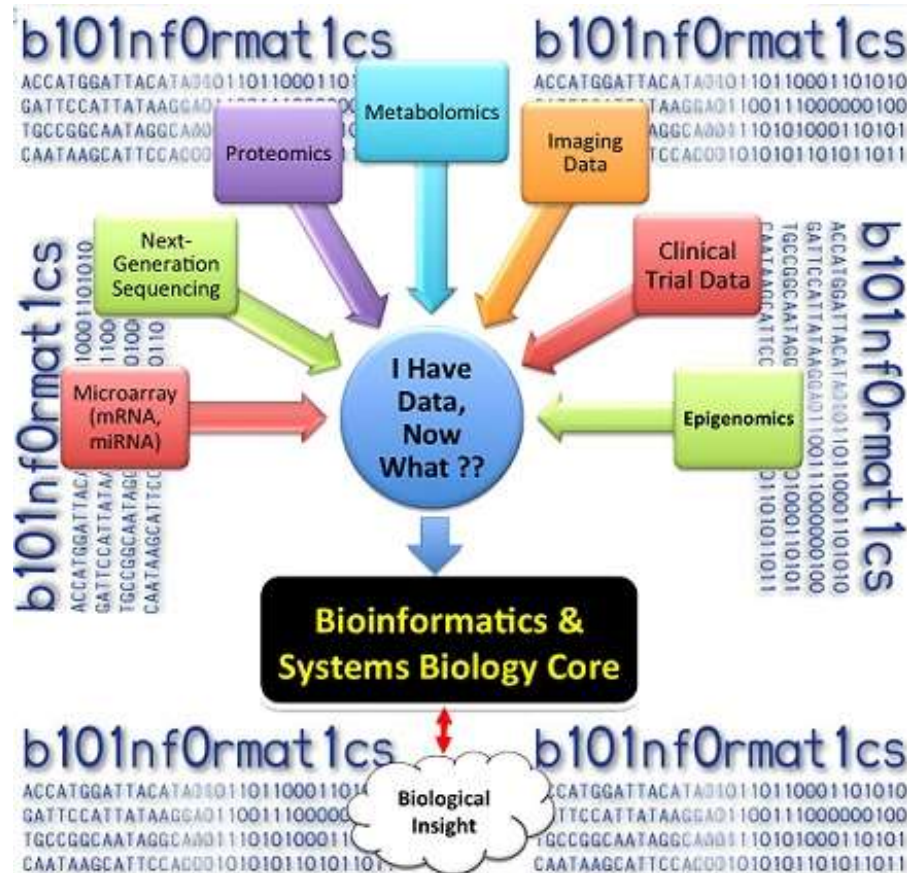
Uma nova geração de instrumentos científicos, sensores, satélites, software de simulação, laboratórios produzem em ritmo exponencial quantidades imensas e diversificadas de dados brutos de pesquisa

Subprodutos dos processos de pesquisa

Existem hoje disciplinas científicas totalmente orientadas por dados, por exemplo:
BIOINFORMÁTICA e ASTROINFORMÁTICA



BIOINFORMÁTICA



BIOINFORMÁTICA é um campo interdisciplinar que desenvolve métodos e **ferramentas de software** para compreender dados biológicos. Como um campo interdisciplinar da ciência, a bioinformática combina **ciência da computação, estatística, matemática, e engenharia** para processar, analisar e **interpretar dados biológicos**.

ASTROINFORMÁTICA



Astroinformática está focada em desenvolver ferramentas, métodos e **aplicações computacionais**, da **ciência de dados** e da **estatística** para pesquisa e educação na área de astronomia orientada por dados.

OBSERVAÇÕES



MODELAGEM
& SIMULAÇÃO



NOVAS
DESCOBERTAS
CIENTÍFICAS



COLETA DE
DADOS



HIPÓTESES

DADOS



DESCOBERTA
DE
CONHECIMENTO



BASES DE DADOS

- ANÁLISE PREDITIVA
- EXTRAÇÃO DE CLUSTER
- DETEÇÃO DE ANOMALIAS
- ANALISE DE CORRELAÇÕES
- ETC.

DATA MINING



PADRÕES E
MODELOS

O QUARTO PARADIGMA CIENTÍFICO

eScience

ACELERAR A PESQUISA CIENTÍFICA E GERAR CONHECIMENTO COM BASE NA EXPLORAÇÃO DESSE ACÚMULO DE DADOS



Ferramentas avançadas de **software e de mineração** de dados ajudam a interpretar e transformar os dados brutos em **configurações ilimitadas de informação e conhecimento**.

Perguntas instigantes e recursivas colocadas perante os vários segmentos científicos podem agora ser endereçadas, pela combinação de **múltiplas fontes de dados provenientes de domínios diferentes**, através da aplicação de modelos complexos e de métodos inéditos de análise.

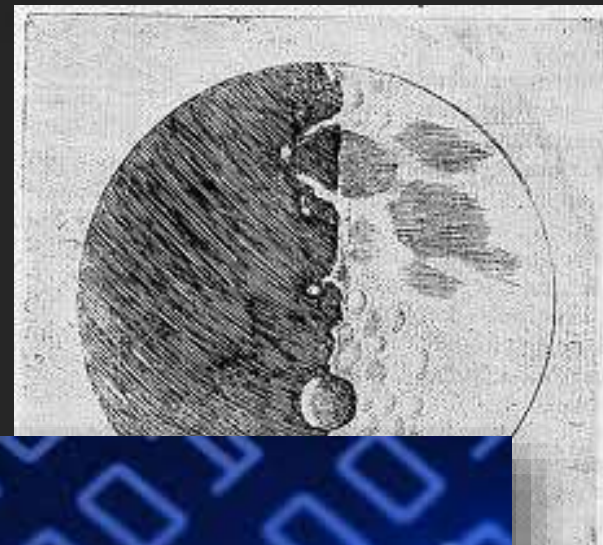
O QUARTO PARADIGMA CIENTÍFICO

eScience

“

CIÊNCIA PRODUZIDA A PARTIR DO USO, ARMAZENAMENTO, PROCESSAMENTO, ANÁLISE E COMPARTILHAMENTO DE DADOS DE PESQUISA (GRAY, 2005).

O poder dos computadores modernos permite que **relações altamente complexas** e até então despercebidas possam ser identificadas e se tornem o motor do quarto paradigma



O QUARTO PARADIGMA CIENTÍFICO

eScience

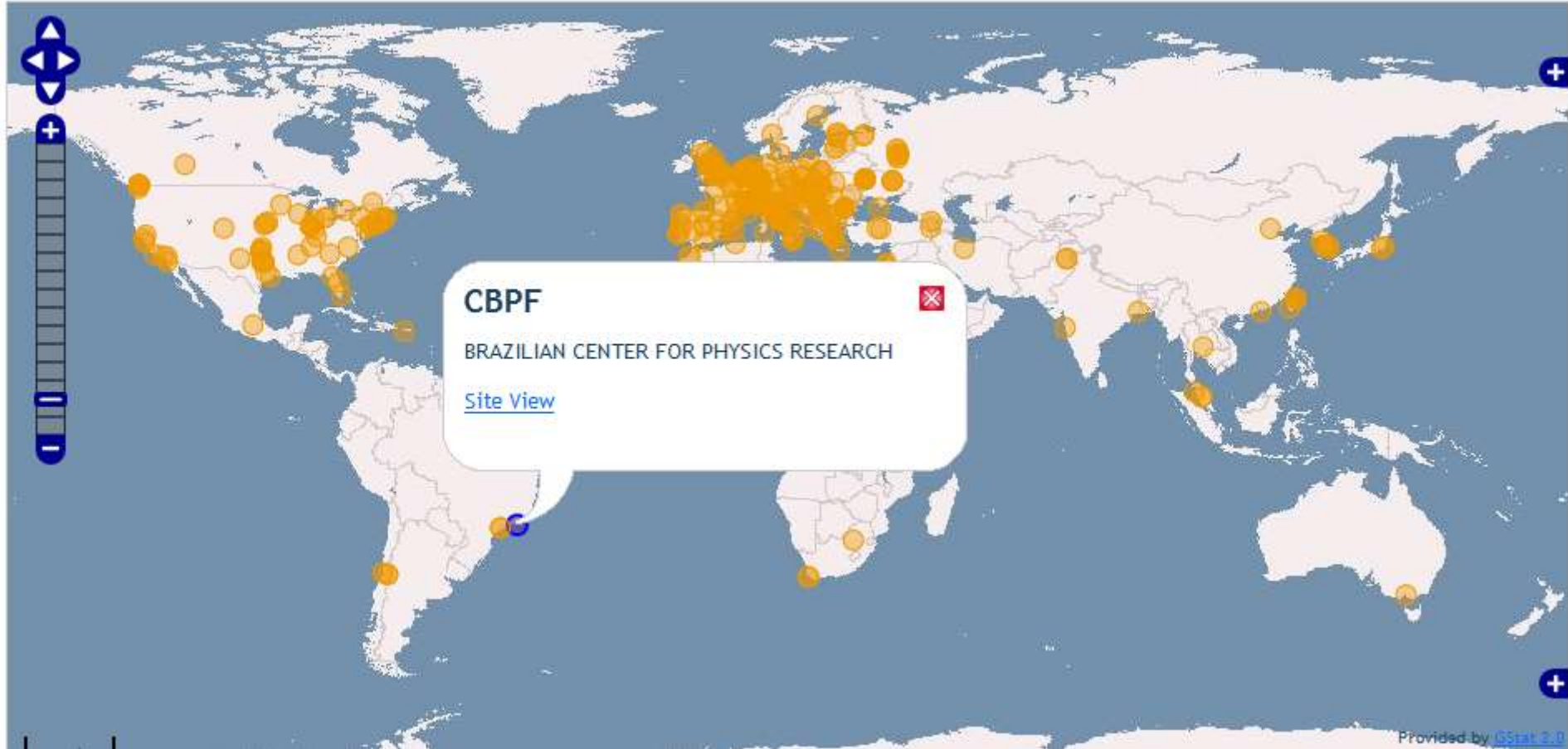
O MODO DE FAZER CIÊNCIA MUDA....

A computação não é mais meramente um suporte para o padrão tradicional de se conduzir a investigação científica em determinadas disciplinas, mas pode mudar fundamentalmente o desenvolvimento dessas disciplinas.

FORMULAÇÃO DE HIPÓTESES

Ao invés de hipóteses serem testadas e desenvolvidas a partir de dados coletados para este propósito, **hipóteses são construídas após a identificação relações nos conjuntos de dados**. Neste abordagem os dados vem primeiro, incorporados numa sequencia de captura de dados, curadoria e análises

WORLDWILDE LHC COMPUTING GRID
GLOBAL COLABORATION
42 COUNTRIES – 170 COMPUTING CENTRES





WLCG

Worldwide LHC Computing Grid

Running jobs: 236092
Transfer rate: 11.41 GiB/sec

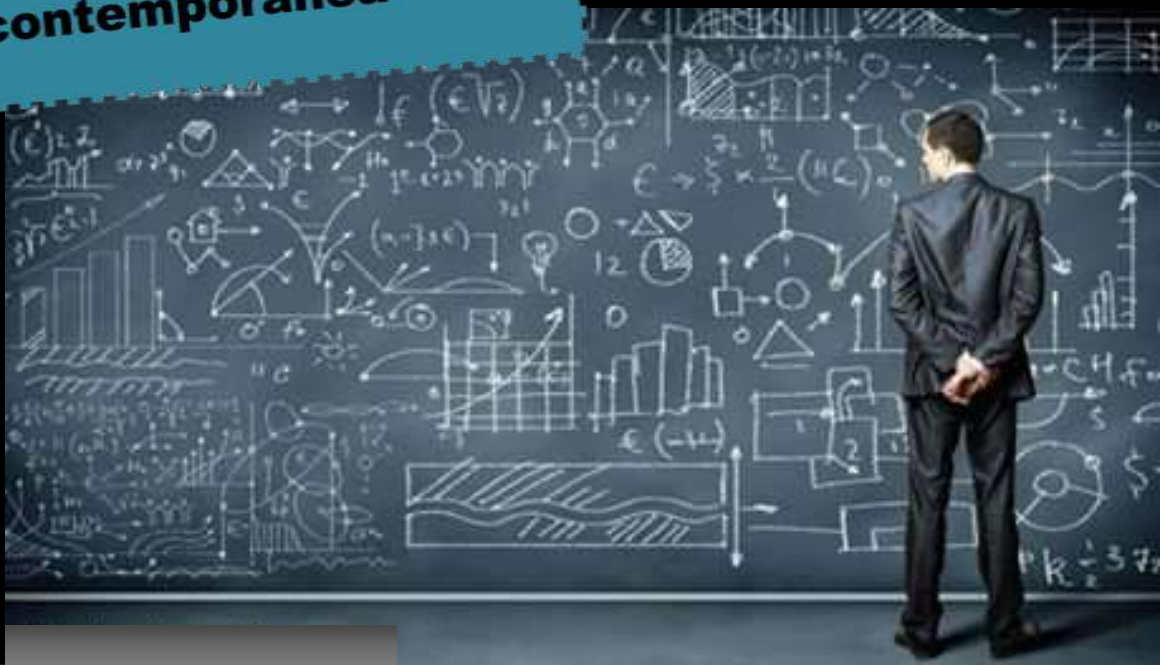


US Dept of State Geographer
© 2013-Google
© 2009 GeoBasis-DE/BKG
Data SIO, NOAA, U.S. Navy, NGA, GEBCO

Google

O QUARTO PARADIGMA CIENTÍFICO eScience

Como traduzir em significado e conhecimento a torrente de dados digitais que caracteriza certos domínios da ciência contemporânea

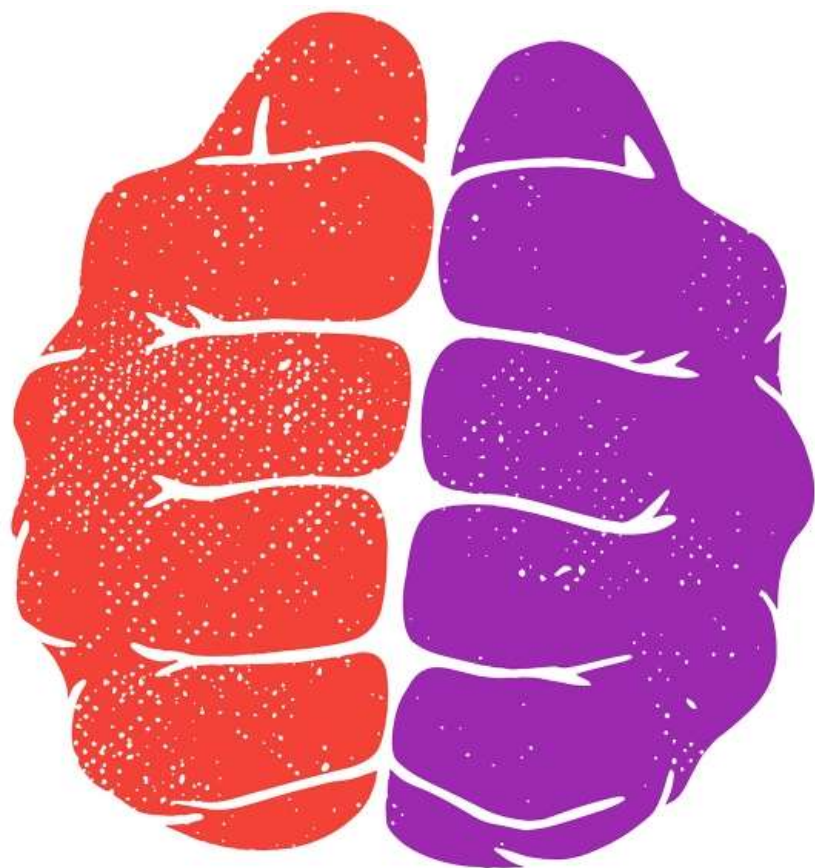


O QUARTO PARADIGMA CIENTÍFICO eScience

Como traduzir em significado e conhecimento a torrente de dados digitais que caracteriza certos domínios da ciência contemporânea



É exatamente a abundância dos dados de pesquisa que desafia os profissionais de informação, de computação e os cientistas de dados a definirem estratégias que possam controlar o fluxo de dados provenientes das atividades de pesquisa, e preservar as evidências únicas para uso futuro

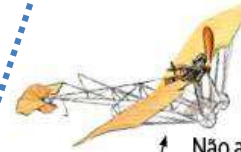
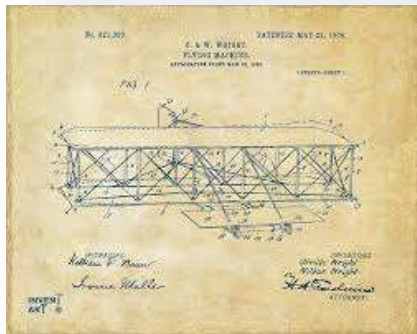


CIÊNCIA
ABERTA

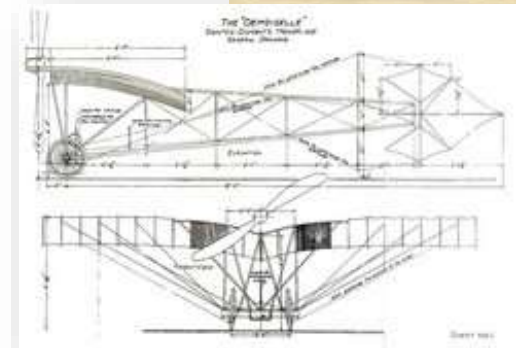
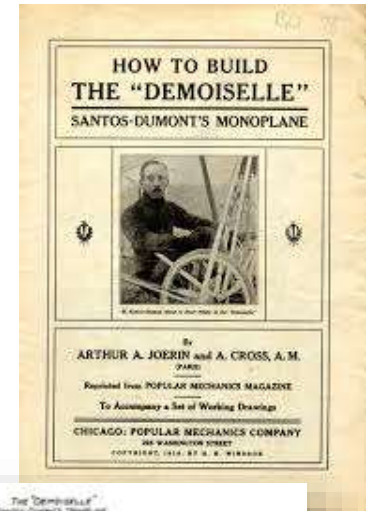
No século XX houve um movimento na direção do fechamento do conhecimento patrocinado por grandes corporações no sentido de privar parte das pessoas do acesso ao conhecimento como forma de gerar receita financeira (KON, 2013)



Com a patente, os irmãos receberam propostas lucrativas e passaram a comercializar as máquinas. Fundaram a **Wright Company** em 1909 e, no ano seguinte, fizeram o primeiro voo comercial da história. Ficaram ricos vendendo aviões.



Não acreditava em patentes e divulgou seus estudos para empresas copiarem de graça, vendo suas invenções como um presente à humanidade. Seu modelo **Demoiselle N° 20** foi o primeiro produzido em série e traz conceitos seguidos até hoje pela indústria aeronáutica

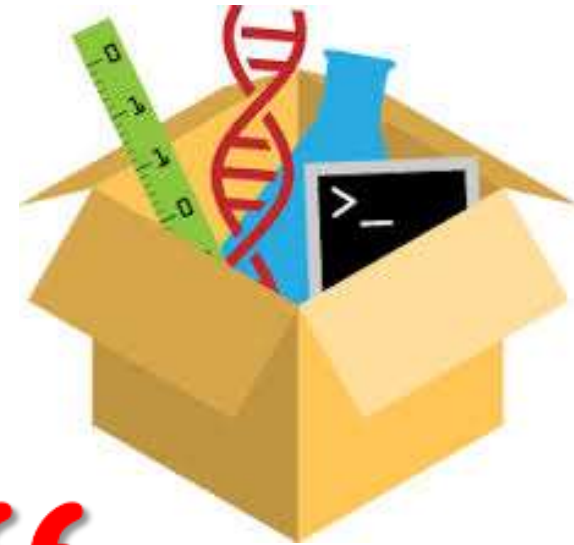


Muitos dos crescimentos marcantes da ciência nos últimos séculos é devido a práticas abertas...

Ciência aberta

O conhecimento científico é um patrimônio da humanidade que, portanto, deve estar disponível livremente para que as pessoas, cientistas ou não, possam usá-lo, reusá-lo, distribuí-lo sem constrangimentos tecnológicos, econômicos, sociais ou legais

O compartilhamento e o intercâmbio permitem descobrir conexões no que estava antes desconectado



“

Quando há compartilhamento de ideias e abertura do conhecimento a ciência avança mais rapidamente

REPRODUTIBILIDADE

Reprodutibilidade dos experimentos científicos é um dos fundamentos da ciência.

TRANSPARÊNCIA NAS METODOLOGIAS

Códigos fontes para reproduzir dos dados; Uso de software livres e formatos abertos; Ferramentas de pesquisa abertas; Dados de entrada e metadados Cadernos de pesquisa abertos

DISPONIBILIDADE DOS DADOS

Os dados científicos devem estar disponíveis para qualquer pessoa sem restrições de copyright, patentes ou outros mecanismos de controle. Dados abertos incentivam o reuso em outras áreas diferentes da original, o que pode levar a descobertas surpreendentes.

ACESSO AOS RESULTADOS

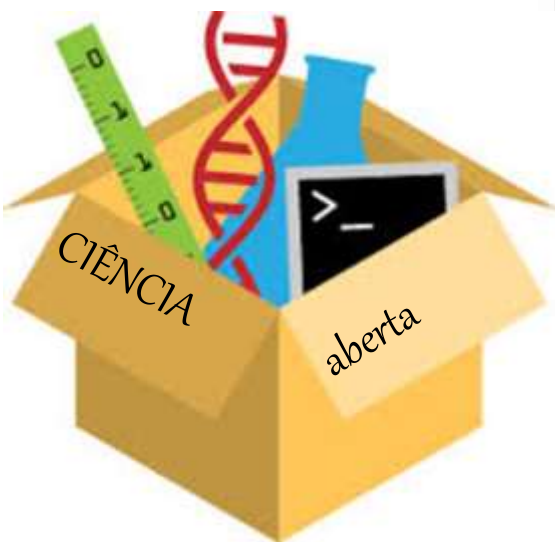
Os pesquisadores devem divulgar suas descobertas de forma que elas estejam acessíveis para todos os usuários potenciais sem qualquer barreira.

PESQUISA 2.0

Colaboração crescente entre cientistas efetivada por meio das mídias sociais e da internet. Um número crescente de cientistas estão encontrando novas estratégias para comunicar seus trabalhos usando wikis, blogs, twitter

AVALIAÇÃO

A avaliação pelas instituições de pesquisa, bem como a aprovação de financiamento pelas agências deve levar em conta a preparação dos dados para disponibilidade na mesma escala em que considera artigos de periódicos e outras publicações, ou seja o nível de transparência.



CAUDA LONGA DA CIÊNCIA



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

EXPERTISES

PESQUISADORES
CIENTISTAS DE DADOS
BIBLIOTECARIOS DE DADOS
ARQUIVISTAS

ORGANIZAÇÕES

UNIVERSIDADES
INSTITUTOS DE PESQUISA
AGÊNCIAS DE FOMENTO
BIBLIOTECAS, ARQUIVOS, MUSEUS
ORGANIZAÇÕES VIRTUAIS;
COMUNIDADES

INSTRUMENTOS CIENTÍFICOS

TELECÓPIOS
SATÉLITES
COLISORES
SENSORES

CIBERINFRAESTRUTURA DE PESQUISA

DADOS

BASES DE DADOS
REPOSITÓRIOS
ACESSO
GESTÃO
CURADORIA
MINERAÇÃO
PRIVACIDADE

RECURSOS COMPUTACIONAIS

SUPERCOMPUTADORES
NUVEM, GRID, CLUSTER;
VISUALIZAÇÃO;
CENTROS DE COMPUTAÇÃO

REDES

REDES DE
PESQUISA/EDUCAÇÃO
NACIONAIS E
INTERNACIONAIS;
SEGURANÇA

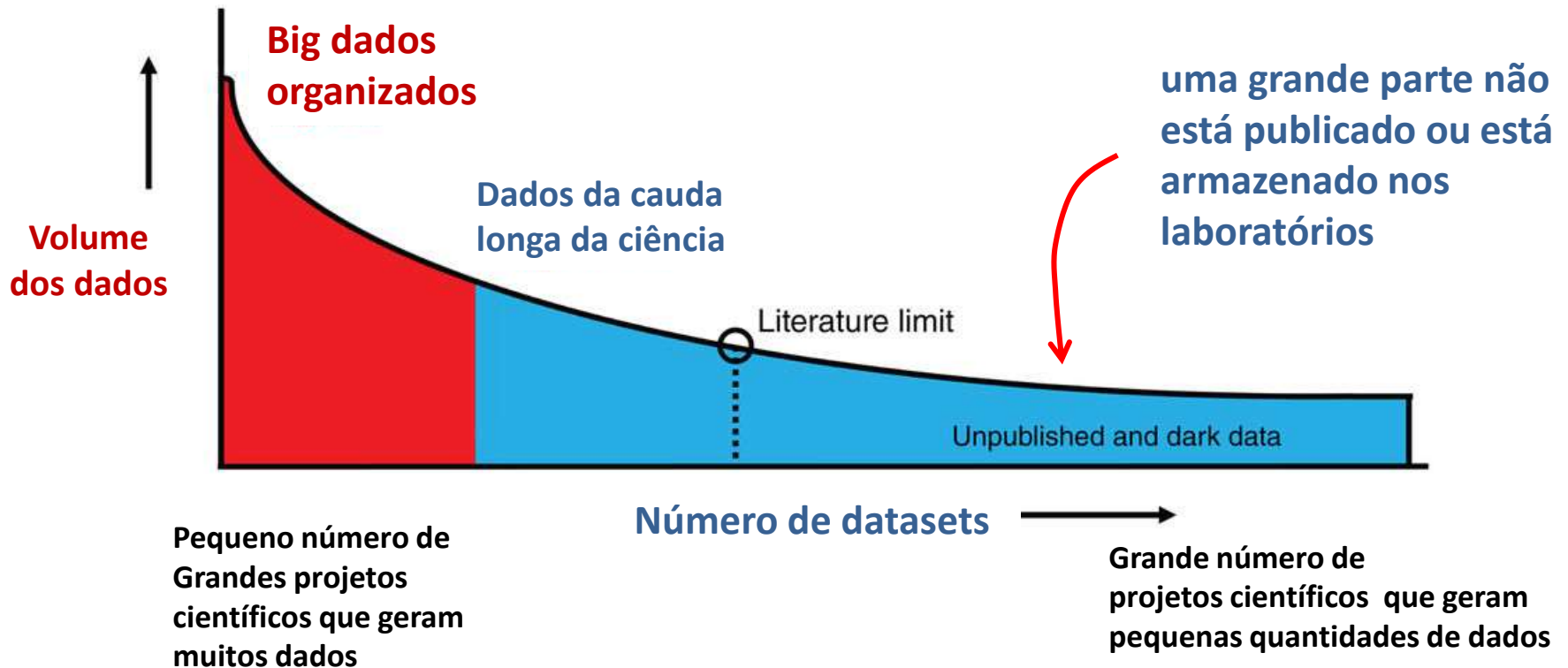
SOFTWARE

APLICAÇÕES;
DESENVOLVIMENTO
E SUPORTE



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



+50%

DOS ACHADOS NÃO FORAM PUBLICADOS



O COMPRTILHAMENTO PODE REVELAR VALORES IMPORTANTES OCULTOS NESSES DADOS

PIRÂMIDE DE GESTÃO DE DADOS

HIERAQUIA DE VALOR & PERMANÊNCIA



**BIG
SCIENCE****X****SMALL
SCIENCE**

CARACTERÍSTICAS		CABEÇA	CAUDA LONGA
UNIFORMIDADE	DIVERSIDADE	homogêneos	heterogêneos
	GERAÇÃO/ COLETA	instrumentos automatizados	gerados/coletados manualmente
	PROCEDIMENTOS	padronizados	específicos
GESTÃO	CURADORIA	Centralizada/ institucionalizada	Individual
	REPOSITÓRIOS DIGITAIS	Disciplinares ou referenciais	Institucionais ou Multidisciplinares
	PRESERVAÇÃO	Preservados	Não preservados
	ARMAZENAMENTO	Sistemas de Storage	Computadores pessoais/ dispositivos portáteis
	ESTRUTURAÇÃO	Banco de dados	Planilhas
COMPARTILHAMENTO	ACESSO	Acesso aberto/ distribuído	Obscuro ou protegido
	REUSO	Imediato/globalizado	Episódico/entre a equipe
INSTITUCIONALIZAÇÃO	FINANCIAMENTO	Fluxo contínuo/ Apoio internacional	Por projeto
	RECONHECIMENTO/ RECOMPENSA	SIM	NÃO



DESCONTINUIDADE NA MEMÓRIA CIENTÍFICA DAS INSTITUIÇÕES

DUPLICAÇÃO DE ESFORÇOS E RECURSOS

PRINCÍPIO DA REPRODUTIBILIDADE DOS EXPERIMENTOS

VALIDAÇÃO E AUTOCORREÇÃO DA PESQUISA

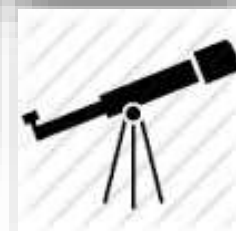
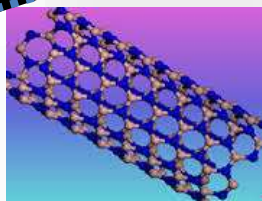
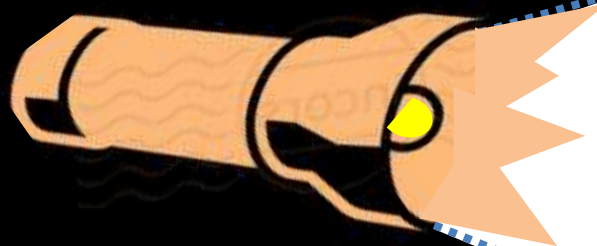
**TORNAR PÚBLICO OS RESULTADOS DAS PESQUISAS FINANCIADAS
POR VERBAS PÚBLICAS**

AVANÇO DO CONHECIMENTO E INOVAÇÃO

NOVAS VISÕES SOBRE ESSES DADOS



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.

AGÊNCIAS FINANCIADORAS DE PESQUISA

PLANOS DE COMPARTILHAMENTO DE DADOS

POLÍTICAS MANDATÓRIAS

Isso garante que os **pesquisadores se comprometem a cuidar dos dados** durante e após a pesquisa no sentido de otimizar o compartilhamento de dados.

PERIÓDICOS CIENTÍFICOS

Os periódicos exigem cada vez mais que os dados que sustentam a pesquisa publicada depositado dentro em uma **base de dados ou repositório** acessível .



INSTITUIÇÕES DE PESQUISA

Financiam/desenvolvem infraestruturas para gestão e serviços de dados para facilitar o compartilhamento dentro de domínios específicos.

PESQUISADORES

Iniciativas como o DataCite - que atribui identificador persistente (DOI) aos dados de pesquisa - ajudam o cientista a tornar seus dados citáveis, rastreáveis e acessíveis de modo que os dados de pesquisa, bem como as publicações baseadas nesses

AGÊNCIAS FINANCIADORAS DE PESQUISA

PLANOS DE COMPARTILHAMENTO DE DADOS

POLÍTICAS MANDATÓRIAS

Isso garante que os pesquisadores se comprometem a cuidar dos dados durante e após a pesquisa no sentido de otimizar o compartilhamento de dados.

PERIÓDICOS CIENTÍFICOS

Os periódicos exigem cada vez mais que os dados que sustentam a pesquisa publicada depositado dentro em uma **base de dados ou repositório** acessível .



nature.com

The world's best science and medicine on your desktop

Availability of data, material and methods

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. **Therefore, a condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications.** Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript, including details of how readers can obtain materials and information. If materials are to be distributed by a for-profit company, this must be stated in the paper.



CALL FOR RESEARCH PROPOSALS - ESCIENCE 2015

Characteristics of the research proposals

Data management plan: A major characteristic of eScience projects is its dependency on data management practices, and the **need of making results public, to allow reuse and collaboration with other groups.** Therefore, all projects should provide indication of how they intend to manage the data produced during the project (where the term "data" is taken on the large, and includes files, algorithms, software, samples, models, curriculum material and others).

COVERED BY

DATA
CITATION
INDEX

THOMSON REUTERS

INSTITUIÇÕES DE PESQUISA

Financiam/desenvolvem infraestruturas para gestão e serviços de dados para facilitar o compartilhamento dentro de domínios específicos.

Fornece um identificador persistente (DOI) aos dados de pesquisa - ajuda a manter os dados citáveis, rastreáveis e acessíveis de modo que os dados de pesquisa, a serem compartilhados, possam ser facilmente encontrados e utilizados em publicações baseadas nesses dados.

CIDADÃO COMUM

Transparência nas atividades de pesquisa
Compreender os benefícios da ciência
abrir a caixa preta da ciência

AGÊNCIA DE FOMENTO

Otimização de recursos públicos
Duplicação de esforços
Validação das pesquisas

CIÊNCIA ABERTA

Expansão do conceito de acesso livre
para dados, metodologia,
códigos, cadernos de laboratório

eSCIENCE

Significado, compartilhamento,
análises, pesquisa distribuída
Pesquisas interdisciplinares

INSTITUIÇÃO DE PESQUISA

Incorporar os dados na memória
Acadêmica. Reuso dos seus dados
Indicador de produtividade

PESQUISADOR

Reconhecimento, Citação,
Recompensa por preparar e
publicar os dados

EDITOR CIENTÍFICO

Expandir seus modelos de
negócio

PROFESSOR

Usar dados no ensino de ciências

DEMANDAM

POLÍTICA NACIONAL DE DADOS PESQUISA

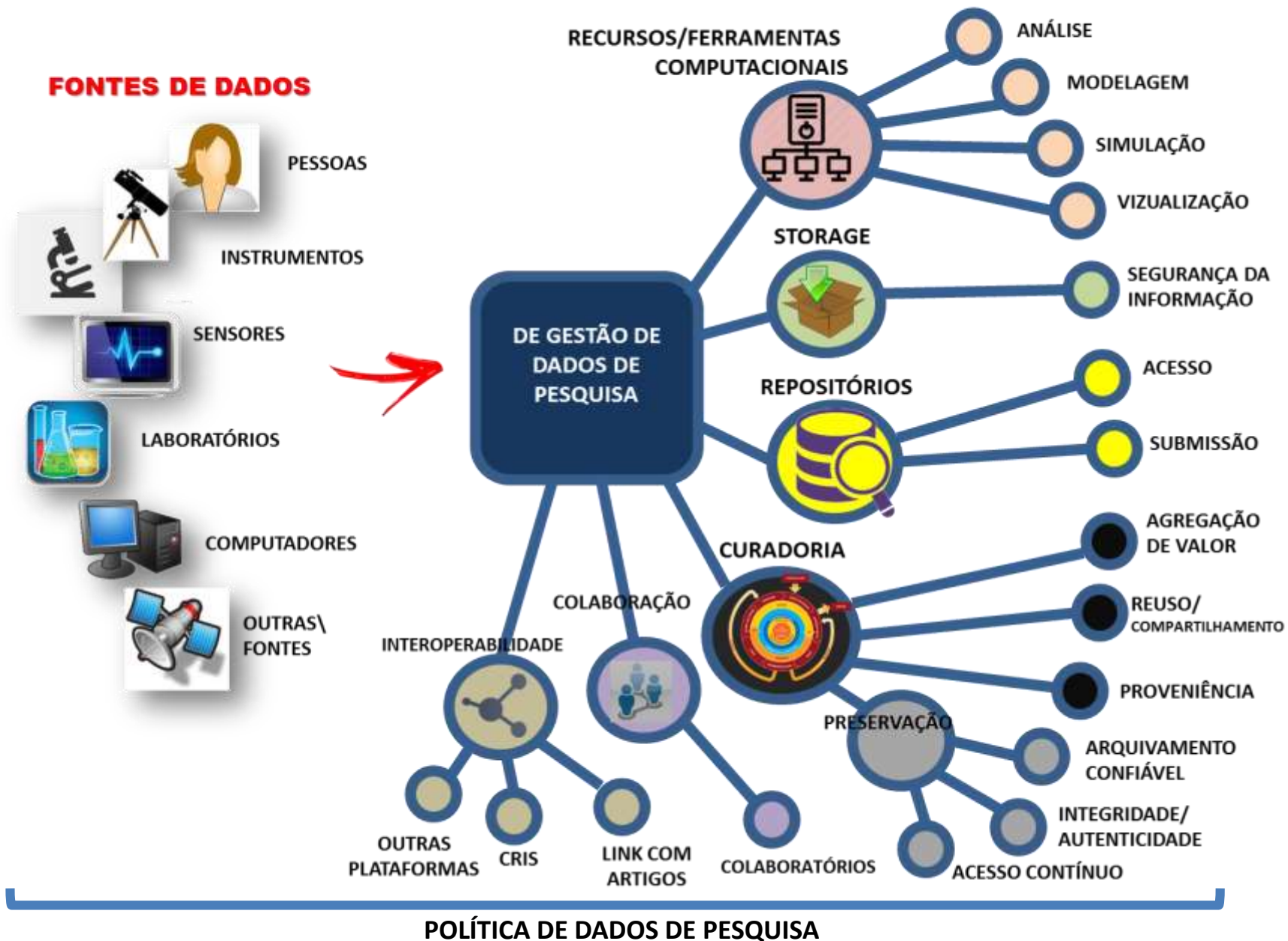
POLITICAS INSTITUCIONAIS

GESTÃO DE DADOS
DE PESQUISA

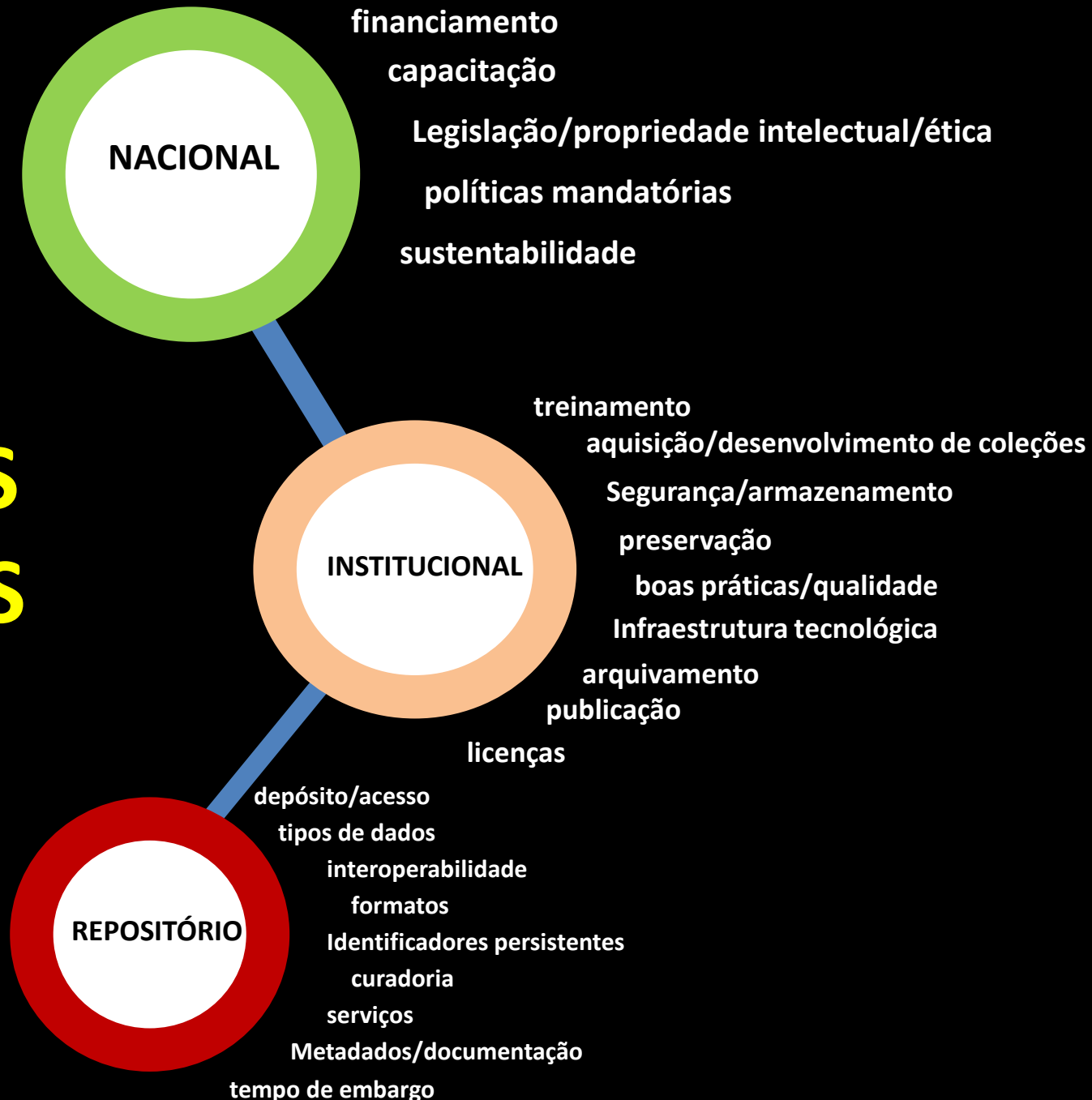
INFRAESTRUTURAS para

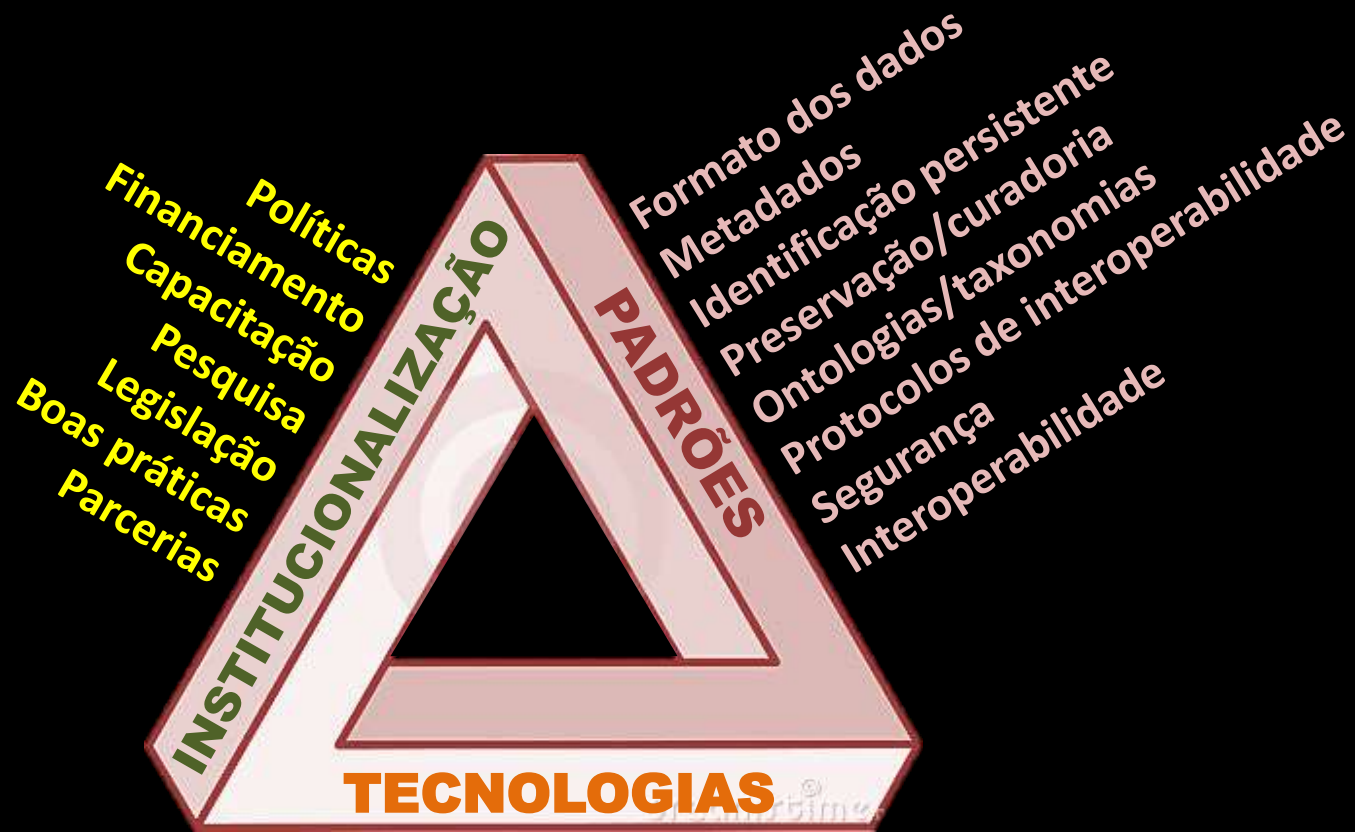
ORGANIZAÇÃO
DOCUMENTAÇÃO
ARQUIVAMENTO
PRESERVAÇÃO
CURADORIA
DESCOBERTA
COMPARTILHAMENTO
SEGURANÇA

CIBERINFRAESTRUTURA DE DADOS DE PESQUISA



POLÍTICAS DE DADOS





- Redes de computadores
- Banco de dados
- Ferramentas de software
- Sistemas de *storage*
- Repositórios confiáveis

COMPARTILHAMENTO




As potencialidades dos acervos digitais podem ser ampliadas se eles forem reconfigurados como matéria-prima para o empacotamento, resignificação, reinterpretação e agregações em novos contextos e com novos propósito, estabelecendo espaços de colaboração e interlocução que definem o conceito de reuso.



REUSO

DOS DADOS DE PESQUISA

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-Winshauer 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 the 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 initiative 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

	0	1	2	3	4
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0001	0.0001	0.0001	0.0001	0.0001
2	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
3	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
4	0.0005	-0.0004	0.0006	0.0009	0.0009
5	0.0000	0.0000	-0.0003	0.0006	0.0001
6	0.0014	0.0004	0.0001	0.0008	0.0018
7	0.0002	0.0002	-0.0006	0.0018	-0.0002
8	0.0000	0.0000	-0.0009	0.0009	0.0004
9	0.0000	0.0000	-0.0009	0.0002	0.0001
10	0.0008	0.0012	-0.0005	0.0002	0.0005
11	0.0000	0.0000	-0.0009	0.0002	0.0001
12	0.0000	0.0000	-0.0009	0.0002	0.0001
13	0.0000	0.0000	-0.0009	0.0002	0.0001
14	0.0000	0.0000	-0.0009	0.0002	0.0001
15	0.0000	0.0000	-0.0009	0.0002	0.0001
16	0.0000	0.0000	-0.0009	0.0002	0.0001
17	0.0000	0.0000	-0.0009	0.0002	0.0001



GESTÃO & CURADORIA

**ANALISADOS
EM NOVOS E
DIFERENTES
CONTEXTOS**

Os pesquisadores começaram a creditar **toda a confiança** nos conteúdos digitais **criados por outros pesquisadores** para dar prosseguimento aos seus empreendimentos

“

A ciência como um todo avança com **maior qualidade, menor custo e mais eficiência** quando abre a possibilidade para que o maior número possível de pesquisadores disponha de **vias de acesso aos dados acumulados por seus antecessores e contemporâneos**. Isso evita, objetivamente, **o custo da duplicação de esforços** e permite **novas interpretações em diferentes contextos científicos** para esses dados e, além do mais, permite que eles **sejam integrados e retrabalhados de forma mais criativa**, descortinando horizontes para **novas pesquisas**.



REUSO DE DADOS DE PESQUISA



Os dados que coletamos hoje podem ser **usados no futuro** de forma que ainda não conseguimos imaginar. Os exploradores de antigamente que **coletavam espécimes de plantas e animais não sabiam nada sobre DNA** e hoje as amostras são submetidas a esse tipo de investigação. Quando você coleta os seus dados, reúne informações que, no futuro, poderão ser analisadas de formas muito diferentes. **São coisas que terão um valor enorme para cientistas que nem nasceram** (POLIAKOFF, 2013).

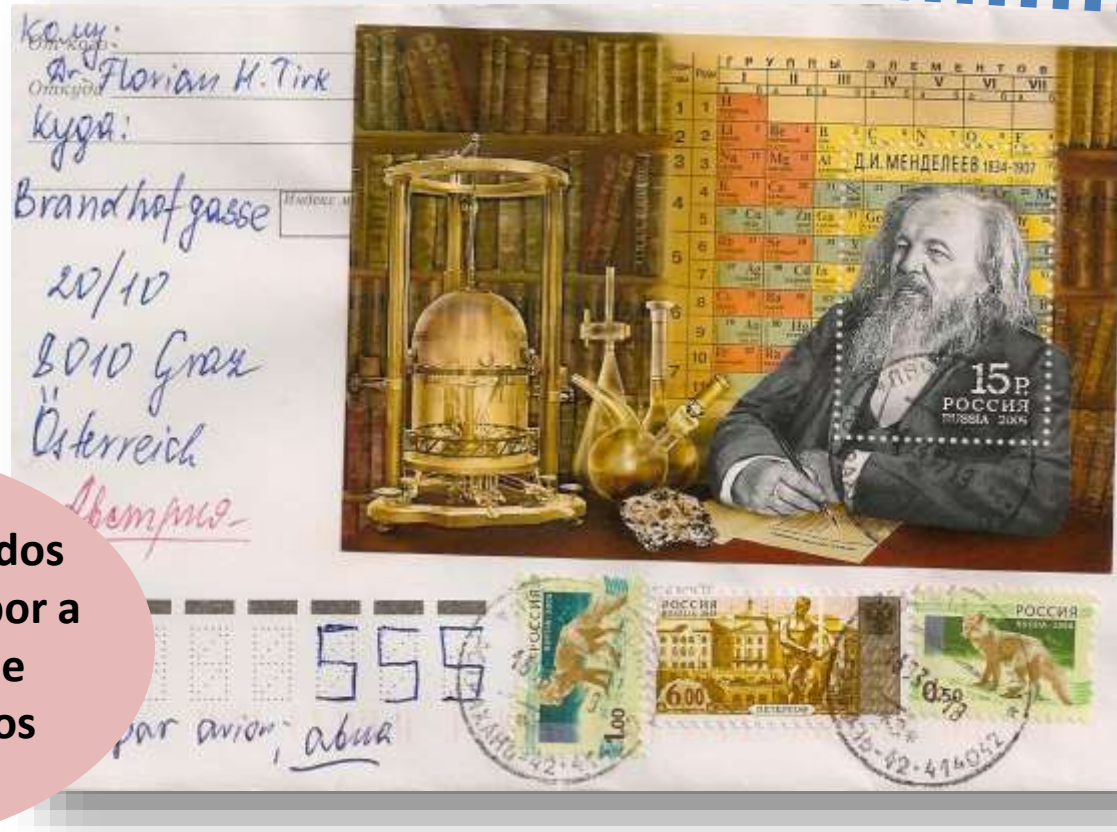


REUSO DE DADOS DE PESQUISA



Alguns cientistas estão procurando por registros astronômicos de 50 a 500 anos atrás...
GRAY, 2002

Mendeleev usou dados publicados para propor a Tabela Periódica de Elementos Químicos
GRAY, 2002



REUSO

EM OUTROS CONTEXTOS

ARQUIVOLOGIA



Centenas de diários de bordo digitalizados, registrando viagens marítimas de três séculos



DATA MINING



se tornam uma base de dados rica sobre a fauna, flora, corrente e ventos oceânicos

CLIMATOLOGIA



Cientistas reconstruem a história dos sistemas dinâmicos da Terra e melhoram as projeções sobre o futuro do clima

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

IMPORTÂNCIA DA PRESERVAÇÃO DOS DADOS DE PESQUISA

No período de 1918 a 1919 a **gripe espanhola** se espalhou pelo mundo inteiro, matando de 20 a 80 milhões de pessoas. De origem viral, não havia tratamento conhecido. Como veio, se extinguiu. Só depois de muito tempo, foi encontrada uma **amostra de tecido humano infectado pelo vírus num hospital militar da Inglaterra**. A partir desses vestígios estão sendo desenvolvidas pesquisas para se descobrir vacinas e meios de tratamento da gripe espanhola. As pesquisas em torno da amostra só se tornaram possíveis graças à **preservação dos arquivos científicos**, datados de 1916, daquele hospital militar (DITADI, 2003).

GRIFE ESPANHOLA



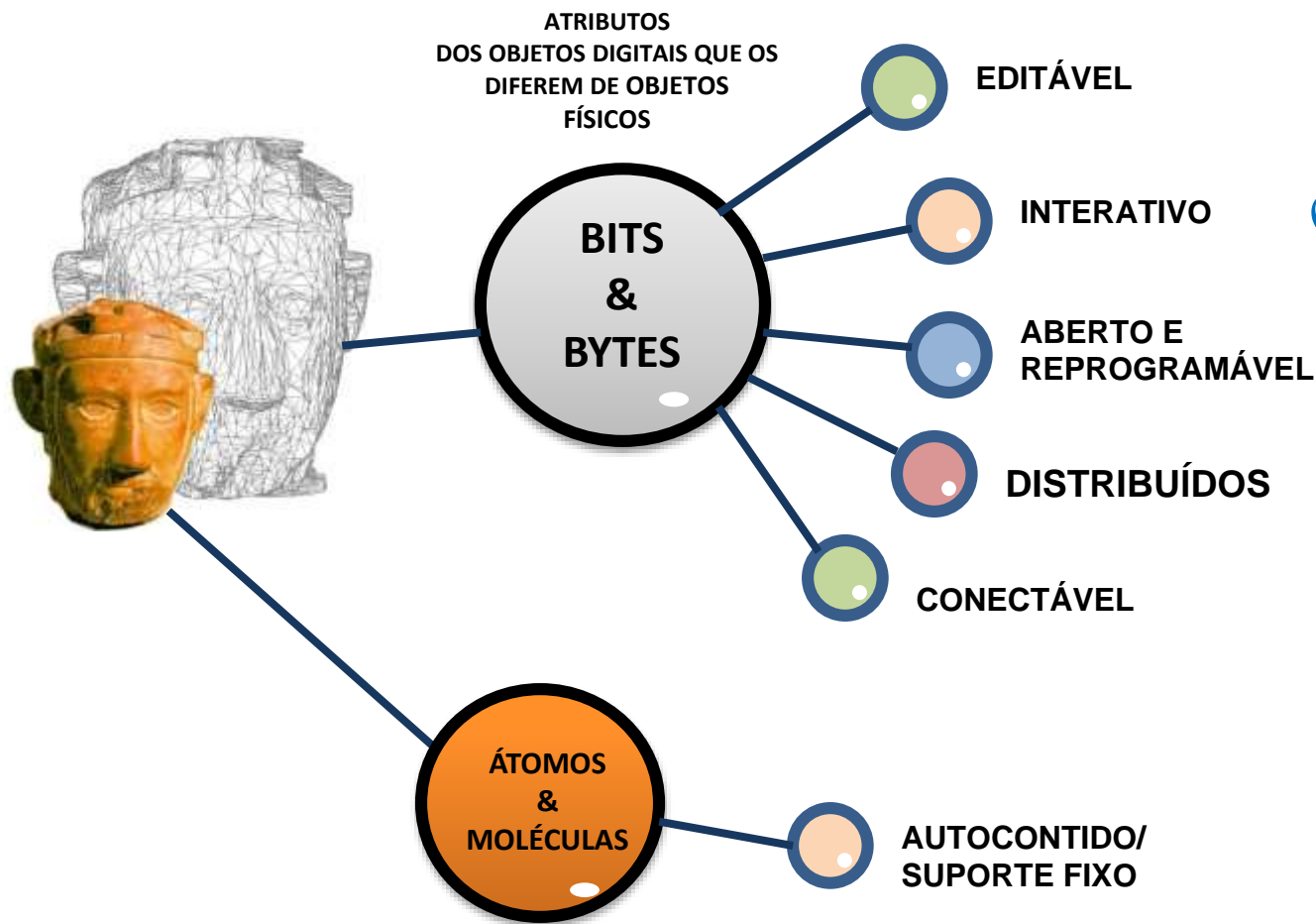
REUSO DE DADOS DE PESQUISA EM OUTROS CONTEXTOS

"Gerar, capturar e gerenciar coleções de dados para uso corrente e futuro não é um desafio trivial. As atividades que são locais e tácitas se tornam globais e explícitas".

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).



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



CRIATIVIDADE
NOVAS FORMAS DE AGREGAÇÃO, APRESENTAÇÃO, CONTEXTUALIZAÇÃO E INTERPRETAÇÃO

“compreender o meio digital como um facilitador de acesso e precursor de novas possibilidades de imersão nos lugares de memória, afastando-se de uma concepção **simplista do digital como mero repositório de informação**” (REIS *et al*, 2016)



REUSO

**COLEÇÕES
DIGITAIS**
00110010100
011001001

AGREGAÇÕES

**ESPAÇOS
COLABORATIVOS**

**CURADORIA
ON-LINE**

EDUCAÇÃO

**PESQUISA
CIENTÍFICA**

**APLICATIVOS
COMPUTACIONAIS**

POR QUE COMPARTILHAR

DADOS CRIADOS A PARTIR DE PESQUISAS SÃO RECURSOS VALIOSOS QPOR QUE PODEM SER USADOS E REUTILIZADOS PARA FUTUROS PROPÓSITOS CIENTÍFICOS E EDUCACIONAIS. O COMPARTILHAMENTO DE DADOS FACILITA NOVAS INVESTIGAÇÕES CIENTÍFICAS, EVITA A DUPLICAÇÃO DE ESFORÇOS NA GERAÇÃO DE DADOS E OFERECE UM RECURSO RICO, PROVENIENTE DA VIDA REAL, PARA A EDUCAÇÃO E TREINAMENTO.

INCENTIVA A **PESQUISA CIENTÍFICA E O DEBATE**

PROMOVE A **INOVAÇÃO** E POTENCIALIZA NOVOS USOS PARA OS DADOS

LEVA A NOVAS **FORMAS DE COLABORAÇÃO** ENTRE USUÁRIOS DE DADOS E CRIADORES DE DADOS

MAXIMIZA A **TRANSPARÊNCIA** E PRESTAÇÃO DE CONTAS

PERMITE O **CONTROLE DOS RESULTADOS** DA PESQUISA CIENTÍFICA

MELHORIA E **VALIDAÇÃO DA PESQUISA E A REVISÃO POR PARES**

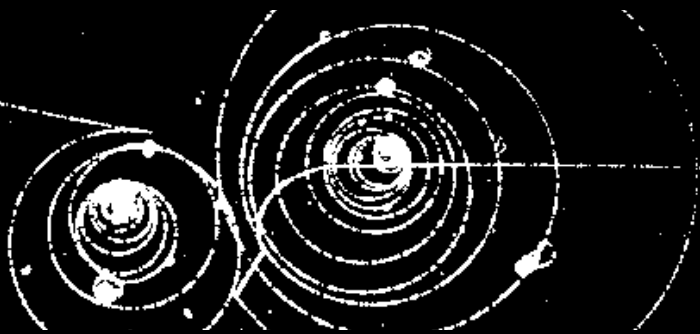
REDUZ O CUSTO DE DUPLICAÇÃO DE GERAÇÃO DE COLEÇÕES DE DADOS

AUMENTA A **VISIBILIDADE E O IMPACTO** DA PESQUISA

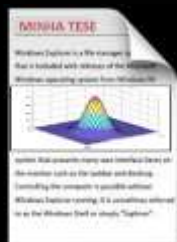
PROMOVE A PESQUISA QUE CRIOU OS DADOS E OS SEUS RESULTADOS

CREDITA AO PESQUISADOR AUTORIA PELOS DADOS ENQUANTO PRODUTO DA SUA PESQUISA

OFERECE RECURSOS IMPORTANTES PARA A **EDUCAÇÃO E TREINAMENTO**



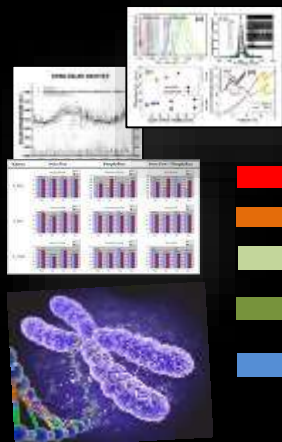
PUBLICAÇÕES



- PERIÓDICOS
- BIBLIOTECAS
- REPOSITÓRIOS INSTITUCIONAIS
- REPOSITÓRIOS TEMÁTICOS

COMO COMPARTILHAR DADOS DE PESQUISA

DADOS



- TORNANDO-OS **DISPONÍVEIS INFORMALMENTE** ENTRE PESQUISADORES DE PESSOA PARA PESSO
- TORNANDO-OS **DISPONÍVEIS NA WEB** NO SITE DO PROJETO OU DA INSTITUIÇÃO
- SUBMETÊ-LOS A UM **PERIÓDICO** PARA APOIAR UMA PUBLICAÇÃO
- DEPOSITANDO-OS EM UM **REPOSITÓRIO INSTITUCIONAL**
- DEPOSITANDO-OS NUM **REPOSITÓRIO DE DADOS**

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

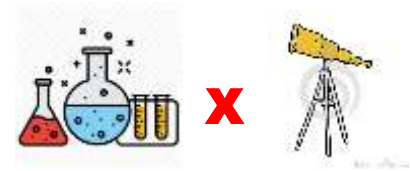
3

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)

**MOTIVOS
PARA O
PESQUISADOR
NÃO
COMPARTILHAR**

**Restrições culturais,
DISCIPLINARES e institucionais**



**INTERESSES ECONÔMICOS
(patentes, acordos comerciais, etc)**



**RESULTADOS NEGATIVOS,
hipóteses não confirmadas**



**CUSTO do tratamento dos dados
(limpeza, catalogação, formatos, etc.)**



**Perda da VANTAGEM COMPETITIVA de
publicar mais baseado nos dados**

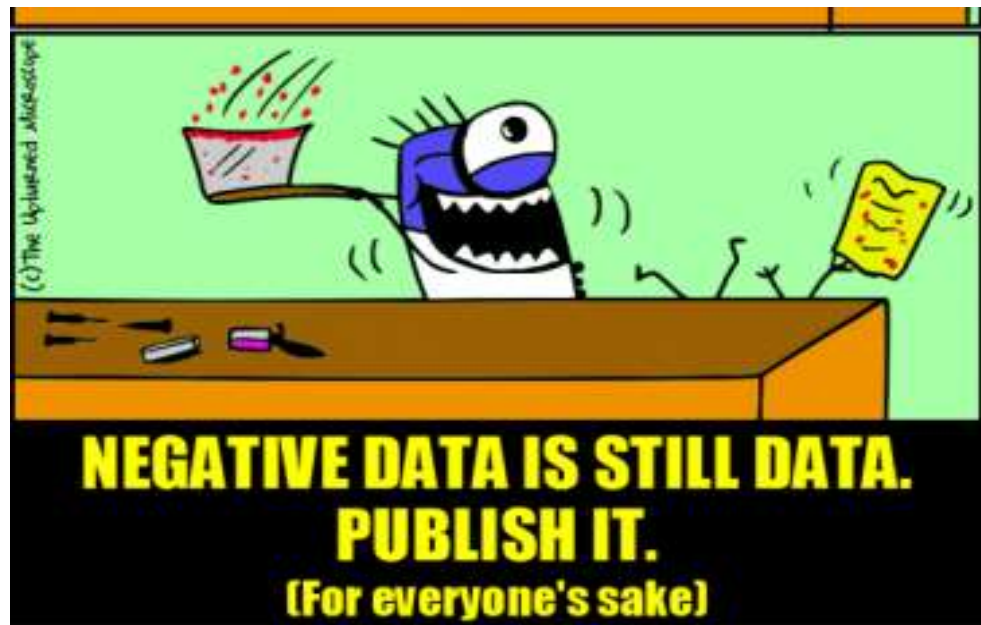
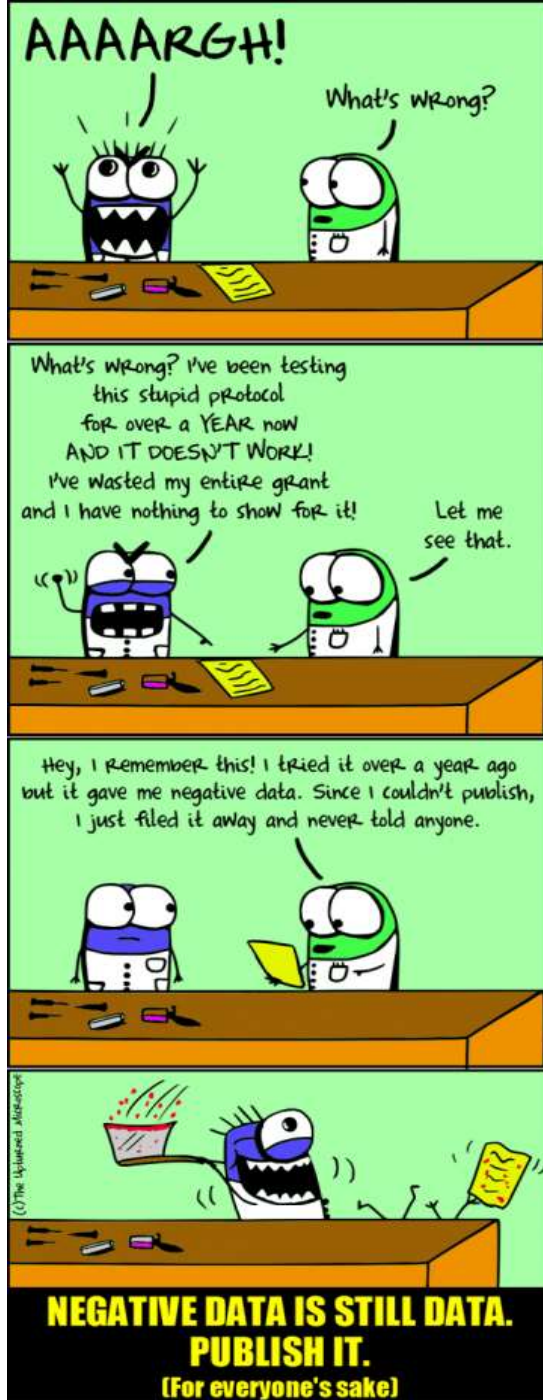


**Preocupação dos dados serem
ERRONEAMENTE INTERPRETADOS por
outros pesquisadores**



**Dificuldade de garantir a PRIVACIDADE
dos dados**

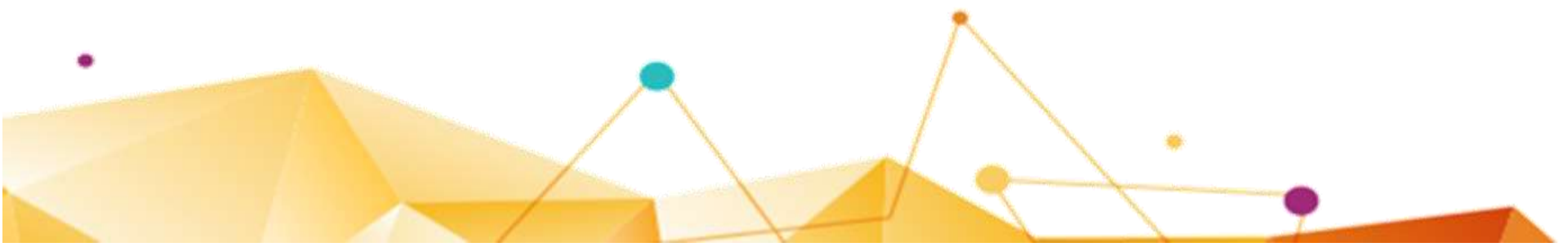




O “viés de publicação do positivo” preocupa há décadas diversos pesquisadores. Partindo da ideia de que a comunidade científica só pode aprender com os resultados negativos se os dados forem publicados, existem alguns **periódicos científicos que investem na publicação do que não deu certo em diversas áreas**. Tais periódicos têm como premissa a concepção de que o suposto “fracasso” é tão importante na ciência como em outros aspectos da vida, e que o progresso científico não depende apenas das realizações de indivíduos isolados, mas requer colaboração, trabalho em equipe e comunicação aberta com todos os resultados, sejam eles positivos ou negativos.

Fonte: <http://www.enago.com.br/blog/motivos-para-publicar-resultados-negativos/>

**A FINAL,
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;

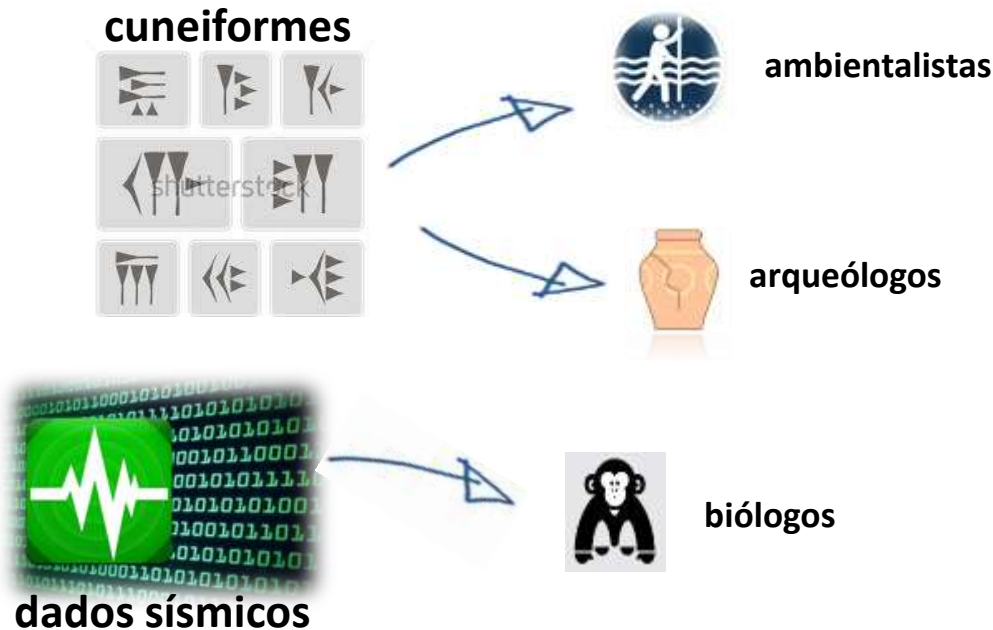


e inscrições em cuneiformes são dados de pesquisa para quem estuda linguagens do Oriente Próximo.

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."

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

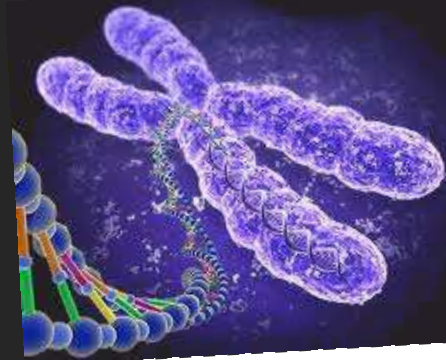


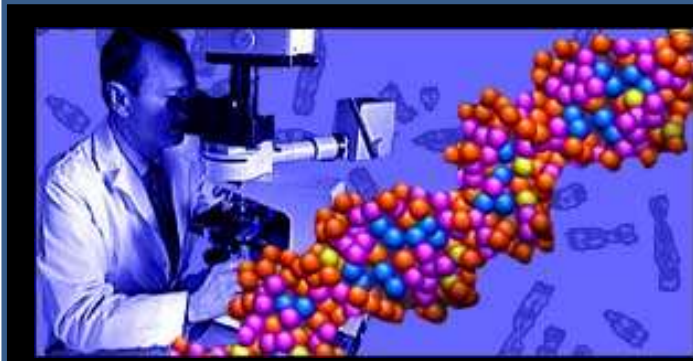
CRITICOS



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

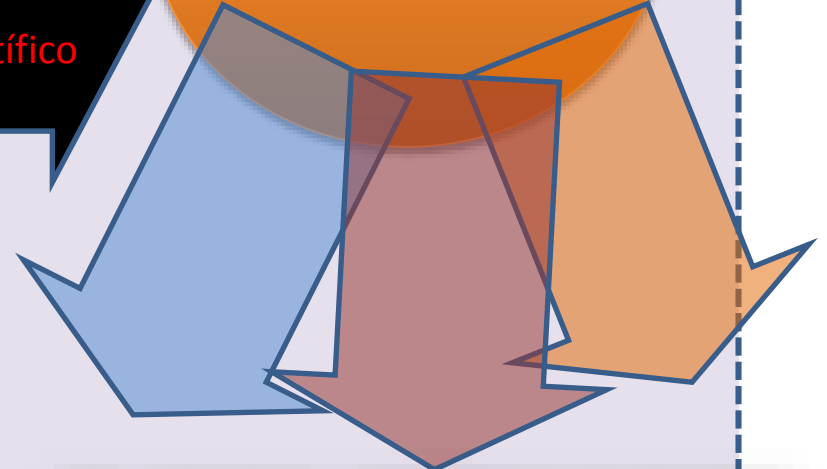




**DADOS BRUTOS
ou
DADOS PRIMÁRIOS**

Dados provenientes
diretamente do
instrumento científico

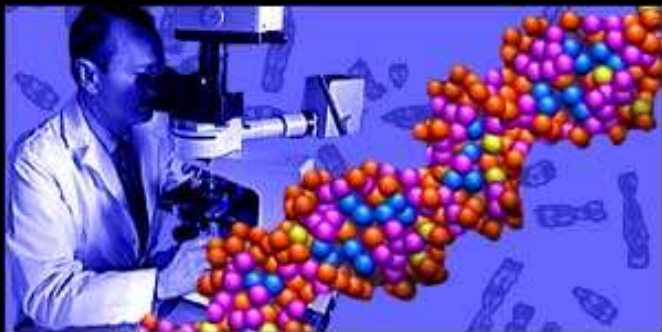
- .PROCESSAMENTO
- . CALIBRAÇÃO
- .VALIDAÇÃO
- .COMBINAÇÃO COM OUTROS DADOS



DADOS DERIVADOS

**DADOS REFERENCIAIS
OU CANÔNICOS**

Coleções de dados consolidados, revisados e geralmente passados por processos de curadoria que estão arquivados em centros de dados. Por exemplo: banco de dados de sequência genética, estruturas química, dados espaciais.



**DADOS BRUTOS
ou
DADOS PRIMÁRIOS**

Dados provenientes
diretamente do
instrumento científico

.PROCESSAMENTO
. CALIBRAÇÃO
.VALIDAÇÃO
.COMBINAÇÃO COM OUTROS
DADOS

DADOS DERIVADOS



DADOS GOVERNAMENTAIS

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

Dados sensíveis



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

Big data

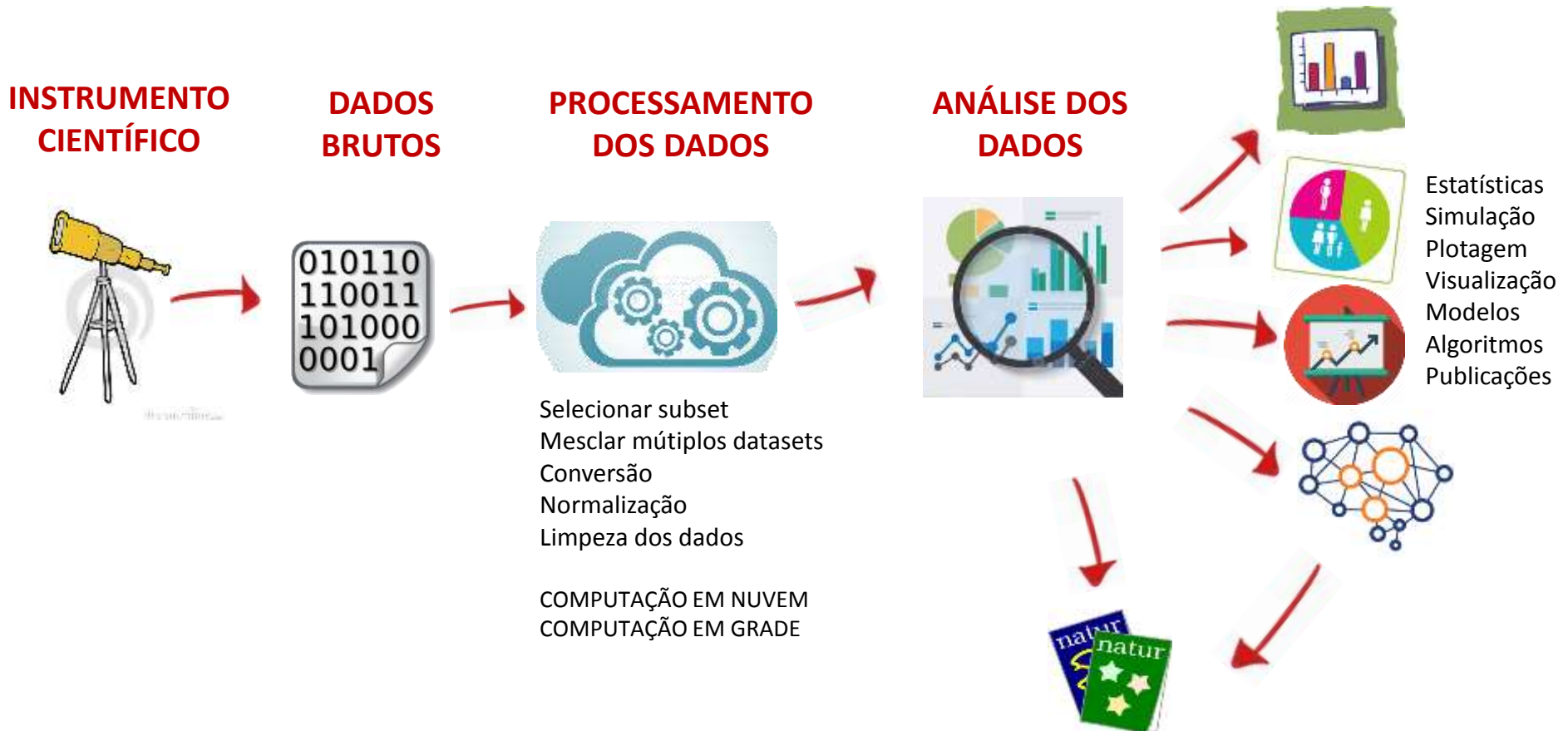
Comércio Transacional baseado em dados de negócios

DADOS REFERENCIAIS OU CANÔNICOS

Coleções de dados consolidados, revisados e geralmente passados por processos de curadoria que estão arquivados em centros de dados. Por exemplo: banco de dados de sequência genética, estruturas química, dados espaciais.

FLUXO DOS DADOS

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



FLUXO DOS DADOS

UM EXEMPLO DE PROCESSAMENTO DE DADOS BRUTOS



```
0711070500276000
0711070600276000
0711070700277003
0711070800282017
0711070900285000
0711071000293000
0711071100301000
0711071200304000
```



Date	time	air temp C	precip mm
11-Jul-07	5:00	27.6	000
11-Jul-07	6:00	27.6	000
11-Jul-07	7:00	27.7	003
11-Jul-07	8:00	28.2	017
11-Jul-07	9:00	28.5	000
11-Jul-07	10:00	29.3	000
11-Jul-07	11:00	30.1	000
11-Jul-07	12:00	30.4	000

Recreated from Michener & Brunt (2000)

FLUXO DOS DADOS



WORKFLOW; INSTRUMENTOS;
MODELOS; FERRAMENTAS;
CÓDIGOS; CADERNO DE
PESQUISA...



DADOS

São números, caracteres ou imagens que designam um atributo de um fenômeno

ELES SE TORNAM INFORMAÇÃO QUANDO...

INFORMAÇÃO

São combinados de maneira que têm o potencial de revelar padrões/ relações em um fenômeno

INFORMAÇÃO SE TRANSFORMA EM CONHECIMENTO...

CONHECIMENTO

Quando ele sustenta afirmações não triviais e verdadeiras sobre um fenômeno

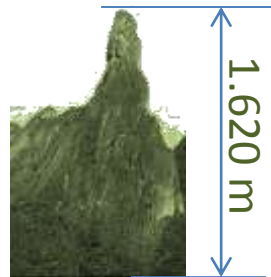


434330200034
095054820403
083110494692

OS NÚMEROS GERADO POR UM TEODOLITO

Medindo a altura de uma montanha são **dados**

Processamento, categorização e apresentação dos dados (empacotamento dos dados em informação utilizável)

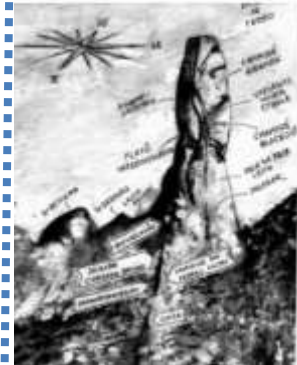


USANDO UMA FÓRMULA

A altura do pico pode ser deduzida a partir dos dados tornando-se

informação

Análise da informação para obter insights e/ou para resolver questões específicas (geração de conhecimento pela análise)

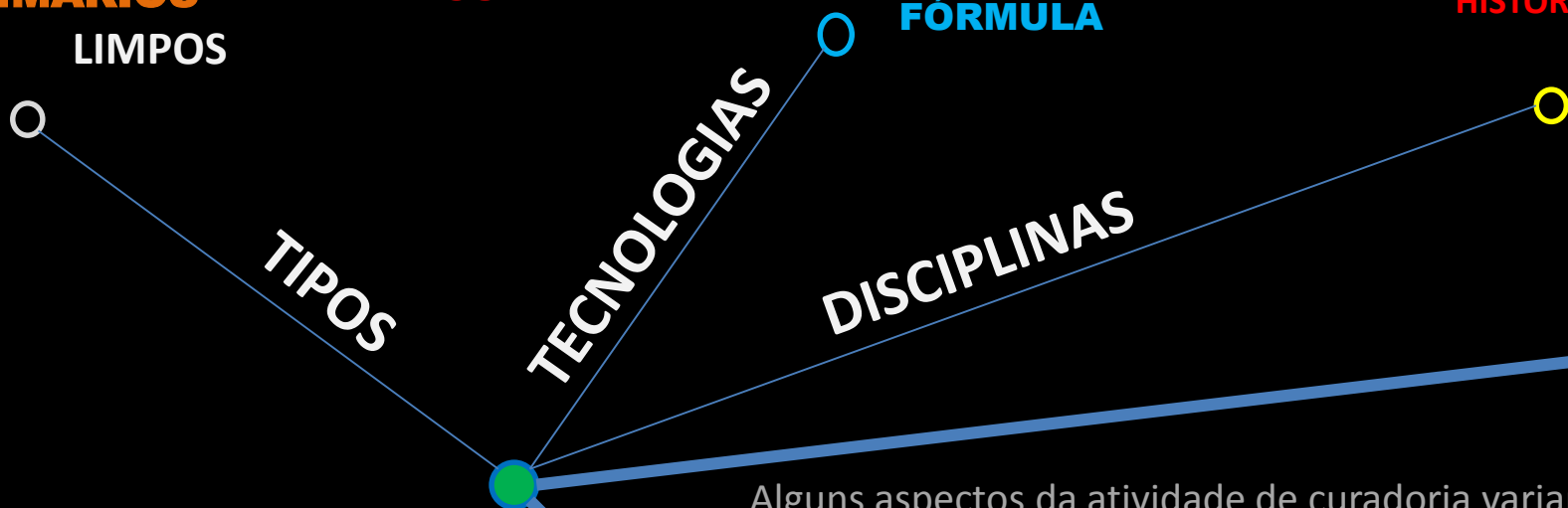


COMBINADOS COM OUTRAS INFORMAÇÕES

por exemplo, sobre as rochas da montanha, cria **conhecimento** sobre a origem da montanha



DINÂMICOS
ESTÁVEIS **EXPERIMENTAIS** **TEXTO** **ASTRONOMIA**
OBSERVACIONAIS **SIMULAÇÃO** **FÍSICA** **MEDICINA**
ÊFEMEROS **TERCIÁRIOS** **VIDEO** **SOFTWARE** **ECOLOGIA**
COMPUTACIONAIS **SUJOS** **NÚMEROS** **GRÁFICOS** **CIÊNCIAS SOCIAIS**
PRIMÁRIOS **DERIVADOS** **FÓRMULA** **HISTÓRIA**
LIMPOS



Alguns aspectos da atividade de curadoria variam amplamente de acordo com o **TIPO DE DADOS**, **TECNOLOGIAS SUBJACENTES AOS DADOS**, e, sobretudo, com o **DOMÍNIO DISCIPLINAR ESPECÍFICO**.

DADOS DE PESQUISA

“

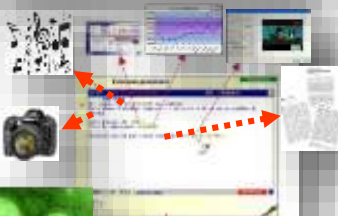
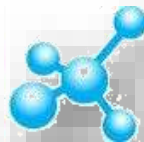
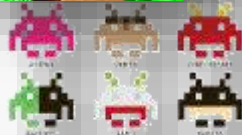
A heterogeneidade intrínseca aos dados de pesquisa implica que é necessário formular **políticas de amplo espectro**, que não só identifiquem, mas efetivamente sustentem os vários tipos de dados e a sua natureza díspar. O reconhecimento dessa idiosincrasia torna-se crucial quando se estabelecem as opções **tecnológicas e gerenciais** para a gestão desses dados.

Texto e números não contam toda história

DADOS DE PESQUISA SÃO OBJETOS COMPLEXOS, DIVERSIFICADOS E HETEROGÊNEOS.

OS OBJETIVOS E OS MÉTODOS USADOS PARA PRODUZI-LOS VARIAM ENORMEMENTE DE ACORDO COM OS CAMPOS CIENTÍFICOS, ASSIM COMO OS CRITÉRIOS PARA COMPARTILHÁ-LOS,

NÍVEIS DE ABSTRAÇÃO



"I've often asked what my term 'post-digital' signifies. It is a term that I created to use to acknowledge a distinction between those that are powered by the machine with computers, and are now driven by the idea based of the technology. It is not an expression of I address use in a world where the data that is 'post-digital' business. It was recorded the book by Nicholas Negroponte 'Being Digital' as an affirmation that the computer has not died, that the 'post-digital' generation will be the growing one that have already been digital, and are now more interested in being human. Being a good computer user. Being a good person is something that cannot be taught. It is the goal of life."

Copyright 2004, John Marko.

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022			
Population	5.3	5.5	5.7	5.9	6.1	6.3	6.5	6.7	6.9	7.1	7.3	7.5	7.7	7.9	8.1	8.3	8.5	8.7	8.9	9.1	9.3	9.5	9.7	9.9	10.1	10.3	10.5	10.7	10.9	11.1	11.3	11.5	11.7	11.9		
GDP	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3	4.5	4.7	4.9	5.1	5.3	5.5	5.7	5.9	6.1	6.3	6.5	6.7	6.9	7.1	7.3	7.5	7.7	7.9	8.1	8.3	8.5

REALIDADE VIRTUAL

GAMES

SIMULAÇÕES

MODELOS EM 3D

ESTRUTURAS QUÍMICAS

SOFTWARE

WEBSITE/MULTIMÍDIA

VIDEOS

FOTOS

GRÁFICOS

ESPECIFICAÇÕES

ENTREVISTAS

FORMÚLAS

TABELAS

ANOTAÇÕES

DADOS NUMÉRICOS

dispositivos de imersão e interativas

apresentações sensoriais

OBJETOS DIGITAIS COMPLEXOS

imagem em movimento

imagens

sons

documentos

letras

símbolos

números

REALIDADE



ABSTRAÇÃO



SIMULAÇÃO



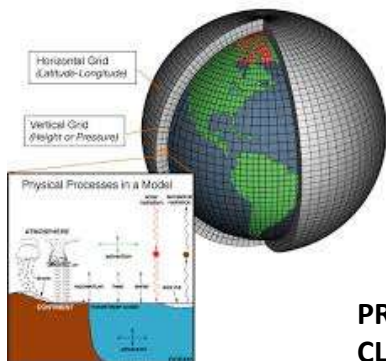
MODELO MATEMÁTICO



ALGORÍTMO NUMÉRICO



SOFTWARE



PREDIÇÕES CLIMÁTICAS

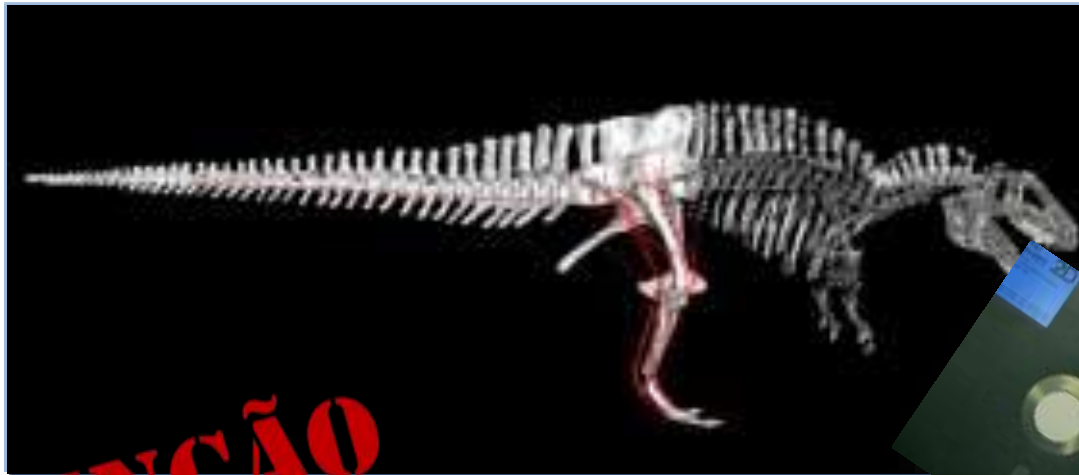
VIRTUAL SIMULATION OF A NUCLEAR POWER PLANT'S CONTROL ROOM AS A TOOL FOR ERGONOMIC EVALUATION.



Leandro Barbosa S. Gatto ^a, Antônio Carlos A. Mól ^{a,b,c}, Isaac J.A. Luquetti dos Santos ^a, Carlos Alexandre F. Jorge ^{a,*}, Ana Paula Legey ^c

INSTITUTO DE ENGENHARIA NUCLEAR - CNEN/IN

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

É NECESSÁRIO COMPARTILHAR MUITO MAIS DO QUE OS DADOS FINAIS PARA A PESQUISA SEJA REPRODUTÍVEL



WORKFLOW; INSTRUMENTOS;
MODELOS; FERRAMENTAS;
CÓDIGOS; CADERNO DE
PESQUISA...

É MAIS DIFÍCIL VALIDAR OS DADOS SE AS INFORMAÇÕES ENTRE A IDEIA INICIAL E RESULTADO FINAL ESTÃO FALTANDO

POR QUE?

QUEM?

O QUE?

COMO?

QUANDO?

ONDE?



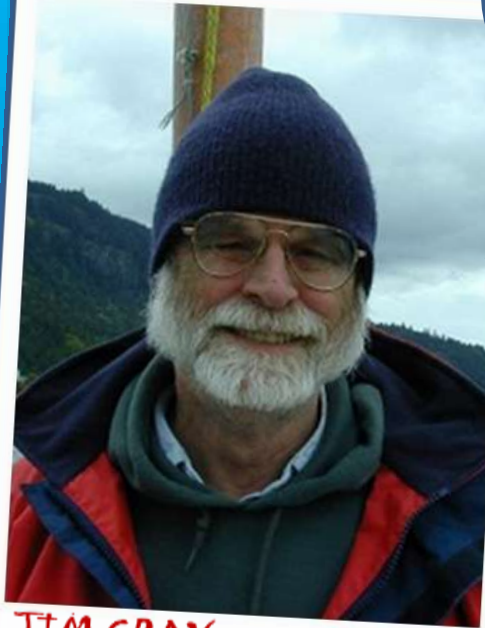
SIGNIFICADO
ESTRUTURA
IDENTIFICAÇÃO
CONTEXTO
PROVENIÊNCIA

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 !!!

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



JIM GRAY
Cientista da Computação
Desaparecido em 2007



04

27

56

01

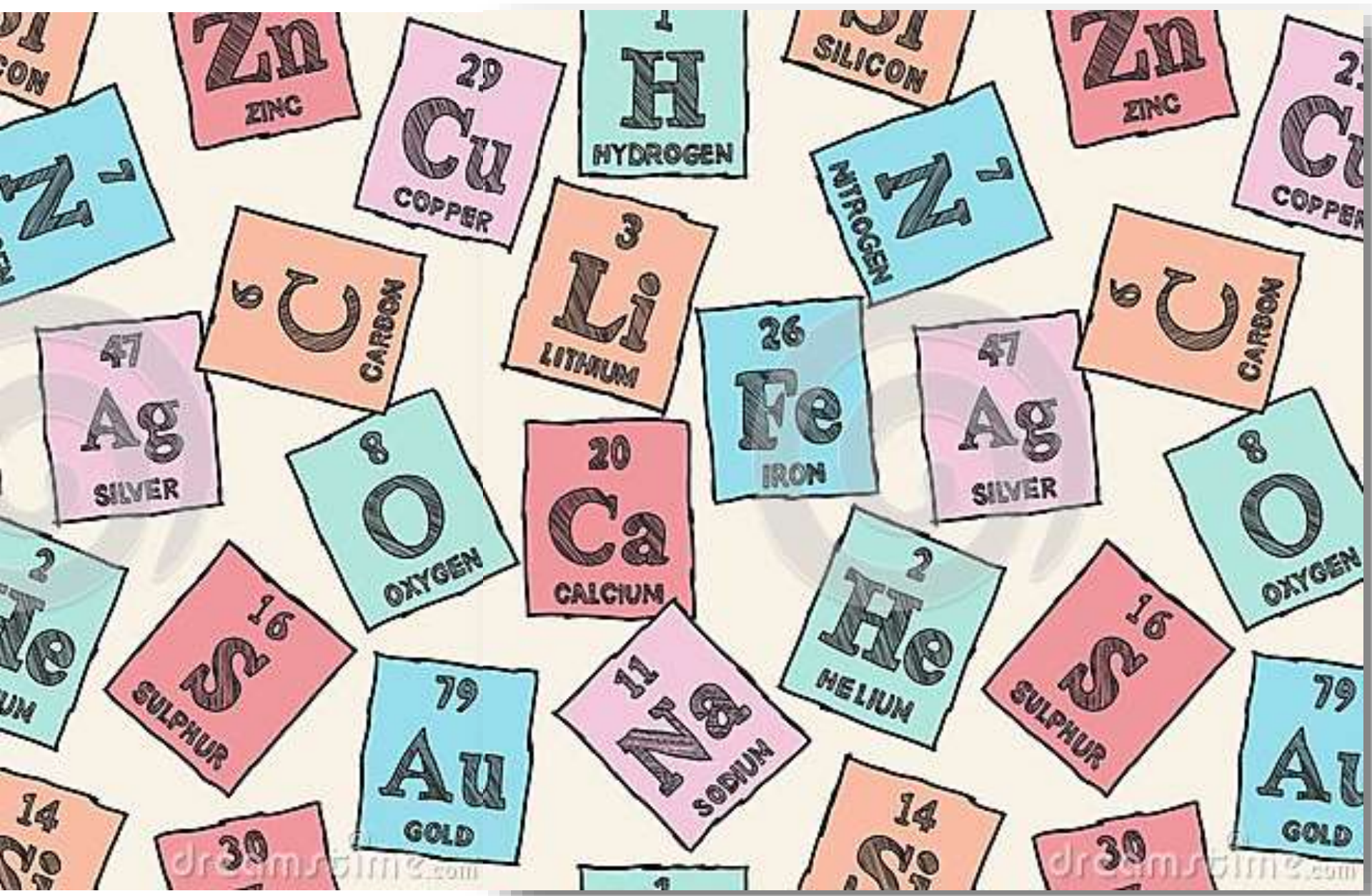
16

44

02

01

17



INFORMAÇÕES CONTEXTUAIS: SEMÂNTICA E ESTRUTURA

1 18
IA IIA ... VIIIA VIIIA

1 H 2 He
1.00794 4.002602

3 Li 4 Be
6.941 9.012182

11 Na 12 Mg
22.98976928 24.304

19 K 20 Ca
39.0983 40.078

37 Rb 38 Sr
85.4678 87.62

55 Cs 56 Ba
132.90545196 137.327

87 Fr 88 Ra
[223] [226]

89 to 103

Metals alcalinos
Metals alcalinos-ferrosos
Metals de transição
Lantanídeos
Actínidos
Outros metais
Não-Metals
Gases nobres
Sólidos
Líquidos
Gases
Sintético

Massas atômicas em parênteses são aquelas do isótopo mais estável ou comum.

Nota: Os números do subgrupo 1-10 foram adotados em 1994 pela International Union of Pure and Applied Chemistry (União Internacional de Química Pura e Aplicada). Os nomes dos elementos 112-118 são os equivalentes tabulares desses números.

57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
138.90547	140.12	140.90768	144.24	[145]	150.36	151.964	157.25	158.92534	162.50	164.93032	167.259	168.93421	173.04	174.967
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr
[227]	232.0377	231.03688	238.02891	[237]	244	243	247	247	251	252	257	288	289	260

Vida, sem destino
sempre escorrendo entre
os dedos nada podemos
fazer tão veloz, sem
piedade onde parar? Não
sabemos

Haroldo de Campos

DESTINO

**SEM
VIDA,**

**SEMPRE
ESCORRENDO
ENTRE**

Vida, sempre escorrendo entre os dedos

NADA

PODEMOS

FAZER

**TÃO VELOZ,
SEM PIEDADE
ONDE PARAR?**

os dedos podemos fazer tão veloz, sem piedade parar? Não

sabemos

**NÃO
SABEMOS**

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



OBJETO DE DADOS

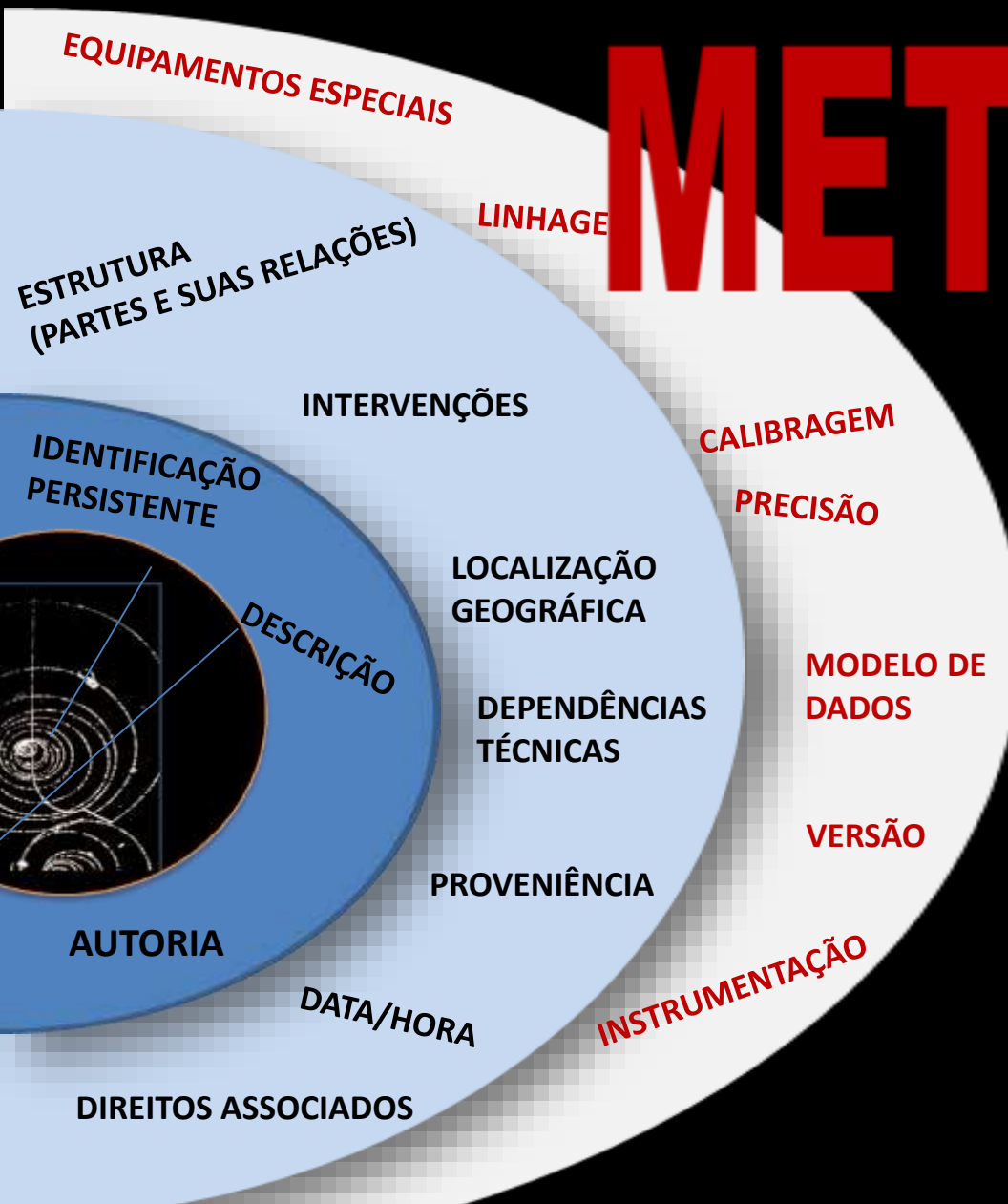


DOCUMENTAÇÃO



OBJETO DE INFORMAÇÃO

METADADOS



Os metadados têm um forte impacto na capacidade dos dados de pesquisa de transmitir conhecimentos e poder ser interpretados e reusados agora e no futuro

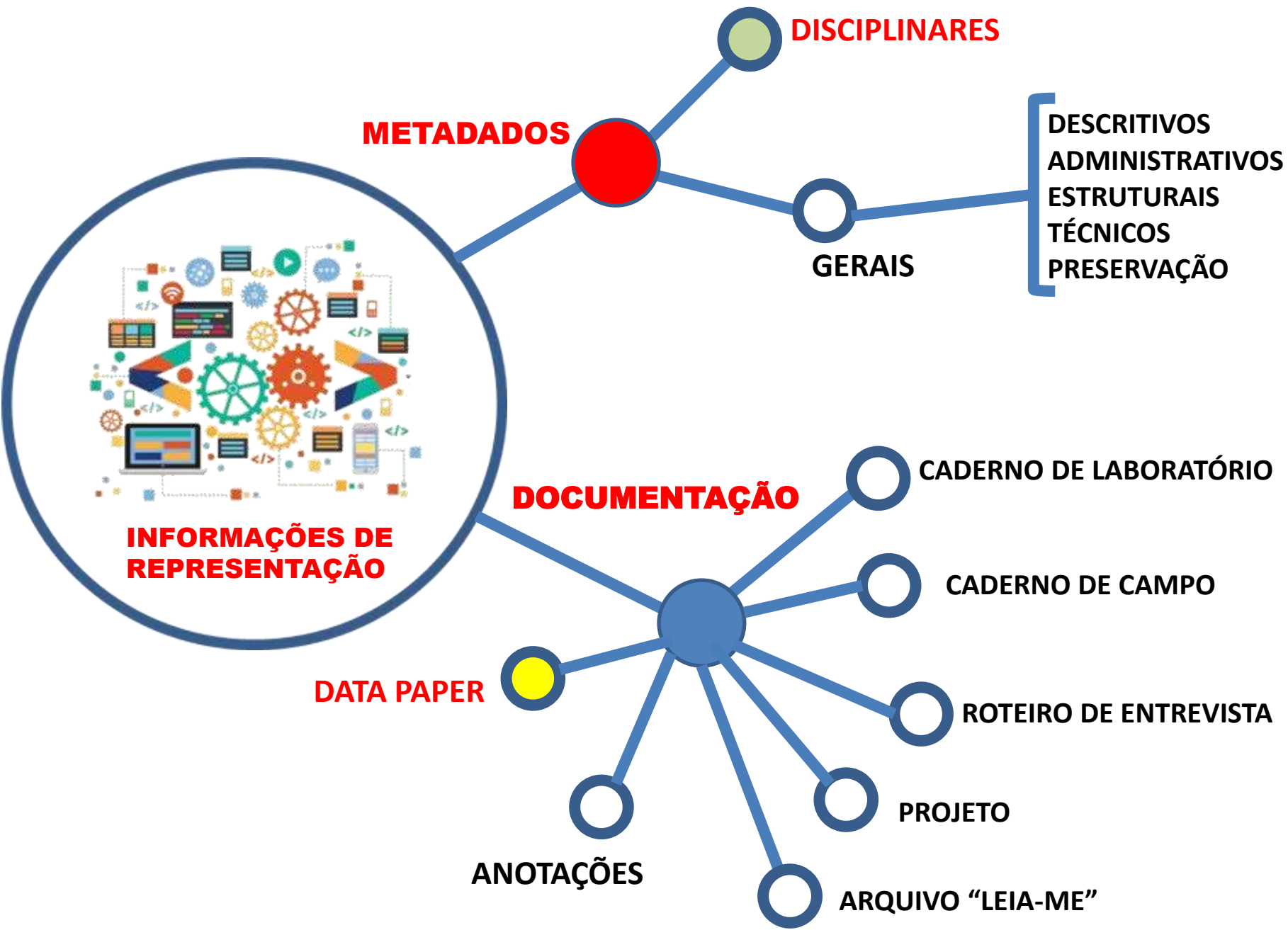
- DESCRITIVOS**
- ADMINISTRATIVOS**
- TÉCNICOS**
- ESTRUTURAIS**
- PRESERVAÇÃO**
- DISCIPLINARES**

CALIBRAGEM

PRECISÃO

MODELO DE DADOS

VERSÃO



DATApaper

Uma publicação periódica científica cujo objetivo principal é **descrever coleções de dados** ao invés de reportar uma investigação científica

DESCREVE

os dados em forma legível por humanos

A **metodologia** sobre a qual os dados forma criados;

Detalha o **potencial de reuso** dos dados

DESCREVE OS DADOS e não hipóteses ou argumentos desenvolvidos sobre os dados

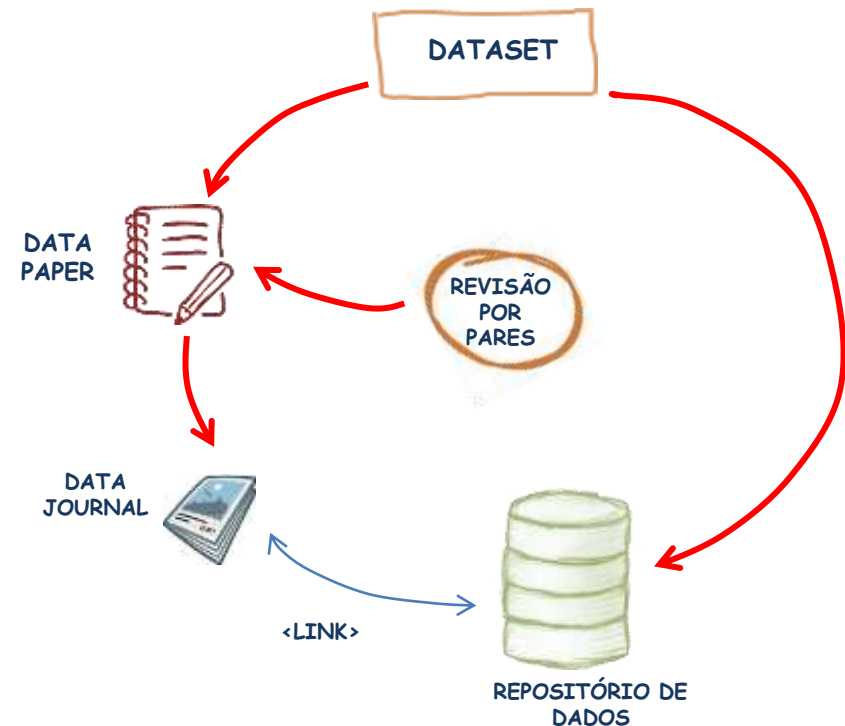
Oferecer uma publicação que **pode ser citada** e que dá **credito ao autor** e o outros envolvidos no processo;

Assegura que os dados estejam **documentados para o reuso**;

Aumenta a visibilidade dos dados na comunidade científica



A novidade interessante introduzida pelos *data journals* é que o modelo propõe um processo de publicação para dados que remete a publicação tradicional [...] A revisão por pares objetiva mensurar a originalidade e qualidade dos dados, ela é aplicada aos dados ao invés da publicação, e a sua “benção” é mandatória para os que os dados sejam publicados (CASTELLI et al, 2013)





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

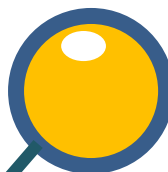
FERRAMENTAS DE APOIO À CITAÇÃO

EXPORTAÇÃO EM FORMATOS DIVERSOS/COMPARTILHAMENTO

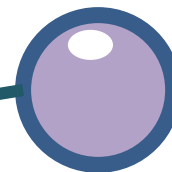
O controle de versões é um processo importante para o fundamento da reprodutibilidade da pesquisa, para a integridade da referência às coleções de dados e para proveniência dos seus conteúdos. Isto por que as coleções de dados podem evoluir no tempo por vários motivos



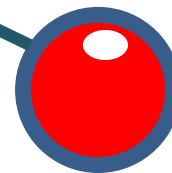
Consórcio internacional que enfrenta o desafio de tornar os dados **CITÁVEIS** de forma harmonizada, interoperável e persistente



**DOI
IDENTIFICADOR
PERSISTENTE**

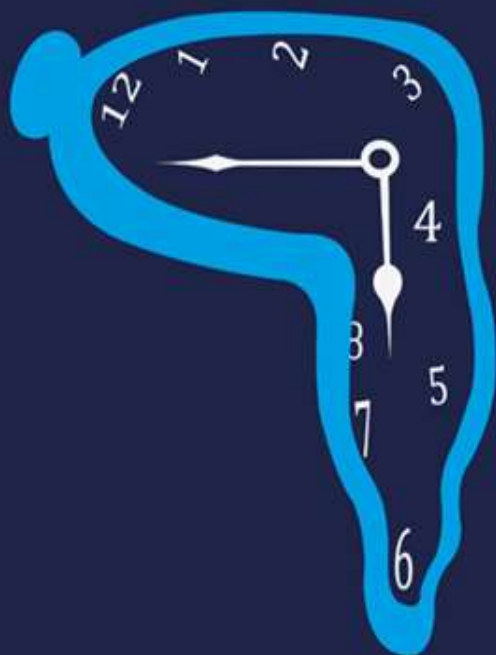


**ESQUEMA DE
METADADOS**



**FORMATADOR
DE
CITAÇÃO**

VERSION 01
VERSION 02
VERSION 03
VERSION 04
VERSION 05



IDENTIFICADAS/CITADAS

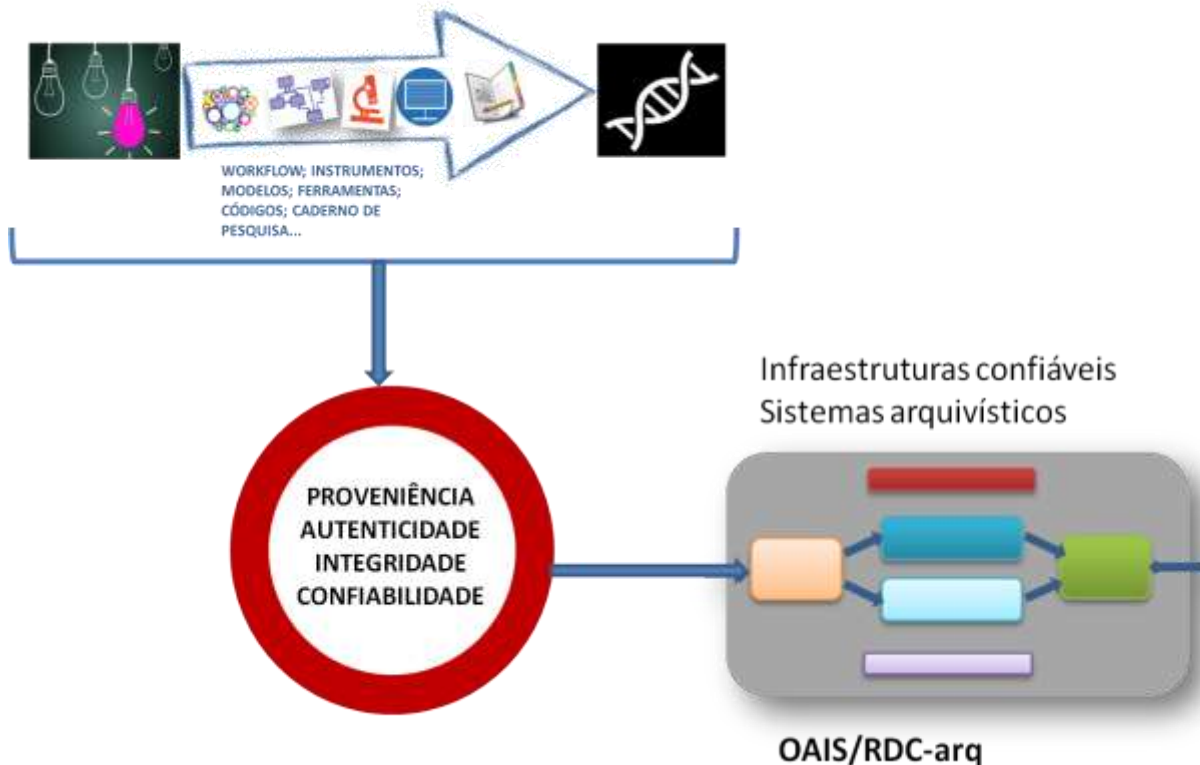


UNF – UNIVERSAL NUMERIC FINGERPRINT



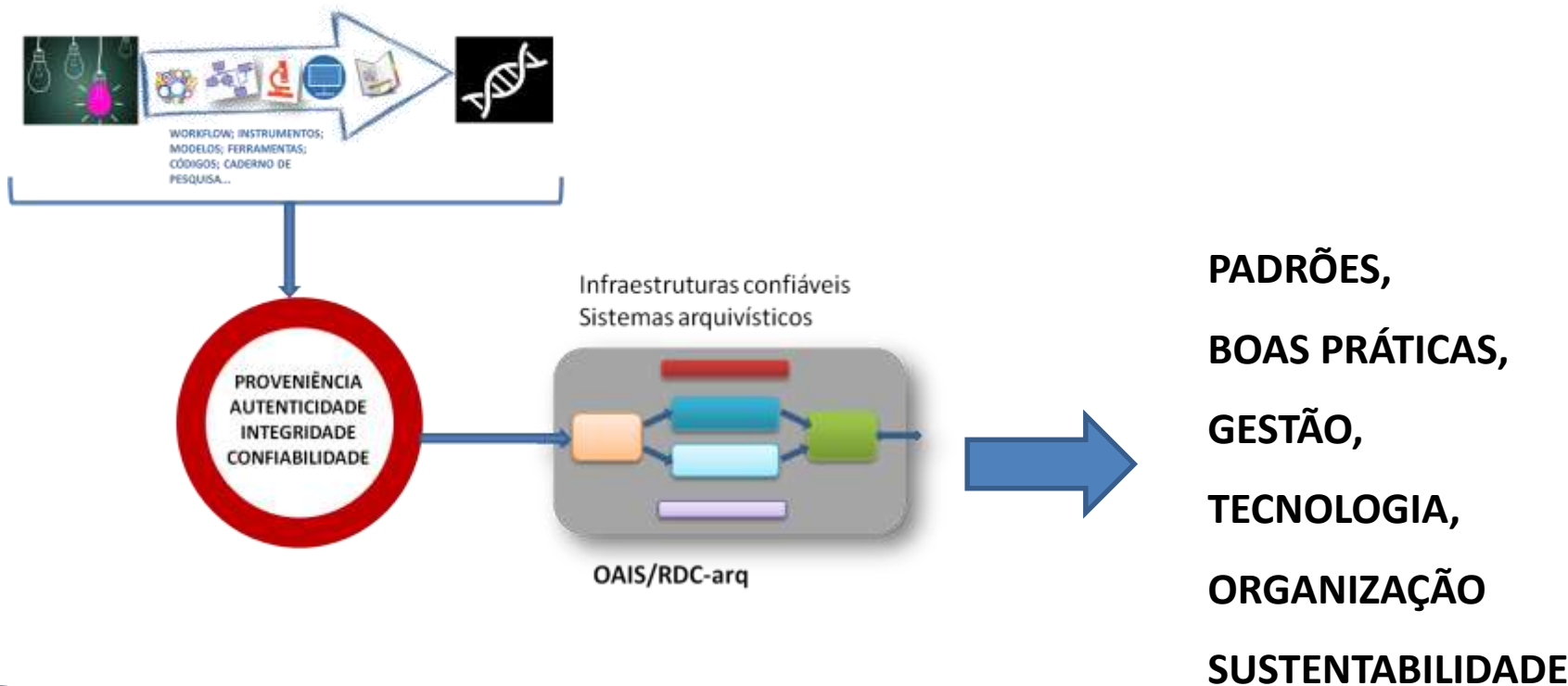
TME STAMP

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**.

A IMPORTÂNCIA DOS PROCESSOS DE AVALIAÇÃO RESIDE NO FATO DELES PROMOVEREM A CONFIANÇA NA USABILIDADE, SUSTENTABILIDADE E PERSISTÊNCIA POR LONGO PRAZO DOS DADOS DISPONÍVEIS PARA COMPARTILHAMENTO.

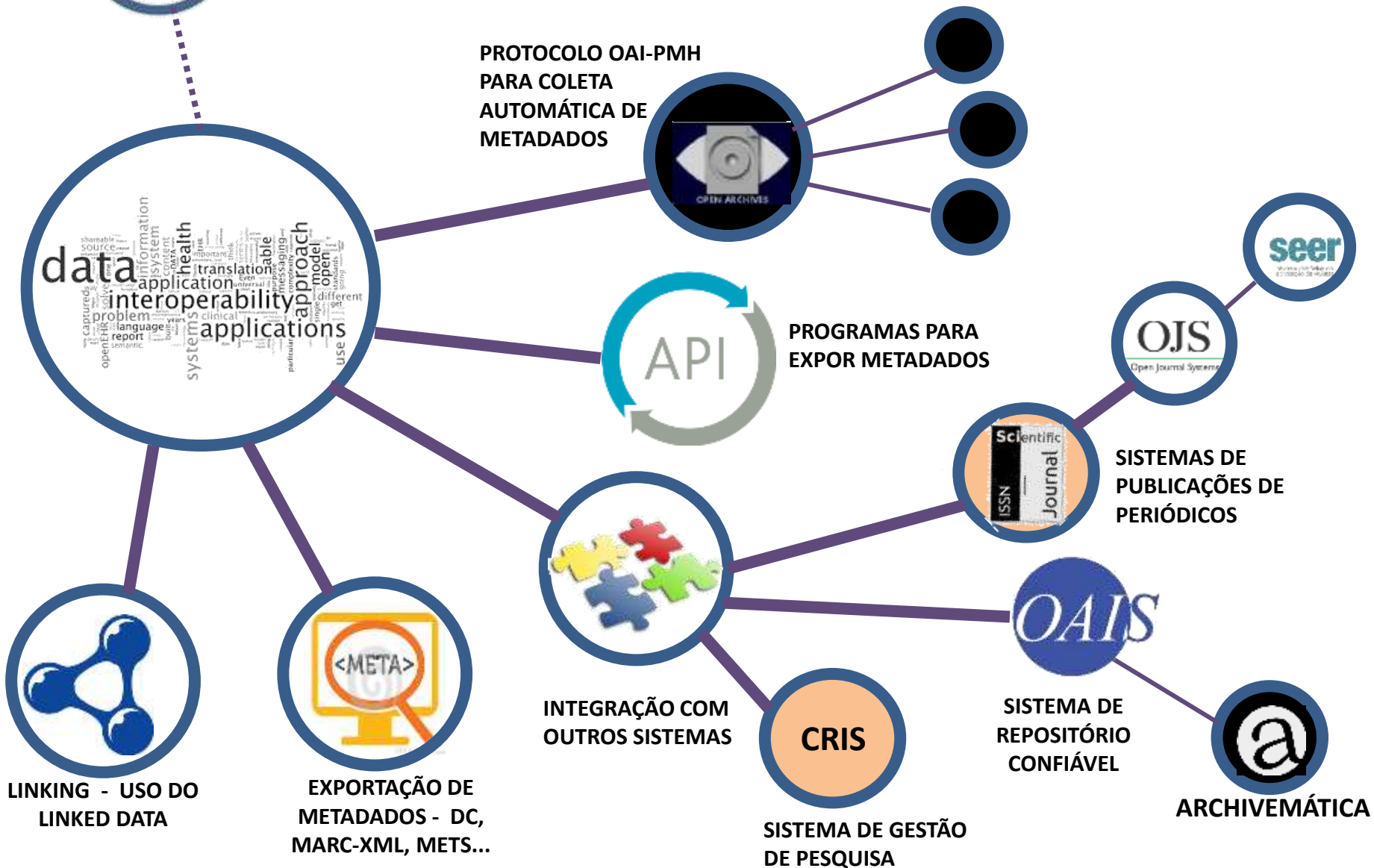




**KEEP
CALM**

**E ACREDITE
NOS
REPOSITÓRIOS
DIGITAIS
CONFIÁVEIS**

A INTEROPERABILIDADE DE SISTEMAS É COMPREENDIDA AQUI COMO A CAPACIDADE DAS PLATAFORMAS DE REPOSITÓRIOS DE DADOS INTERCAMBIAREM INFORMAÇÕES – **DADOS E METADADOS** - COM SISTEMAS EXTERNOS DE FORMA HARMÔNICA E INTEGRADA E COM PROPÓSITOS ESPECÍFICOS.



LINKING - USO DO LINKED DATA

EXPORTAÇÃO DE METADADOS - DC, MARC-XML, METS...

PROTOCOLO OAI-PMH PARA COLETA AUTOMÁTICA DE METADADOS

PROGRAMAS PARA EXPOR METADADOS

INTEGRAÇÃO COM OUTROS SISTEMAS

CRIS SISTEMA DE GESTÃO DE PESQUISA

OAIS SISTEMA DE REPOSITÓRIO CONFIÁVEL

ARCHIVEMÁTICA

SISTEMAS DE PUBLICAÇÕES DE PERIÓDICOS

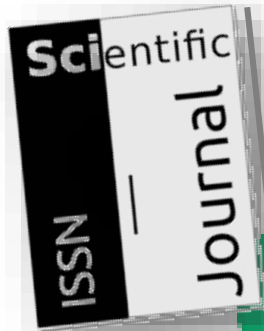
OJS Open Journal Systems

seer

Sistema de Seleção e Avaliação de Periódicos

Scientific Journal ISSN

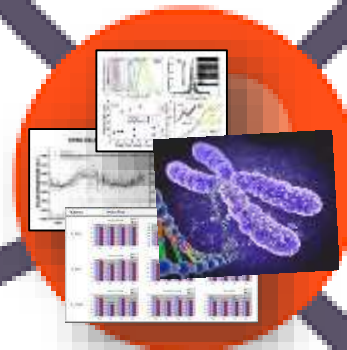
INTEROPERABILIDADE POSSÍVEL



SISTEMAS DE PERIÓDICOS



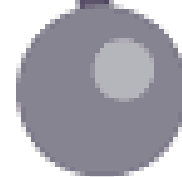
REPOSITÓRIOS INSTITUCIONAIS



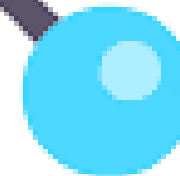
REPOSITÓRIOS DE DADOS



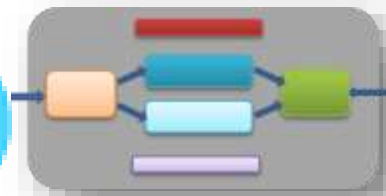
SISTEMAS CRIS
GESTÃO DE PESQUISA
GESTÃO ACADÊMICA
PLATAFORMA LATTES



CLOCKSS



REPOSITÓRIOS CONFIÁVEIS



PUBLICAÇÕES AMPLIADAS



TRANSMITIR CONHECIMENTO!

Disponibilizar os dados Internet é apenas uma das etapas de um ciclo complexo, e que isoladamente não garante que os dados possam ser acessados, reusados, e, sobretudo, terem seus significados e estruturas recompostos agora e no futuro.

(DADOS INTELIGENTEMENTE ABERTOS)

TRANSMITIR CONHECIMENTO!

1

ACESSÍVEL

Localizado e acessado

2

INTELIGÍVEL

Deve se apresentar de forma inteligível para aqueles que desejam entendê-lo ou analisá-lo. Deve ser diferente para diferentes audiências, mesmo para o cidadão comum.

3

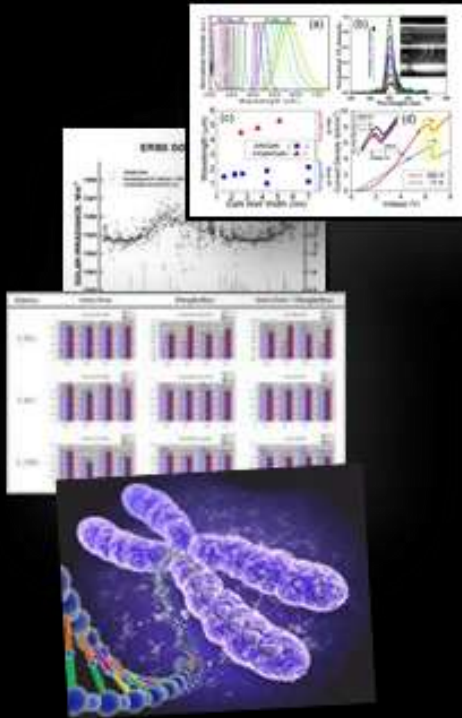
AVALIÁVEL

Natureza, evidências, confiabilidade, fonte competente, financiador, objetivos

4

UTILIZÁVEL

Deve ser capaz de ser reusado em diferentes contextos e, minimamente, por outros cientistas.



TRANSMITIR CONHECIMENTO!

1

ACESSÍVEL

Localizado e acessado

2

INTELIGÍVEL

Deve se apresentar de forma inteligível para aqueles que desejam entendê-lo ou analisá-lo. Deve ser diferente para diferentes audiências, mesmo para o cidadão comum.

3

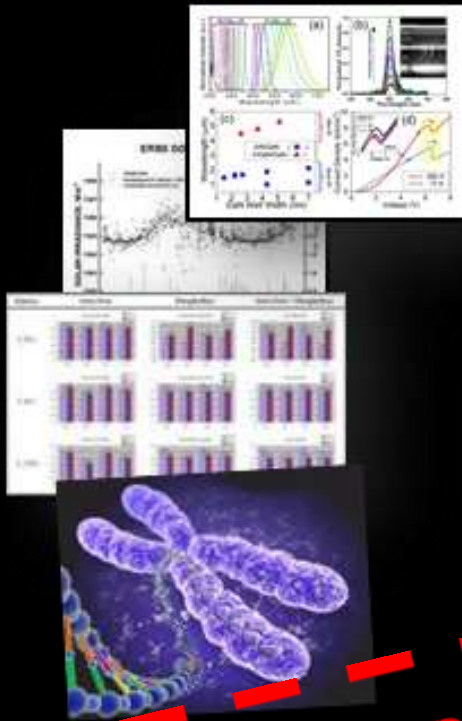
AVALIÁVEL

Natureza, evidências, confiabilidade, fonte competente, financiador, objetivos

4

UTILIZÁVEL

Deve ser capaz de ser reusado em diferentes contextos e, minimamente, por outros cientistas.



INTEROPERÁVEL!

PRINCÍPIO FAIR

Open
data
is about
MORE
THAN
DISCLOSURE
it must be
Fair

- Findable
- Accessible
- Interoperable
- Reusable

ENCONTRÁVEL:

Fácil de achar por humanos e computadores por meio de metadados que facilitem a busca por datasets específicos.

ACESSÍVEL:

Armazenado por longo prazo de forma que ele pode ser facilmente acessado e/ou baixado com licenças e condições de acesso bem definidas (acesso aberto quando possível)

INTEROPERÁVEL

Pronto para combinar com outros dados por seres humanos ou por computadores

REUSÁVEL

Pronto para ser usado para pesquisas futuras, e para ser processado usando métodos computacionais.



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

DADO DE PESQUISA MANEIRO!



**ARQUIVADO
PRESERVADO**

**LOCALIZADO
RECUPERADO
ACESSADO**

**INTERPRETADO
CONTEXTUALIZADO
AVALIADO
PROVENIÊNCIA**

REUSADO

**COMPARTILHADO
ON-LINE**

**ANOTADO
ATIVA COLABORAÇÃO**

INTEROPERÁVEL

**LINKADO COM
PUBLICAÇÃO**

LICENÇA APROPRIADA

CONSIDERA PRIVACIDADE/ÉTICA

**IDENTIFICADO
CITADO
VISÍVEL**

GESTÃO DE DADOS DE PESQUISA



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



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

GESTÃO/CURADORIA



GESTÃO DE DADOS DE PESQUISA

CONJUNTO DE ATIVIDADES GERENCIAIS E TECNOLÓGICAS, APOIADAS POR POLÍTICAS GERAIS E ESPECÍFICAS DESTINADAS A GARANTIR:

- ARQUIVAMENTO**
- CURADORIA, PRESERVAÇÃO E OFERTA DE ACESSO**
- CONTINUO AOS DADOS DE PESQUISA**



PÚBLICO-ALVO

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.

PLATAFORMAS DE GESTÃO DE DADOS DE PESQUISA



VISIBILIDADE



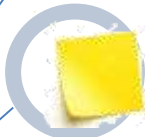
disponibilidade *on-line*
descoberta
acesso



COMPARTILHAMENTO/REUSO/INTERAÇÃO



CRÉDITO AO AUTOR



MEMÓRIA CIENTÍFICA | TRANSPARÊNCIA



CURADORIA DIGITAL



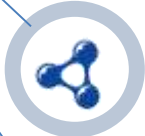
Preservação
Arquivamento
Anotação



SEGURANÇA DOS DADOS



SERVIÇOS INOVADORES



INTEROPERABILIDADE | REDE DE REPOSITÓRIOS



REVISÃO/VALIDAÇÃO/REPRODUTIBILIDADE



INDICADOR DE QUALIDADE E PRODUTIVIDADE

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

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 **perdido** 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.



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



CADA MECANISMO DE COLETA E ARMAZENAMENTO DE DADOS COLOCA DESAFIOS ESPECÍFICOS SOBRE:
ACESSOS NÃO AUTORIZADOS, PERDA OU DANOS ACIDENTAIS, ROUBO, HACKEAMENTO E ELIMINAÇÃO



**PESQUISAS
CORRENTES**

**SENSÍVEIS/PESSOAIS
COMPARTILHADOS
VERSIONADOS
ATIVOS
ACESSO REMOTO
VOLUME
FORMATOS
AMOSTRAS FÍSICAS
DOCUMENTAÇÃO**

O COORDENADOR DA PESQUISA DEVE INFORMAR ADEQUADAMENTE A INSTITUIÇÃO A NATUREZA DOS DADOS DE PESQUISA COLETADOS:
**MECANISMOS DE COLETA, PROCESSAMENTO E ANÁLISE;
QUESTÕES LEGAIS E ÉTICAS,
MEDIDAS DE SEGURANÇA**

**POLÍTICA DE
PROTEÇÃO &
SEGURANÇA
DE DADOS**

**Criptografia
Anonimização
Consentimento**

**Checksum
Assinatura digital
Backup
Segurança física**

**Sistema de *Storage*
Nuvem institucional
Nuvem pública
Servidor
Dispositivos pessoais**

**Anti-virus
Atualização de soft/mídias
firewall**

**Amostras físicas
Cadernos de lab**

GESTÃO DE DADOS DE PESQUISA



PÓS-PUBLICAÇÃO → PRÉ-PUBLICAÇÃO

As bibliotecas de pesquisa tem que capturar dados em diferentes estágios da geração e processamento dos dados de pesquisa.

ALINHAMENTO COM OS PROCESSOS DE PESQUISA



WORKFLOW; INSTRUMENTOS;
MODELOS; FERRAMENTAS;
CÓDIGOS; CADERNO DE
PESQUISA...

PLANO DE DADOS DE PESQUISA

O planejamento da gestão de dados se torna parte do processo de investigação científica

CAPTURA DE DADOS

Inserção de coleções de dados – brutos ou derivados - provenientes de experimentos, simulações, observações, questionários, levantamentos etc. Os dados podem ser submetidos pelos próprios autores ou por equipes especializadas vinculadas ao serviço.

CATALOGAÇÃO DAS COLEÇÕES DE DADOS

Descrição, atribuição de metadados e inclusão de documentação que assegurem que os dados possam ser acessados e interpretados no tempo e no espaço.

ARQUIVAMENTO E PRESERVAÇÃO

Arquivamento seguro que garanta a gestão de curto e longo prazo das coleções de dados orientadas por um plano/política de preservação digital

INTEROPERABILIDADE

Intercâmbio e compartilhamento e *linkage* com outros repositórios de dados e outros sistemas de informação (repositórios institucionais, bibliotecas digitais de publicações acadêmicas, editoras científicas)

RECUPERAÇÃO, ACESSO E REUSO

Interface web para a descoberta, acesso e download de coleções de dados relevantes para o usuário ou para aplicações computacionais, como visualização e mapeamento, que podem prover serviços a partir dessas coleções; vinculado a uma política de acesso estabelecida pela instituição que inclui: tempo de embargo, direito de acesso, pagamentos, restrições sobre determinadas coleções, acesso somente aos metadados, registros de usuários e termos de uso dos dados.





GESTÃO DE DADOS DE PESQUISA

- **Avaliação/seleções** de coleção de dados (appraisal)
- Verificação do **enquadramento no escopo** do repositório;
- Verificação dos **formatos de arquivos** aceitáveis para submissão;
- Verificação dos **direitos associados às coleções** (copyright e licenças);
- Verificação de **dados sensíveis** (dados não anonimizados, confidenciais, pessoais);
- Verificação do **volume e quantidade** de arquivos;
- Verificação dos **metadados gerais e disciplinares** que acompanham os dados;
- **Normalização** para elenco de formatos padronizados aceitos para arquivamento e disseminação;
- **Controle de qualidade** dos dados;
- Definição de **tempo de embargo**.





GESTÃO DE DADOS DE PESQUISA



- **Atribuição de metadados descritivos, estruturais, administrativos, técnicos** (que inclui os relativos às dependências técnicas dos objetos digitais)
- **Atribuição de metadados de preservação**, que assegurem a proveniência, autenticidade e integridade dos dados ao longo do tempo;
- Uso de **taxonomias especializadas** e disciplinares;
- **Atribuição de identificador persistente** (DOI, Handles, UNF, URN, etc.) que permita que os dados possam ser localizados de forma persistente e citados como as publicações acadêmicas;
- **Identificação do autor** (ORCID ID, Scopus Author ID, ResearcherID etc.)
- **Inclusão de documentação sobre os dados**, incluindo descrição do projeto, dos arquivos e dos parâmetros; cadernos de laboratório e de campo, protocolos de pesquisa ou metodologia etc.;
- **Vinculação (por links)** a publicações e a dados relacionados internos e externos ao repositório



GESTÃO DE DADOS DE PESQUISA



- **Armazenamento** em sistemas seguros;
- **Gestão da preservação de curto prazo** (backups, backups redundantes *offsite*; checagem de integridade, armazenamento seguro, criptografia, compressão);
- **Gestão de longo prazo** (migração, emulação, reformatação para formatos padronizados, aplicação de normas pertinentes (OAIS, TRAC), informação de fixidade voltada para validar a autenticidade e integridade de um objeto digital (*checksums*, assinatura digital);
- Implementação de **trilhas de auditoria**;
- e **compartilhamento e linkage** com outros repositórios de dados e outros sistemas de informação (repositórios institucionais, bibliotecas digitais de publicações acadêmicas, editoras científicas)



GESTÃO DE DADOS DE PESQUISA

INTEROPERABILIDADE



- Intercâmbio
- Disponibilização de metadados segundo o **protocolo OAI-PMH**;
- Agregação para **formação de publicações ampliadas** segundo o padrão OAI-ORE;
- Uso dos padrões, **web service**, **linked data** e outros
- Empacotamento de metadados para intercâmbio segundo o **padrão METS**.
- **API**



GESTÃO DE DADOS DE PESQUISA

RECUPERAÇÃO, ACESSO E REUSO



- Disponibilização de interfaces web para **recuperação, acesso e *download***;
- Oferta de **aplicações e serviços** sobre as coleções para seres humanos e aplicações

SERVIÇOS DE GESTÃO DE DADOS DE PESQUISA

GESTÃO DE DADOS & PLANO DE COMPARTILHAMENTO

Apoio aos pesquisadores para preparar os seus **planos de gestão de dados** e de compartilhamento.

ASPECTOS LEGAIS E ÉTICOS

Este serviço inclui alertar os pesquisadores sobre **implicações legais e éticas** da criação, compartilhamento e utilização de dados.

MELHORES FORMATOS E MELHORES PRÁTICAS

Apoio aos pesquisadores para decidir quais são os melhores formatos e prática para **produzir e documentar dados específicos**. Este serviço pode também incluir a prestação de apoio para o **projeto de banco de dados**.

ARMAZENAMENTO SEGURO

inclui infraestrutura que permite **armazenar dados de pesquisa**, que ofereçam serviços de *backup* e controle de versão entre outras coisas.

METADADOS

Ferramentas e suporte que permitam o pesquisador **descrever seus dados** a partir do momento da criação.

DESCOBERTA E ACESSO

Serviço de apoio e ferramentas para ajudar os pesquisadores a **localizar e acessar dados de pesquisa**.

SERVIÇOS DE GESTÃO DE DADOS DE PESQUISA

COMPUTAÇÃO, ANÁLISE E VISUALIZAÇÃO

Ferramentas de software e outros recursos de computação que permitem a **análise e visualização de dados de pesquisa**, bem como a capacitação necessária para dotar os pesquisadores das competências adequadas para uso das ferramentas.

COMPARTILHAMENTO RESTRITO

Infraestrutura técnica para **compartilhar** dados de pesquisa com **indivíduos ou grupos selecionados**.

LIMPEZA DOS DADOS

Apoio para **limpar e preparar os dados para o padrão exigido para publicação**. Este serviço deve incluir ajuda para **anonimização dos dados**.

PUBLICAÇÃO

Infraestrutura que permite aos pesquisadores **publicar dados documentados e ligá-los por hyperlink a artigos e a outros materiais localizados em outros repositórios**.

AVALIAÇÃO

Um dos principais desafios com dados de pesquisa é decidir quais os dados que **precisam ser mantido e por quanto tempo**.

PRESERVAÇÃO

Este serviço seria responsável por **cuidar dos dados no longo prazo**, aplicando as medidas necessárias para que os dados são acessíveis através do tempo.

ADICIONAR VALOR

Uma vez que os dados são armazenados com os metadados associados, o valor pode ser adicionado através da organização de dados semelhantes em grupos, promovendo-o, ligando-o a outros materiais ou permitir anotações.

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,



PAPÉIS NA GESTÃO DE DADOS DE PESQUISA

É NECESSÁRIO CONHECIMENTO DISCIPLINAR PARA A PRÁTICA DA CURADORIA DE DADOS



PROVER SERVIÇOS
AGREGAR VALOR
GERENCIAR
TOMAR DECISÕES
PRESERVAR



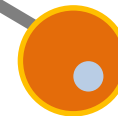
CONHECIMENTO SOBRE

- . domínio disciplinar
- . problemas
- . processos/fluxos



CAMPO CIENTÍFICO

Terminologia
Metodologias
Tipos de dados/formatos
Padrões de aceitação
Normas da comunidade

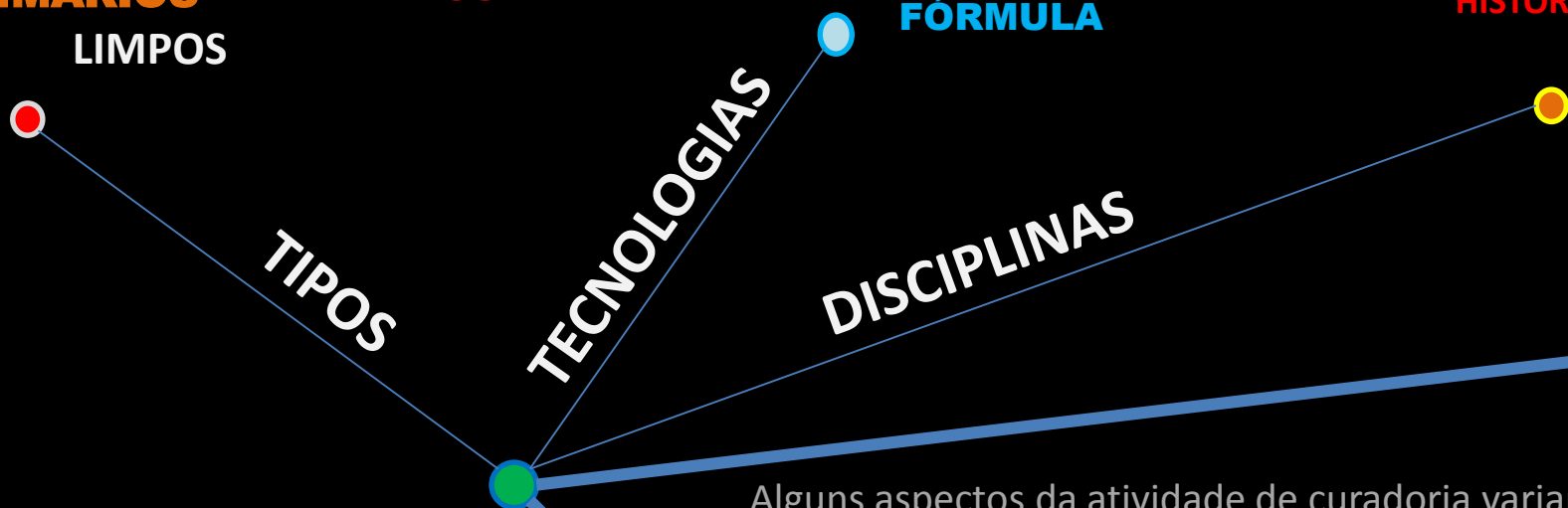


CAMPO COMERCIAL

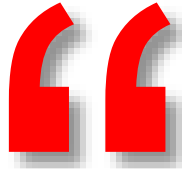
Cenários de competidores
Arcabouço regulatório
Nomeclatura da área



DINÂMICOS
ESTÁVEIS **EXPERIMENTAIS** **TEXTO** **ASTRONOMIA**
OBSERVACIONAIS **SIMULAÇÃO** **FÍSICA** **MEDICINA**
ÊFEMEROS **TERCIÁRIOS** **VIDEO** **SOFTWARE** **ECOLOGIA**
COMPUTACIONAIS **SUJOS** **NÚMEROS** **GRÁFICOS** **CIÊNCIAS SOCIAIS**
PRIMÁRIOS **DERIVADOS** **FÓRMULA** **HISTÓRIA**
LIMPOS



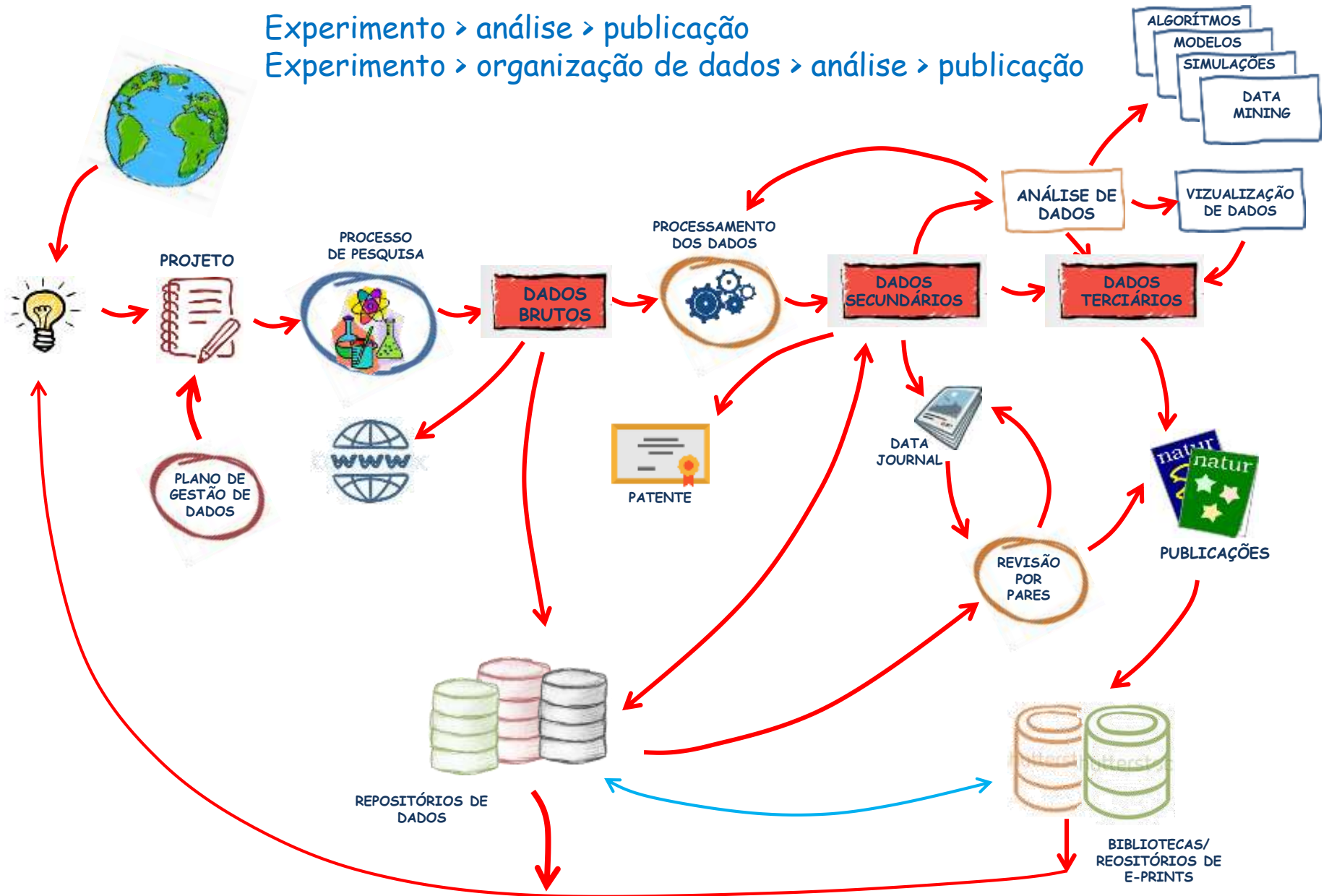
Alguns aspectos da atividade de curadoria variam amplamente de acordo com o **TIPO DE DADOS**, **TECNOLOGIAS SUBJACENTES AOS DADOS**, e, sobretudo, com o **DOMÍNIO DISCIPLINAR ESPECÍFICO**.



O sucesso dos novos serviços de informação para a pesquisa está relacionado à sua capacidade de dar apoio às **práticas e culturas** das comunidades científicas da instituição.

Experimento > análise > publicação

Experimento > organização de dados > análise > publicação



PROJETO DE PESQUISA

CAPTURA DE DADOS

LIMPEZA DOS DADOS

ANÁLISES E RESULTADOS

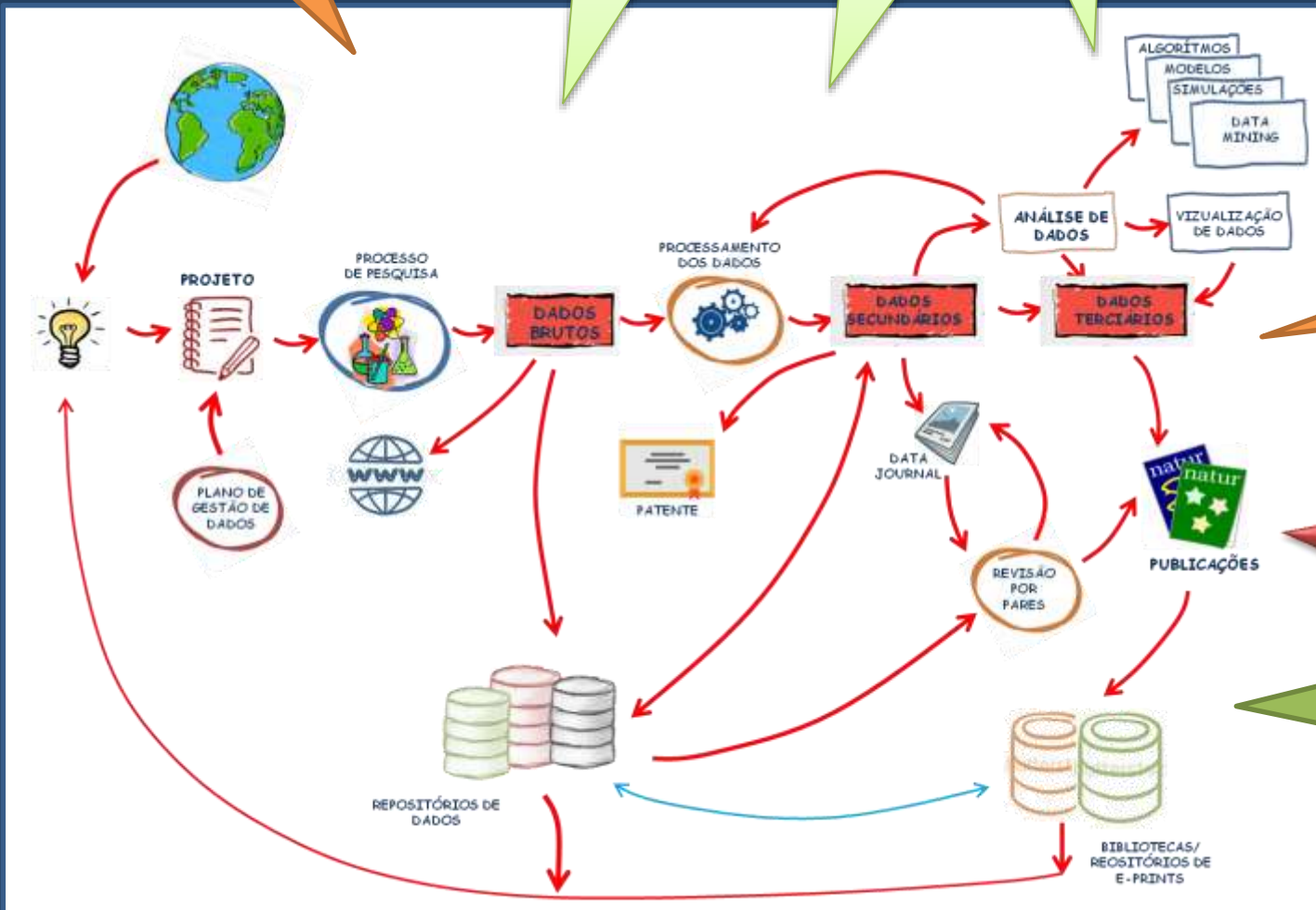
FORMULAÇÃO DE HIPÓTESES !

PROCESSOS COMPUTACIONAIS

CATALOGAÇÃO

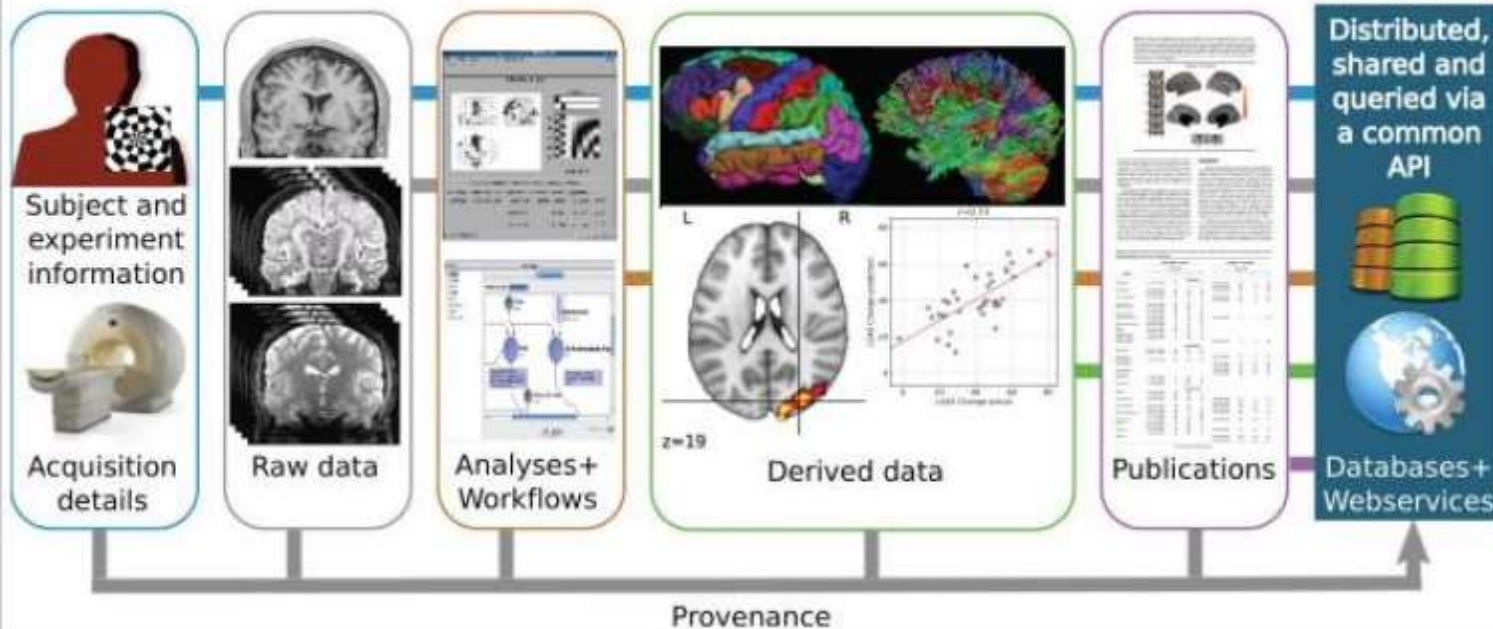
INTEGRAÇÃO COM O SISTEMA DE PUBLICAÇÃO

CURADORIA ARQUIVAMENTO PRESERVAÇÃO

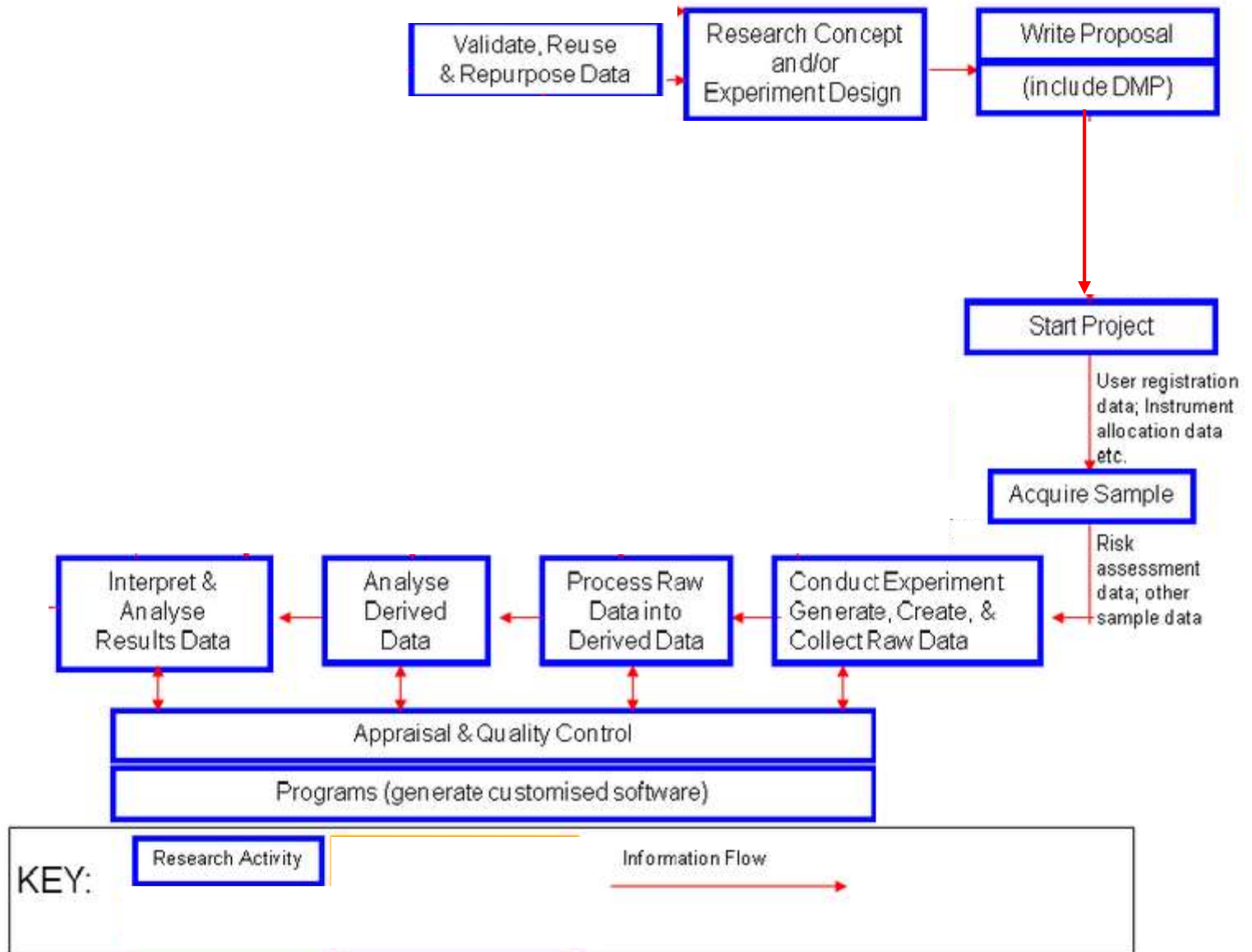


POLÍTICAS – SUSTENTABILIDADE – CONFORMIDADE LEGAL E ÉTICA

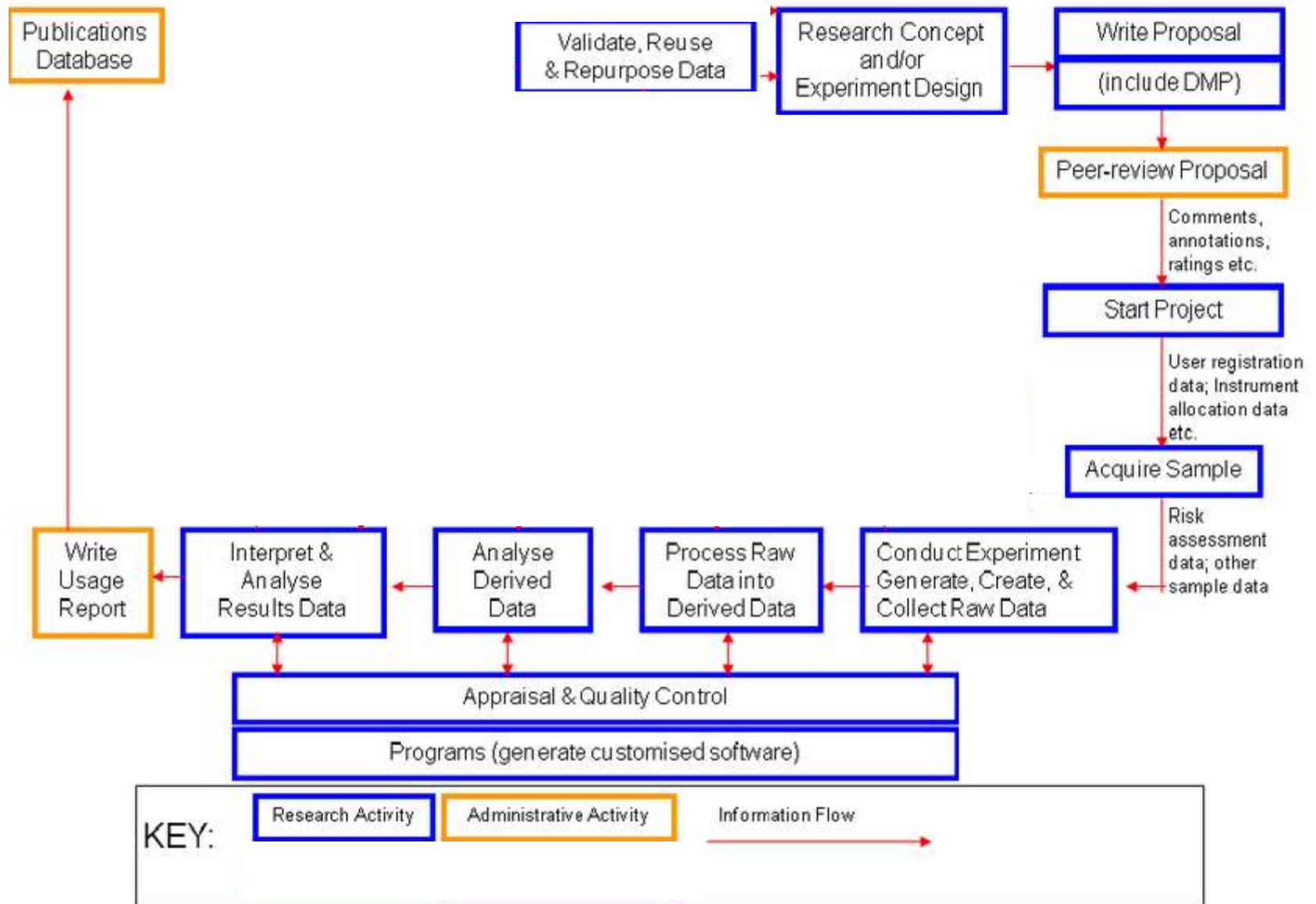
Stages of Electronic Data Capture



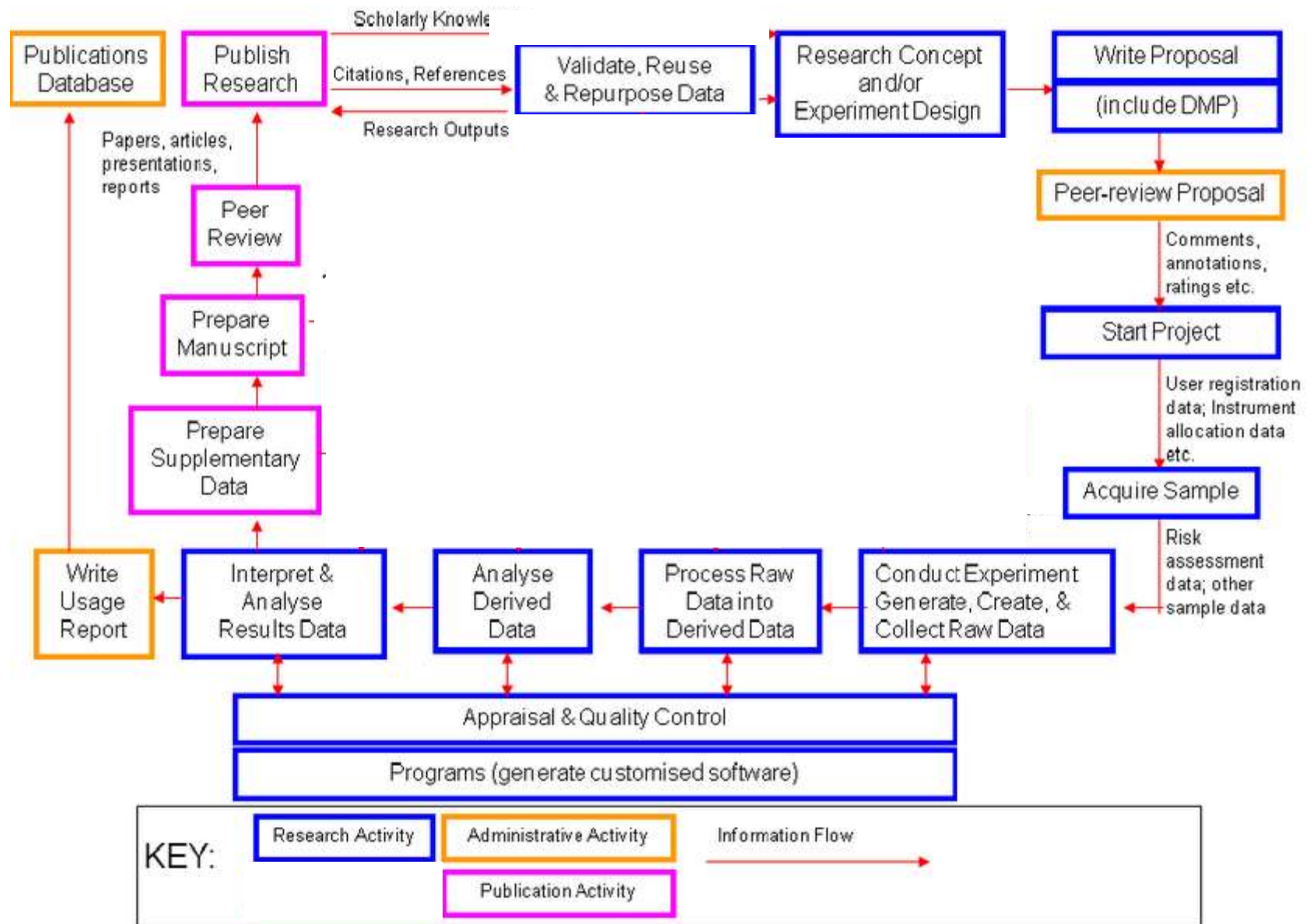
An Idealised Scientific Research Activity Lifecycle Model



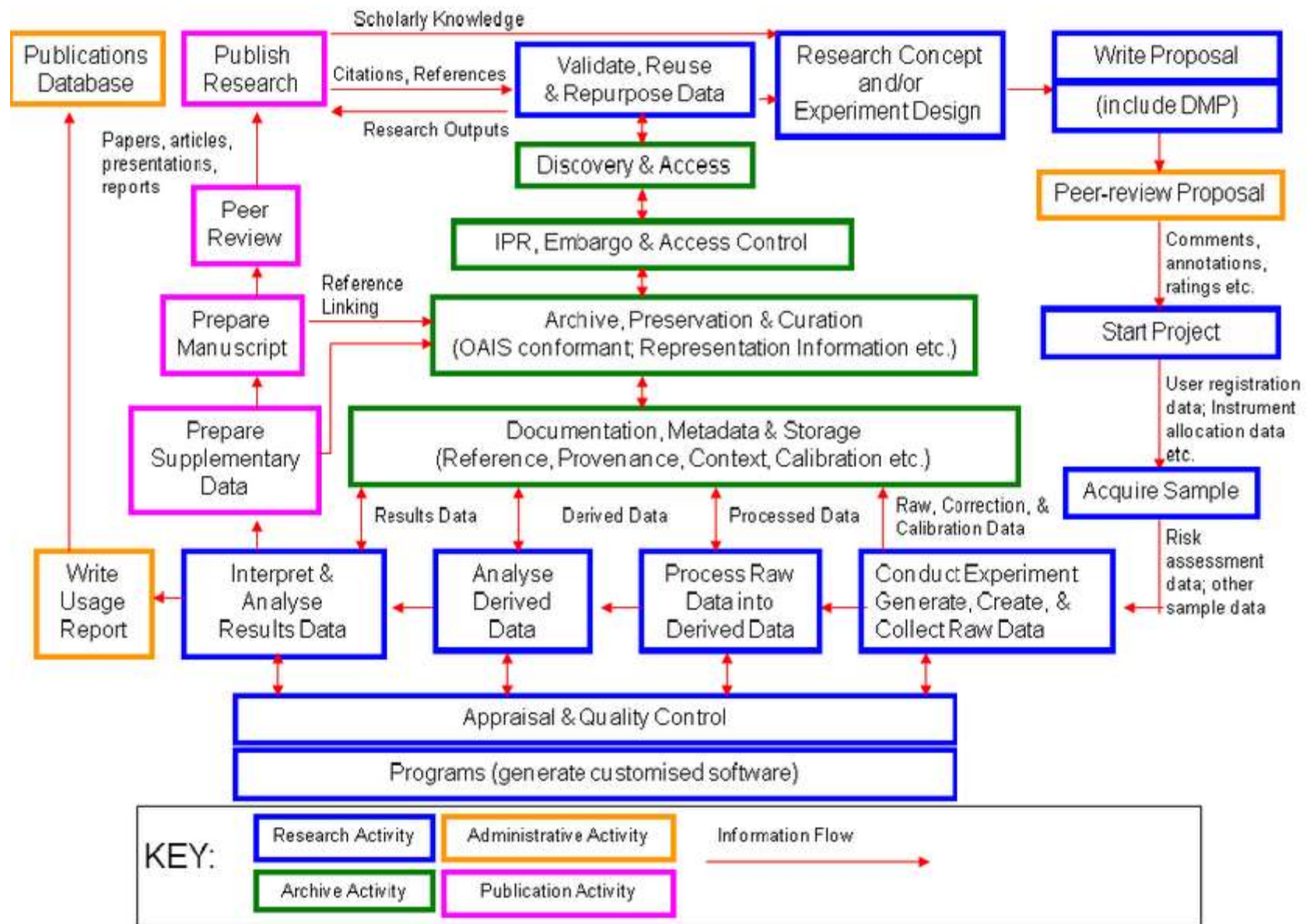
An Idealised Scientific Research Activity Lifecycle Model



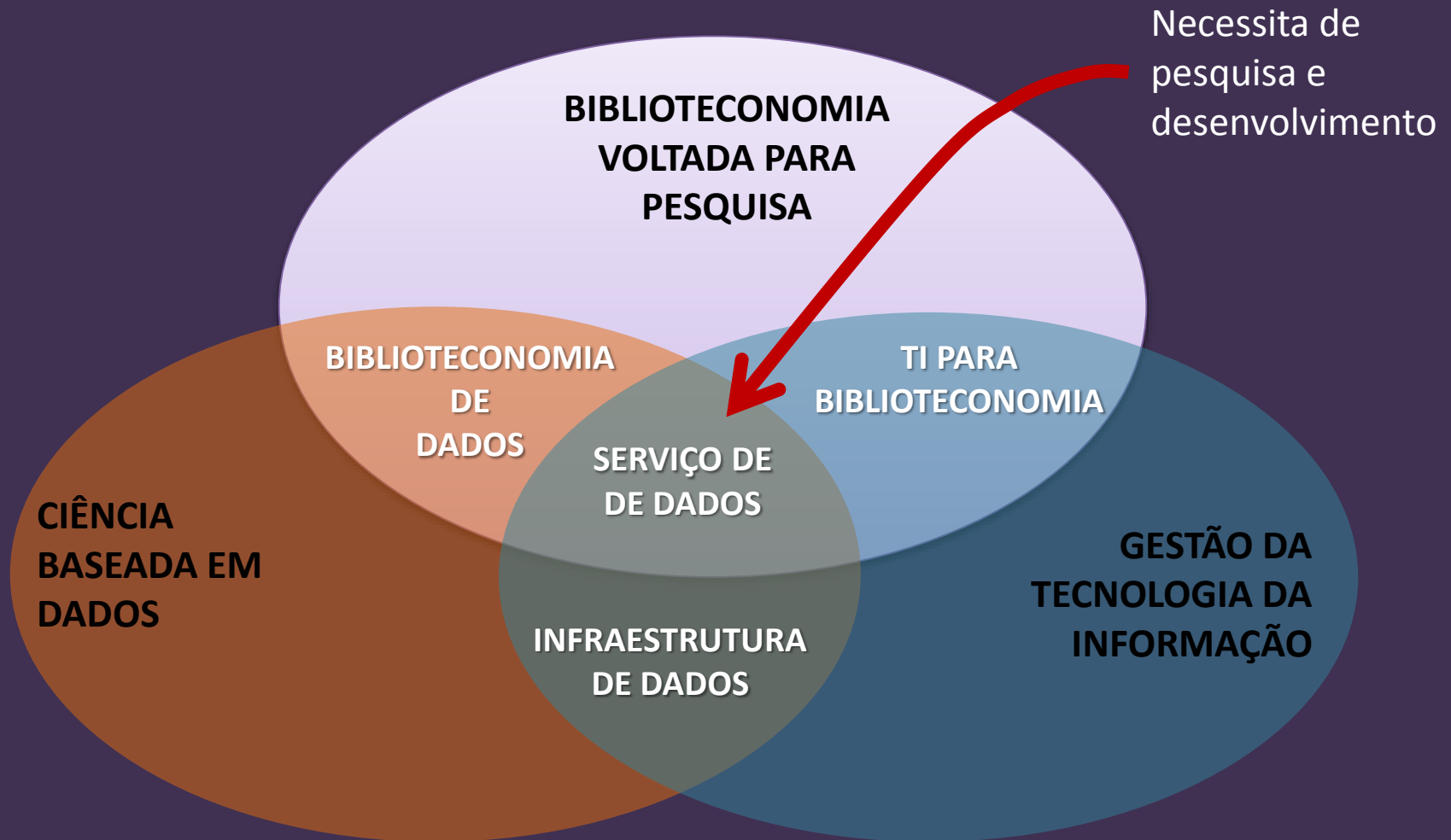
An Idealised Scientific Research Activity Lifecycle Model

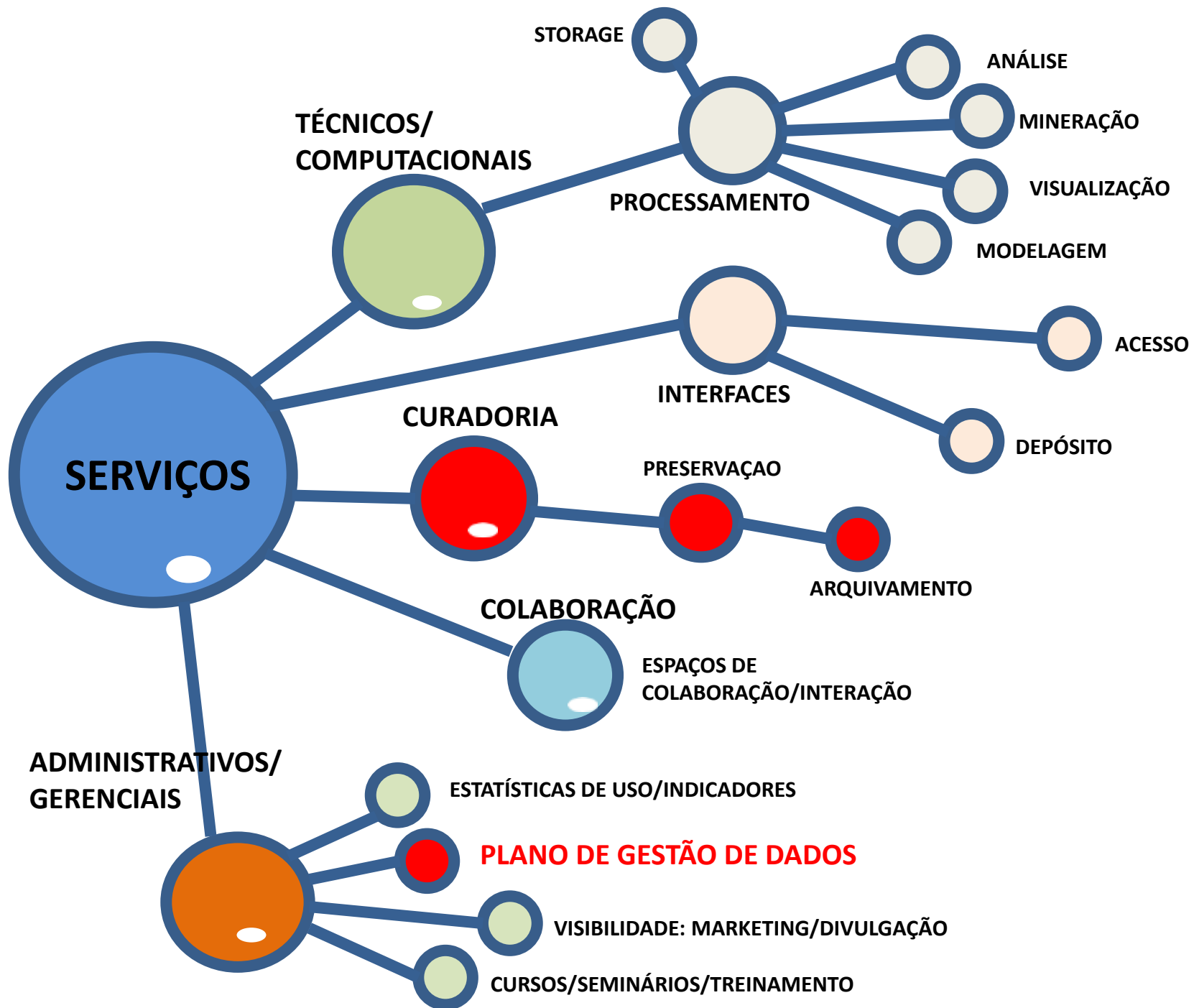


An Idealised Scientific Research Activity Lifecycle Model



INFRAESTRUTURA DE DADOS E A BIBLIOTECA DE PESQUISA





**CURADORIA
DE DADOS
DE PESQUISA**

CURADORIA: A PRÓXIMA FRONTEIRA



CURADORIA

O COLETIVO
CARANDAÍ 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.

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).

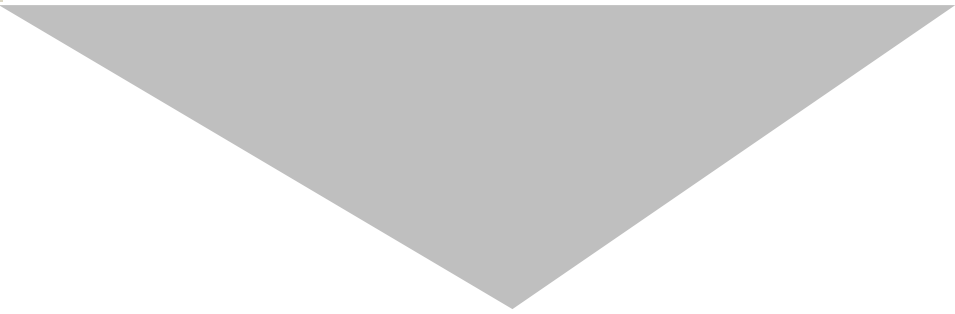
OBJETOS FÍSICOS

- 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



CURADORIA DIGITAL

imensas e sempre crescentes quantidades de conteúdos digitais

objetos complexos e heterogêneos, que dependem de tecnologias específicas e pouco duradouras,

grande diversidade de contextos organizacionais em que a curadoria de conteúdos digitais ocorre;

uma audiência que pode ser indefinida e localizada no futuro

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).

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 ESTÁ 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
LINKS

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

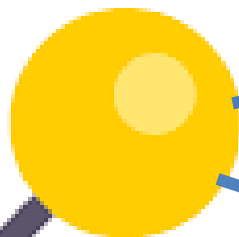
**VOLTADO PARA
UM PÚBLICO-ALVO**

**ALINHADO COM O
FLUXO DE PESQUISA**



GESTÃO ATIVA

Distingue curadoria de simplesmente coletar e armazenar dados



ATIVIDADES GERENCIAIS

Políticas | avaliação de risco | identificar acervos |
regulação | recursos | pessoal



ATIVIDADES TÉCNICAS

hardware e software | repositórios | metadados |
disseminação

ADICIONAR VALOR

Medidas par aumentar o valor dos dados para uso atual e futuro



USOS

ATUAL

Diversidade de usos e metodologias



FUTURO

Atualização de tecnologias, metodologias, software



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

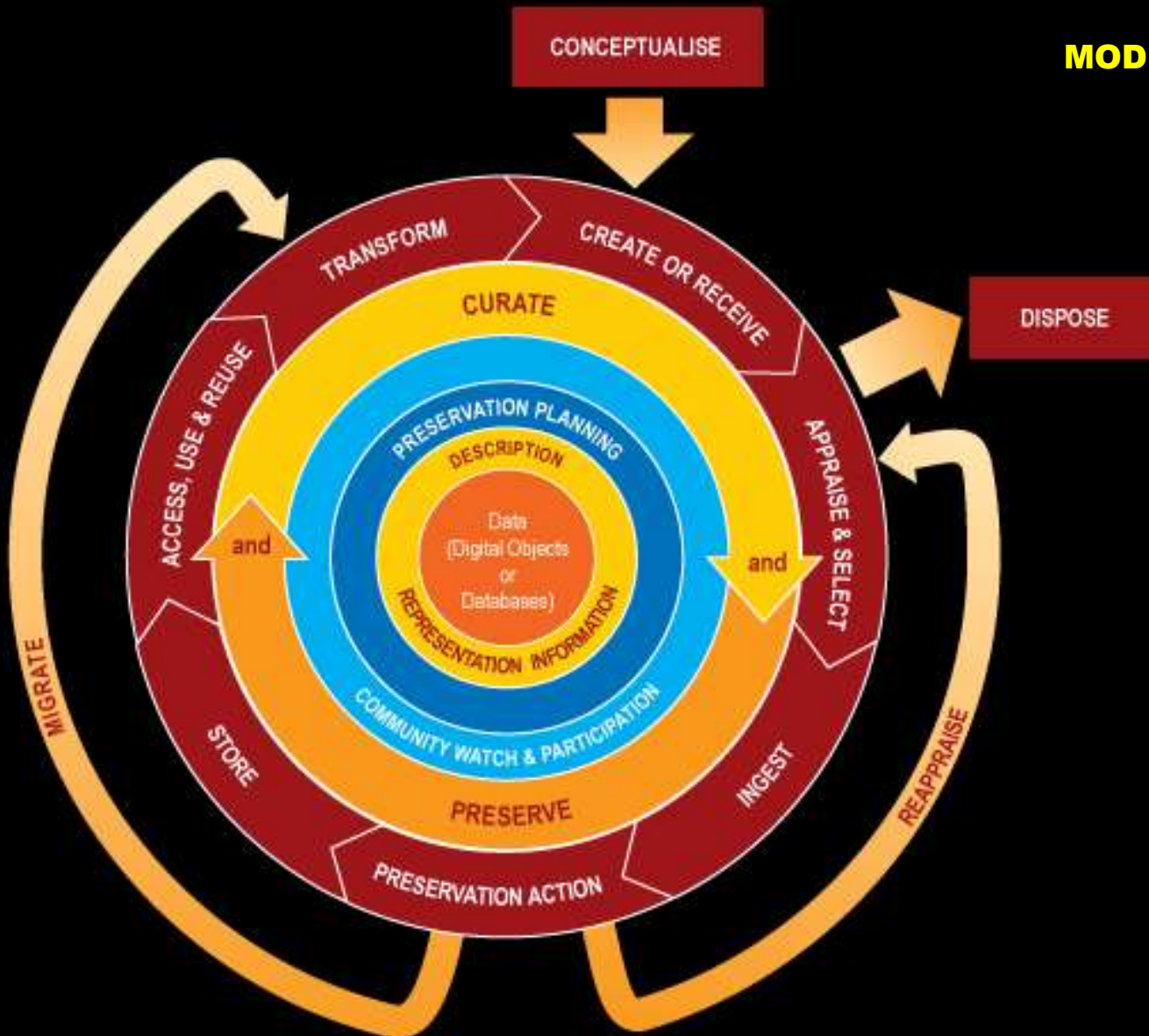
ATIVIDADE INCLUÍDA NA CURADORIA ATRAVÉS DA QUAL ITENS ESPECÍFICOS DE DADOS SÃO MANTIDOS AO LONGO DO TEMPO DE FORMA QUE ELES POSSAM SER ACESSADOS E COMPREENDIDOS NO FUTURO A DESPEITO DAS MUDANÇAS TECNOLÓGICAS



D | C | C

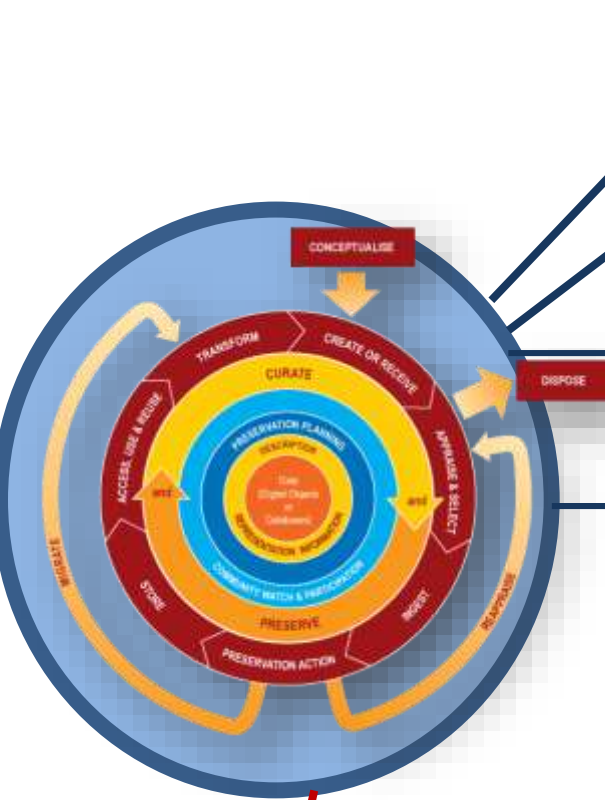
because good research needs good data

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**

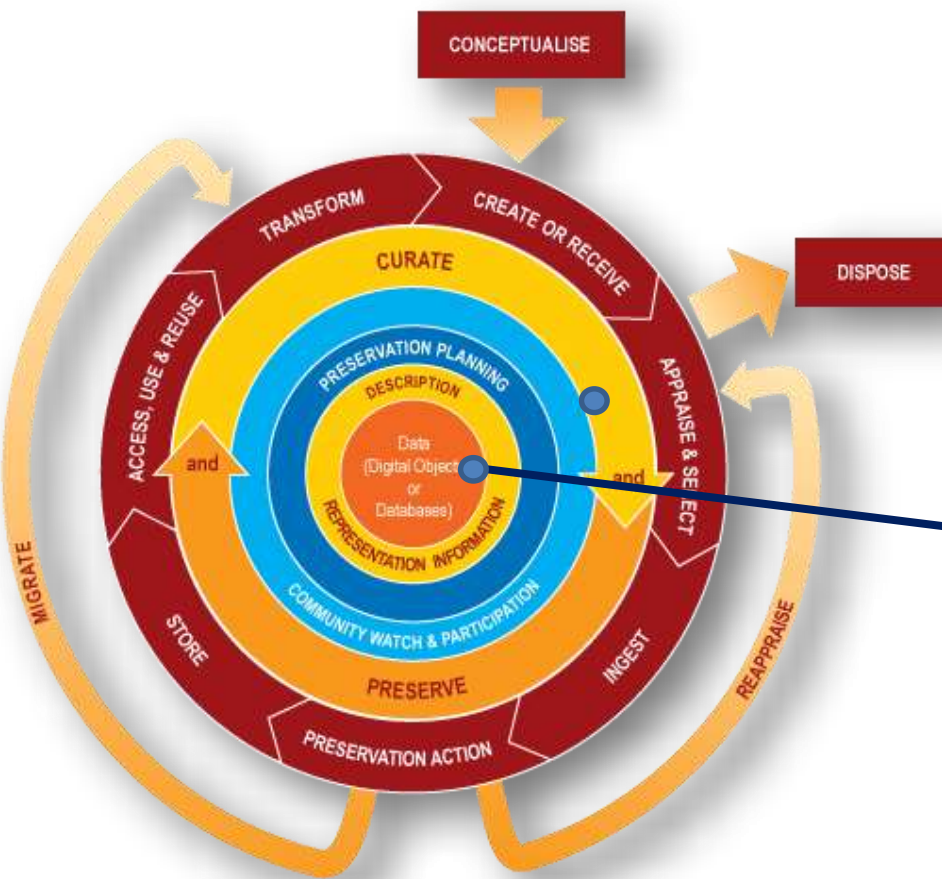
Extensível para permitir que atividades de curadoria e preservação possam ser planejadas em diferentes níveis de **granularidade**.

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.



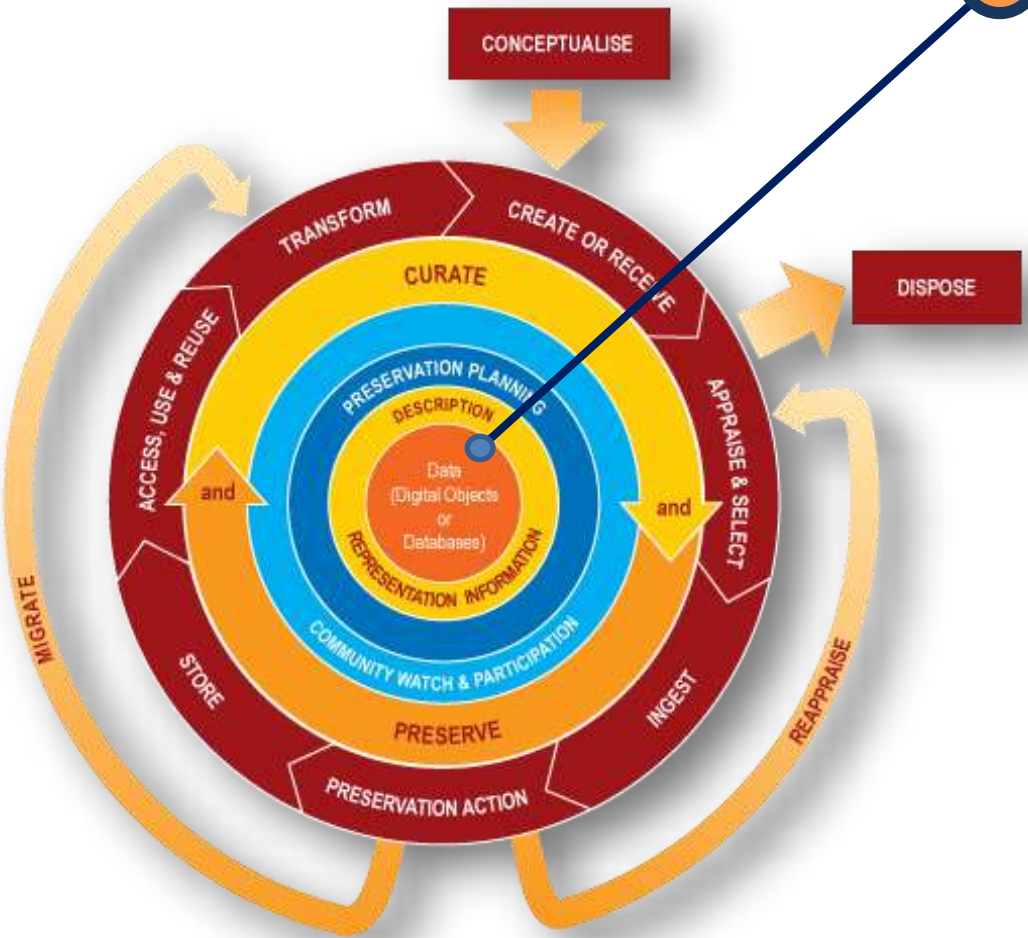
DADO

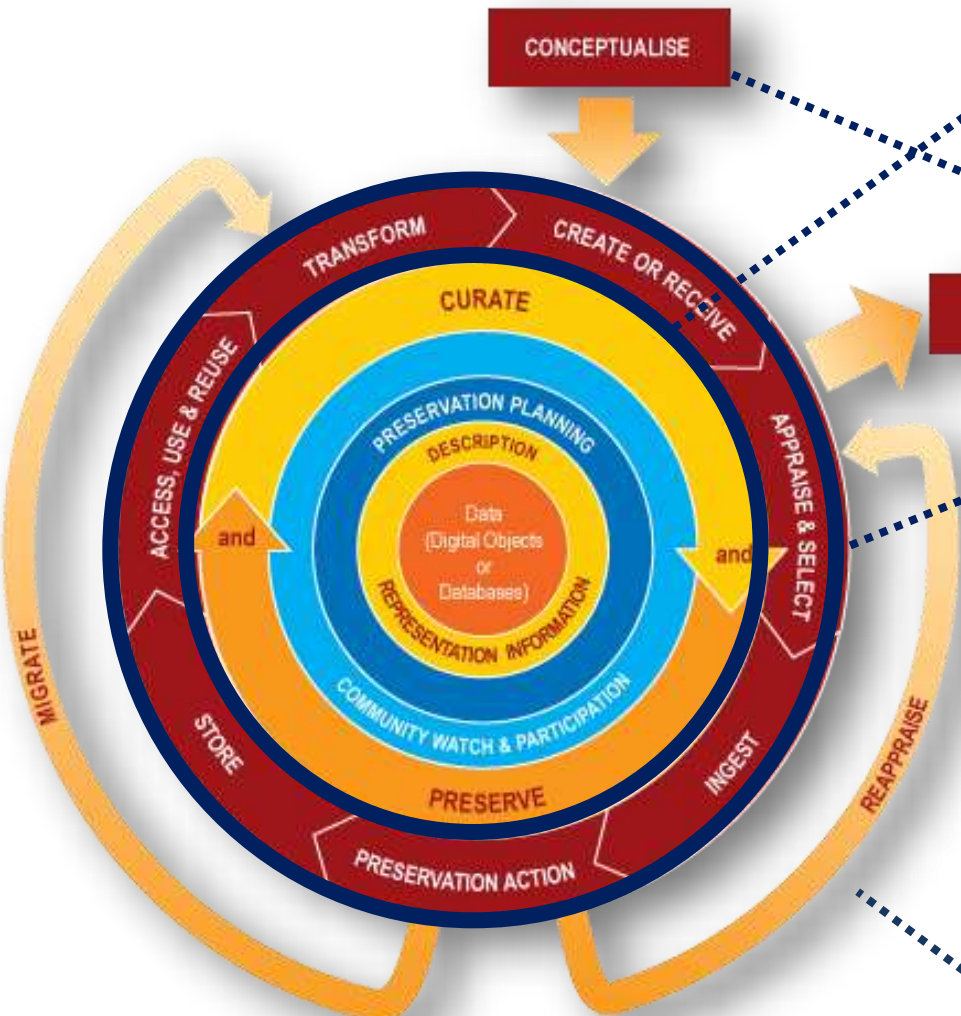
Objetos digitais { simples
 Bases de dados { complexos

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

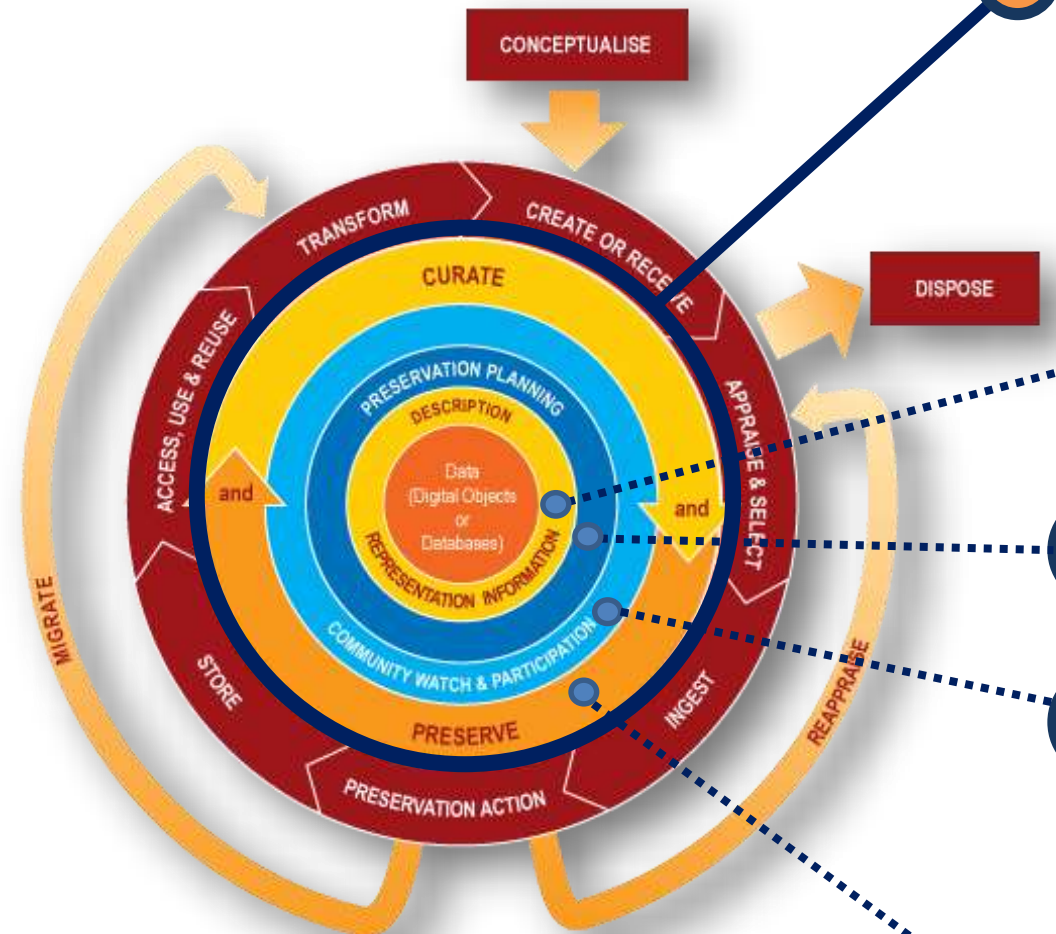
AÇÕES SEQUENCIAIS

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

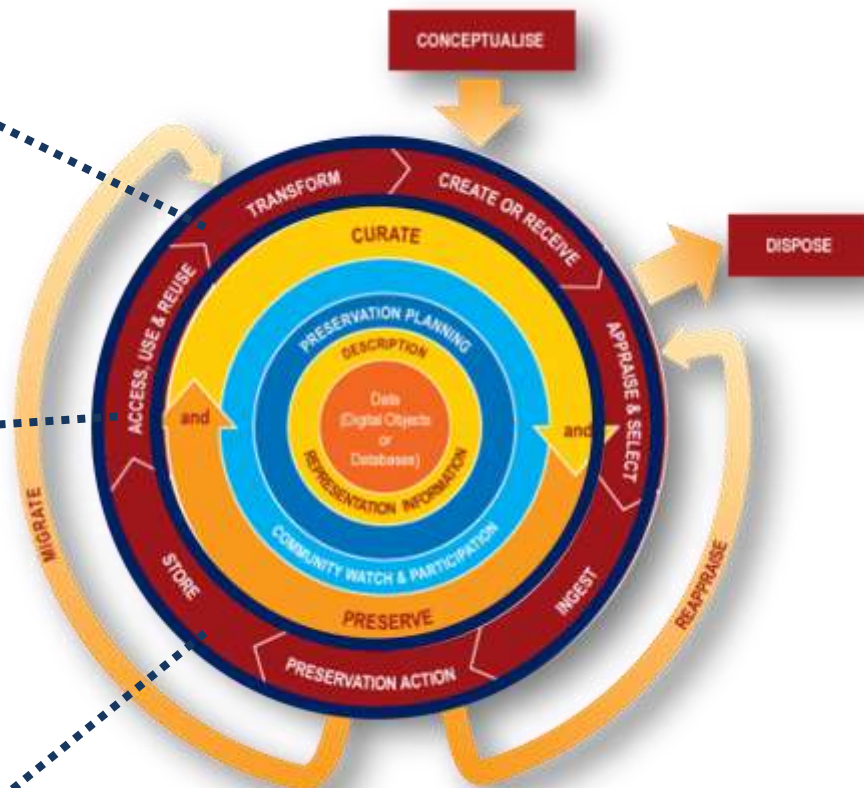
criar novos dados a partir do original como, por exemplo, a migração para um formato diferente ou a geração de resultados derivados por seleção ou consulta.

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

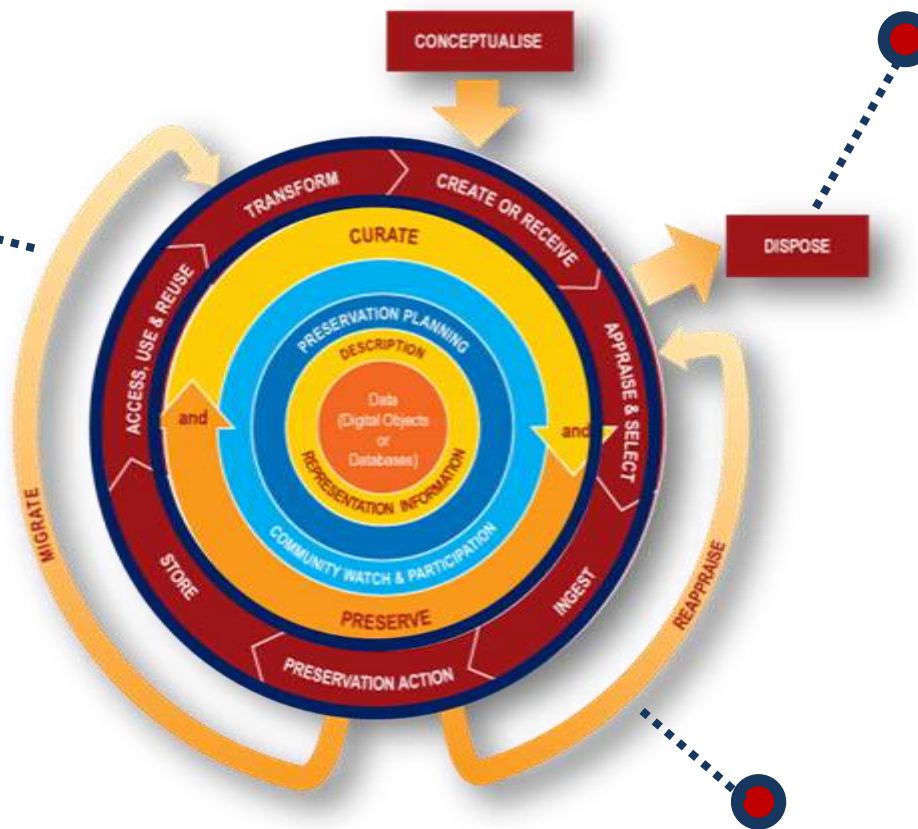
dos dados de forma segura e aderente aos padrões relevantes,



AÇÕES OCASIONAIS

Ações aplicadas eventualmente, quando necessário

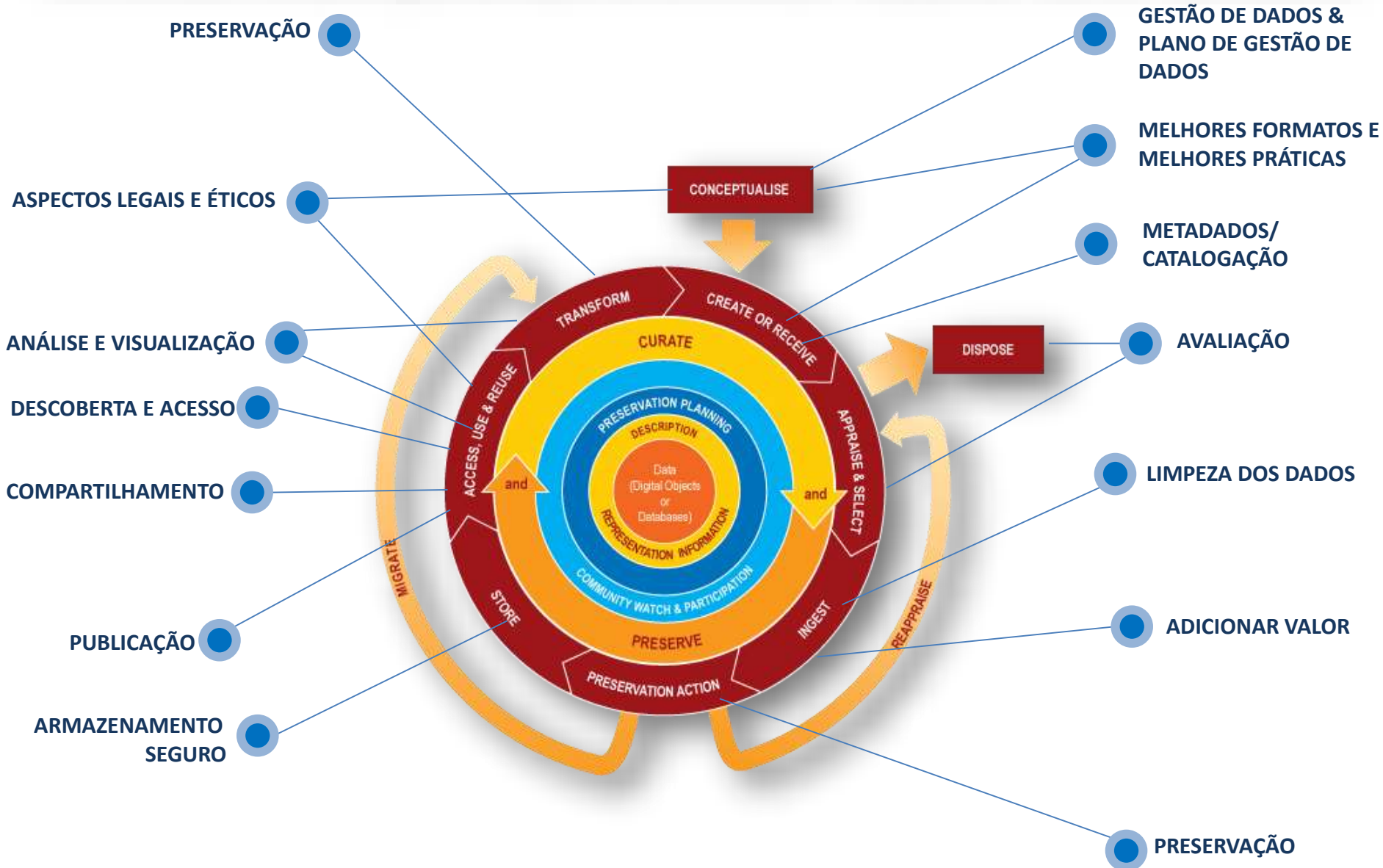
MIGRAÇÃO
do dado para um formato diferente para compatibilizá-lo com o ambiente de armazenamento, ou para garantir a imunidade à obsolescência de *hardware* e *software*.



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.

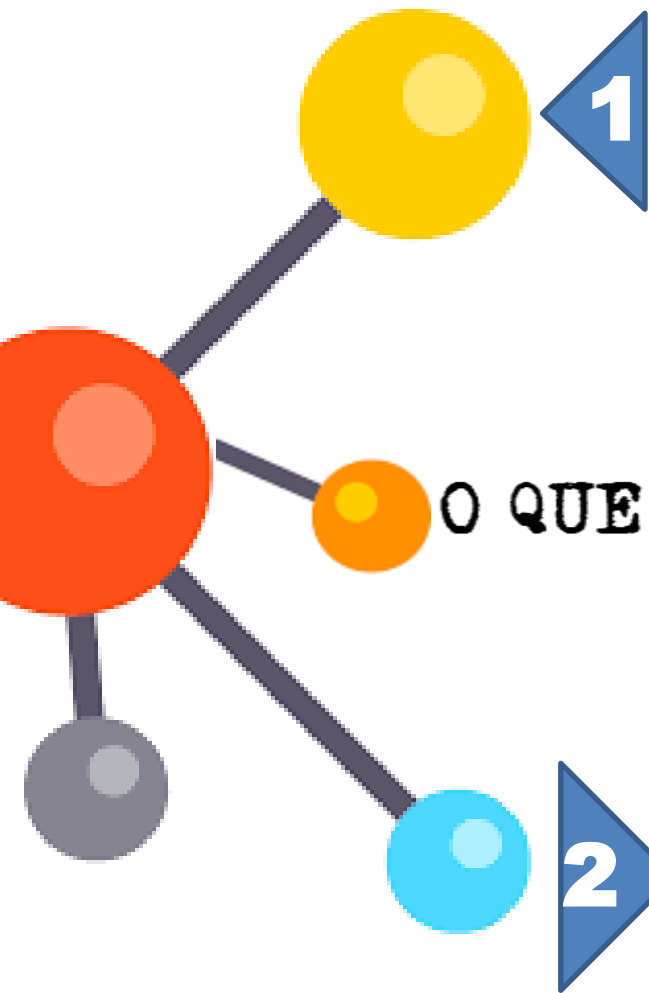
REAVALIAÇÃO
dos dados retornados por falhas nos procedimentos de validação para nova avaliação e possível seleção.

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



REPOSITÓRIO DIGITAL DE DADOS DE PESQUISA

A primeira exigência para a curadoria é o dado estar disponível em um repositório



DEPOSITAR & COMPARTILHAR

infraestruturas que assegurem o máximo de **confiabilidade, estabilidade e acessibilidade** e que facilitem o trabalho de **arquivamento, compartilhamento** e **reconhecimento de autoria** para os seus dados

O QUE PRECISAM OS PESQUISADORES?

DESCOBRIR E ACESSAR

precisam **encontrar coleções de dados** de pesquisa, saber como acessá-las e sob que condições podem reutilizar esses dados e assim dar prosseguimento às suas pesquisas **confiando na autenticidade e proveniência dos dados** coletados ou gerados por outros pesquisadores.

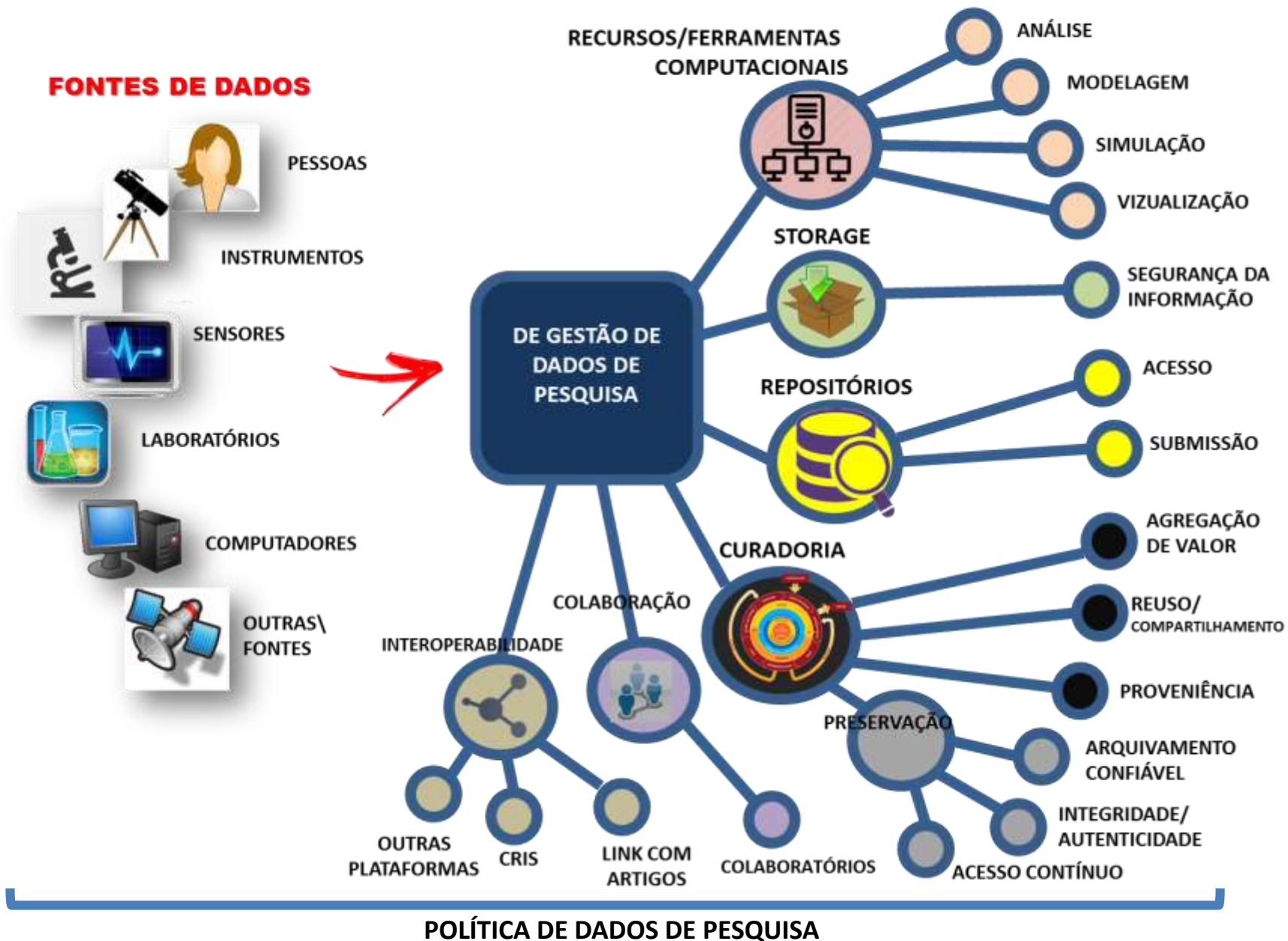
REPOSITÓRIOS DE DADOS DE PESQUISA



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)



CIBERINFRAESTRUTURA DE DADOS DE PESQUISA



○ REPOSITÓRIO INSTITUCIONAL DE DADOS DE PESQUISA

Extensão dos modelos de dados dos RI's

Links entre os recursos

Multidisciplinares

CARPE DIEM (IEN/CNEN); Edimburgh Data Share

○ REPOSITÓRIO TEMÁTICO DE DADOS DE PESQUISA

Domínios ou tipos de dados específicos

Bio Models Data Base (modelos computacionais na área de biologia)

GenBank

○ REPOSITÓRIO MULTIDISCIPLINAR DE DADOS DE PESQUISA

Dados gerados/coletados por várias áreas

Gestão mais complexa: heterogêneos, licenças, formatos...

Dryad, Figshare, Zenodo e o Dataverse

○ REPOSITÓRIO DE DADOS DE PESQUISA DE UM PROJETO ESPECÍFICO

Coleções de dados resultados de projetos de pesquisa ou resolução de problemas específicos.

Scientific Drilling Database



TIPOS DE PLATAFORMAS DE GESTÃO DE DADOS

As **PLATAFORMAS DISCIPLINARES** se voltam para domínios específicos ou para tipos particulares de dados. Em geral possuem modelos de dados adequados à representação das coleções de dados e oferecem uma **CARTEIRA DE SERVIÇOS** mais orientadas, como curadoria e visualização.

Essas plataformas estão abertas para publicar qualquer tipo de dados, e são especialmente desenvolvidas para dar apoio a publicação de *datasets* produzidas no âmbito da ciência chamada de **“CAUDA LONGA”** – domínios científicos nos quais um grande número de relativamente pequenos laboratórios ou de pesquisadores individuais produzem a maioria dos resultados científicos.

BioModels Database

Advanced

BioModels Database is a repository of computational models of biological processes. Models described from literature are manually curated and enriched with cross-references. All models are provided in the Public Domain. More information about BioModels Database can be found in the [FAQ](#).

Models published in the literature

Browse

Alternative access



[Manually curated](#)
(630 models)



[Non curated](#)
(983 models)



[Gene
Ontology
classification](#)



[Gene
Ontology tree](#)



[Advanced
search](#)

Models automatically generated from pathway resources (Path2Models)

Browse

Alternative access



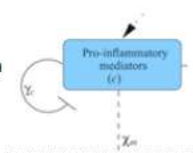
[Metabolic](#) (112,898 models)
[Non-metabolic](#) (27,531 models)



Model of the month

May, 2017

A mathematical model on inflammation resolution: prediction suggests that an effective treatment protocol would be to target macrophage phagocytosis alongside neutrophil apoptosis.



[Access this model of the month.](#)

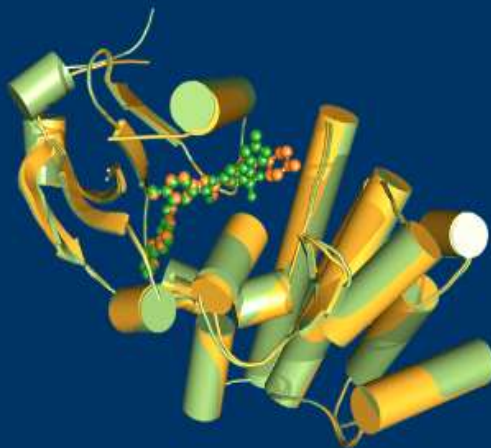
News

- 16 January 2017** [Metabolic network and logical models](#)
Archives of [metabolic network](#) and [logical models](#) are available for download.
- 9 January 2017** [Models on neurodegeneration: review article](#)
Published in [CPT:PSP Mechanistic models on neurodegenerative disease processes](#)
- 9 January 2017** [New Models in BioModels](#)
BioModels now provides 1603 literature-based models.

Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Learn more about PDB **HISTORY** and **FUTURE**.



Validate Structure

or View validation reports



Deposit Structure

All Deposition Resources



Download Archive

wwPDB Members

wwPDB data centers serve as deposition, annotation, and distribution sites of the PDB archive. Each site offers tools for searching, visualizing, and analyzing PDB data.

BMRB

- › **Biological Magnetic Resonance Data Bank**



Collects NMR data from any experiment and captures assigned chemical shifts, coupling constants, and peak lists for a variety of macromolecules; contains derived annotations such as hydrogen exchange rates, pKa values, and relaxation parameters.

PDBe

- › **Protein Data Bank in Europe**



Rich information about all PDB entries, multiple search and browse facilities, advanced services including PDBePISA, PDBeFold and PDBeMotif, advanced visualization, and validation of NMR and EM structures.

wwPDB Resources

Data Dictionaries

- › **Macromolecular Dictionary (PDBx/mmCIF)**
- › **Small Molecule Dictionary (CCD)**
- › **Peptide-like antibiotic and inhibitor molecules (BIRD)**

Annotation

- › **Procedures and policies**
- › **Improvements for consistency and accuracy**

Community Input: Task Forces and Working Groups

- › **Validation Task Forces (X-ray, NMR, 3DEM)**
- › **Small Angle Scattering Task Force**
- › **PDB/mmCIF Working Group**
- › **Hybrid/Integrative Methods Task Force**
- › **Ligand Validation Workshop**

PDB Data Growth & Usage Statistics

News & Announcements

March 31, 2016

- › **Zika Virus Structure Released**

The 3D structure of the Zika virus, determined by cryo-electron microscopy, has been released. Public availability of the Zika virus atomic coordinates to medical researchers worldwide will accelerate new antiviral drug and vaccine development.

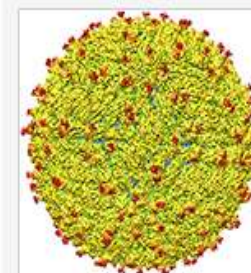


Photo Credit: Purdue University

[Read more](#)



- 1 DEPOSIT DATA**

- 2 GET PERMANENT IDENTIFIER**

- 3 WATCH YOUR CITATIONS GROW!**

- 4 RELAX, YOUR DATA ARE DISCOVERABLE AND SECURE**


<http://dx.doi.org/10.5061/dryad.20>

[Submit data now](#)

[How and why?](#)

Search for data

Enter keyword, author, title, DOI, etc [Go](#)

[Advanced search](#)

Browse for data

Recently published

Popular

Recently published data

Méndez-Janovitz M, Macías García C (2017) Data from: Do male fish prefer them big and colourful? Non-random male courtship effort in a viviparous fish with negligible paternal investment. *Behavioral Ecology and Sociobiology*
<https://doi.org/10.5061/dryad.9s380>

Dominoni DM, Åkesson S, Klaassen R, Spoelstra K, Bulla M (2017) Data from: Methods in field chronobiology. *Philosophical Transactions of the Royal Society B*
<https://doi.org/10.5061/dryad.511h1>

Becker DJ, Streicker DG, Altizer S (2017) Data from: Using host species traits to

Latest from @datadryad

Tweets by @datadryad

 Dryad Retweeted

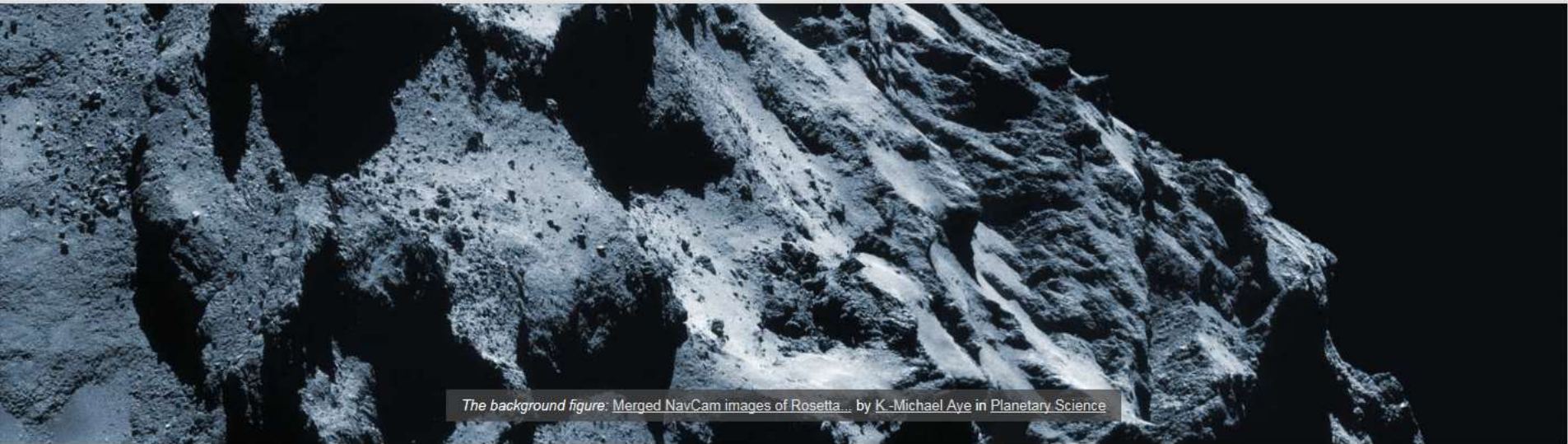
 **UC Curation Center**
@UC3CDL

OrgID Working Group issues Request for Information. Interested in developing open identifier registry? Submit Nov15
uc3.cdlib.org/2017/10/10/org...

  6h

 Dryad Retweeted

 **Simon Hodson**



The background figure: Merged NavCam images of Rosetta... by K-Michael Aye in Planetary Science

Discover research from figshare

FEATURED CATEGORIES

[Agricultural and Veterinary Sciences](#)[Astronomy, Astrophysics, Space Science](#)[Biological Sciences](#)[Built Environment and Design](#)[Chemistry](#)[Commerce, Management, Tourism and Services](#)[Earth and Environmental Sciences](#)[Engineering](#)[Health Sciences](#)[Humanities](#)[Information And Computing Sciences](#)[Language, Communication and Culture](#)[Mathematics](#)[Meta Science](#)[Physics](#)[Psychology](#)[Social Science](#)[Studies in Creative Arts and Writing](#)[Studies in Human Society](#)



Share, archive, and get credit for your data. Find and cite data across all research fields.

 Find

[Advanced Search](#)

 Add Data

 **Dataverses (2,348)**

 **Datasets (74,726)**

 **Files (353,618)**

Dataverse Category

- [Research Project \(716\)](#)
- [Researcher \(673\)](#)
- [Organization or Institution \(195\)](#)
- [Journal \(116\)](#)
- [Research Group \(63\)](#)

[More...](#)

Metadata Source

- [Harvested \(51,643\)](#)
- [Harvard Dataverse \(25,431\)](#)

Publication Date


- [2015 \(15,557\)](#)
- [2011 \(9,626\)](#)
- [2012 \(8,151\)](#)
- [2016 \(4,658\)](#)
- [2009 \(4,020\)](#)

[More...](#)

Subject

- [Social Sciences \(33,161\)](#)
- [Medicine, Health and Life Sciences \(1,709\)](#)

1 to 10 of 77,074 Results

 Sort ▾

Replication Data for: Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach 

Oct 11, 2017 - American Journal of Political Science (AJPS) Dataverse



Brendan Nyhan; Christopher Skovron; Rocio Titiunik, 2016, "Replication Data for: Differential Registration Bias in Voter File Data: A Sensitivity Analysis Approach", doi:10.7910/DVN/LCDBRU, Harvard Dataverse, V2, UNF:6:dfpHOaXg54q+BviR8l0bDA==

The widespread availability of voter files has improved the study of political participation in American politics, but the lack of good data on non-registrants creates difficult inferential issues. Most notably, observational studies that examine turnout rates among registrants o...


Replication Data for Institutions, Norms and Accountability: A Corruption Experiment with Northern and Southern Italians 

Oct 11, 2017 - Journal of Experimental Political Science



Zhang, Nan, 2017, "Replication Data for Institutions, Norms and Accountability: A Corruption Experiment with Northern and Southern Italians", doi:10.7910/DVN/XCSQ9C, Harvard Dataverse, V1, UNF:6:QYDNL1Y9qm/rzgiHk0SA==

Anti-corruption research has highlighted the potential for grassroots monitoring to improve governance outcomes, but the conditions under which citizens are willing to report bribery remain under-studied. Are individuals from some societies socialized into a "culture of corruptio...

Replication Data for: "Adoption of Distributed Solar across India" 

Oct 10, 2017 - Chao-yo Cheng Dataverse



Michaël Aklin; Cheng, Chao-yo; Johannes Urpelainen, 2017, "Replication Data for: "Adoption of Distributed Solar across India"", doi:10.7910/DVN/GITJQ4, Harvard Dataverse, V1, UNF:6:NRwNn3JA+Gg4Ct6TcS6ORw==

This package includes all datasets and codes (in R) to replicate all analyses and figures in "Adoption of Distributed Solar across India," a forthcoming article on Energy for Sustainable Development. To cite the dataset: Aklin, Michaël; Cheng, Chao-yo; Urpelainen, Johannes. Repli...

Chao-yo Cheng Dataverse (UCLA) 



Oct 10, 2017

Recent uploads

August 30, 2017 (v2) Working paper Open Access

View

Introducing Parsl: A Python Parallel Scripting Library

Babuji, Yadu; Brizius, Alison; Chard, Kyle; Foster, Ian; Katz, Daniel S.; Wilde, Michael; Wozniak, Justin

Researchers frequently rely on large-scale and domain-specific workflows to conduct their science. These workflows may integrate a variety of independent software functions and external applications. However, developing and executing such workflows can be difficult, requiring complex...

Uploaded on September 15, 2017

1 more version(s) exist for this record

September 6, 2017 (v4) Software Open Access

View

geodynamics/pylith: PyLith v2.2.1

Brad Aagaard; Charles Williams; Matthew Knepley

Added new examples. examples/3d/subduction: New suite of examples for a 3-D subduction zone. This intermediate level suite of examples illustrates a wide range of PyLith features for quasi-static simulations. examples/2d/subduction: Added quasi-static spontaneous rupture earthquake cycle...

Uploaded on September 6, 2017

3 more version(s) exist for this record

August 31, 2017 (v1) Figure Open Access

View

Fig. 1 in Vanmanenia orcicampus | a new species of loach from the Plain of Jars, Laos (Teleostei: Gastromyzontidae)

Maurice Kottelat

Fig. 1. Vanmanenia orcicampus | MHNG 2767.094 | holotype | 47.5 mm SL; Laos: Nam Ngum on Plain of Jars.

Uploaded on September 6, 2017

Zenodo now supports DOI versioning!



Read more about it, in our newest blog post.

Using GitHub?



Just Log in with your GitHub account and click here to start preserving your repositories.

Zenodo in a nutshell

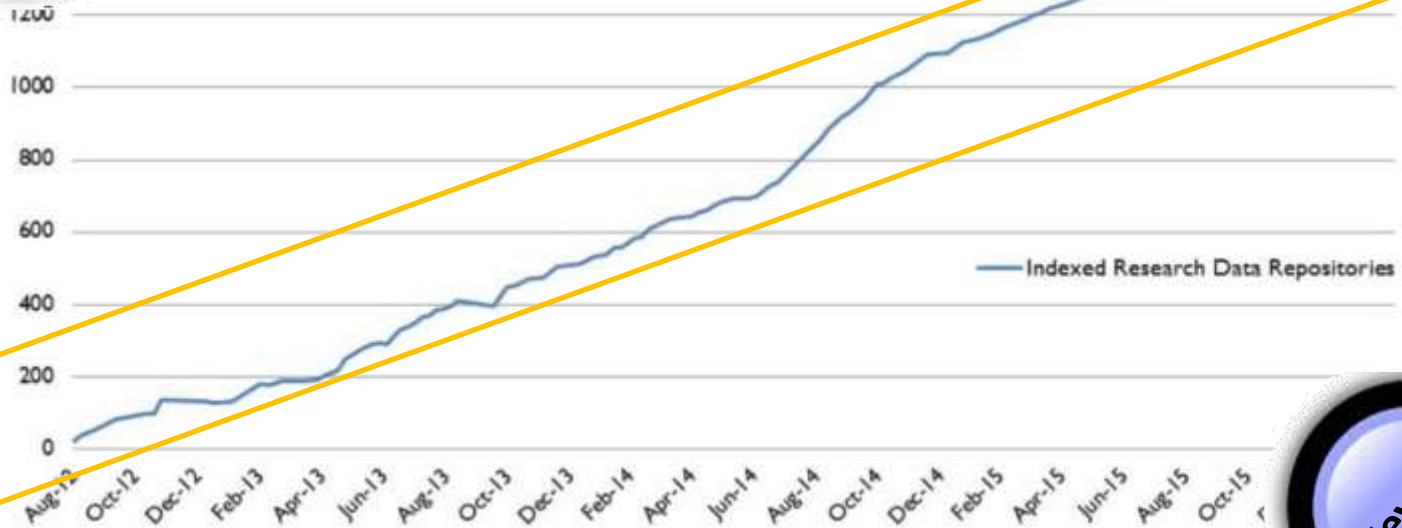
- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- **Flexible licensing** — because not everything is under Creative Commons.
- **Safe** — your research output is stored safely for the future in the same cloud infrastructure





1600

Indexed Research Data Repositories



1500



Feb-16

COUNTRY: Brazil

Brazil Exploration and Production Database

WorldClim - Global Climate Data

Integrated Ocean Discovery Program

Global Collaboration Engine

Repositório de Dados de Levantamentos Biológicos

PLATAFORMAS PARA GESTÃO DE DADOS DE PESQUISA



INVENIO é um framework para **bibliotecas digitais, repositórios digitais, arquivos multimídia ou repositório de dados de pesquisa** disponíveis na web. <http://invenio-software.org/>

Seguem exemplos de utilizadores:

Zenodo, um serviço do CERN que permite o compartilhamento de dados de pesquisa, utiliza o Invenio – <https://zenodo.org/>

A **biblioteca digital das Nações Unidas** também utiliza este framework em seu repositório – <https://digitallibrary.un.org/>



CKAN (Comprehensive Knowledge Archive Network) é um sistema de depósito e gerenciamento de dados que oferece ferramentas para publicação, compartilhamento, descoberta e uso de dados. É um sistema **voltado a governos nacionais e regionais, companhias e organização que desejam disponibilizar seus dados** (IBICT). O **CKAN** é uma [aplicação web](#) de catalogação de dados desenvolvido pela [Open Knowledge Foundation](#). O CKAN mantém uma descrição dos dados nele armazenados, bem como outras informações úteis, como em que formatos eles estão disponíveis, quem é o seu dono e se eles estão livremente disponíveis, e com quais assuntos ele se relaciona. Tal como ocorre num [wiki](#), outros usuários podem melhorar ou acrescentar informações, enquanto o CKAN guarda um histórico versionado (Wikipedia).



DATAVERSE is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility. <http://dataverse.org>

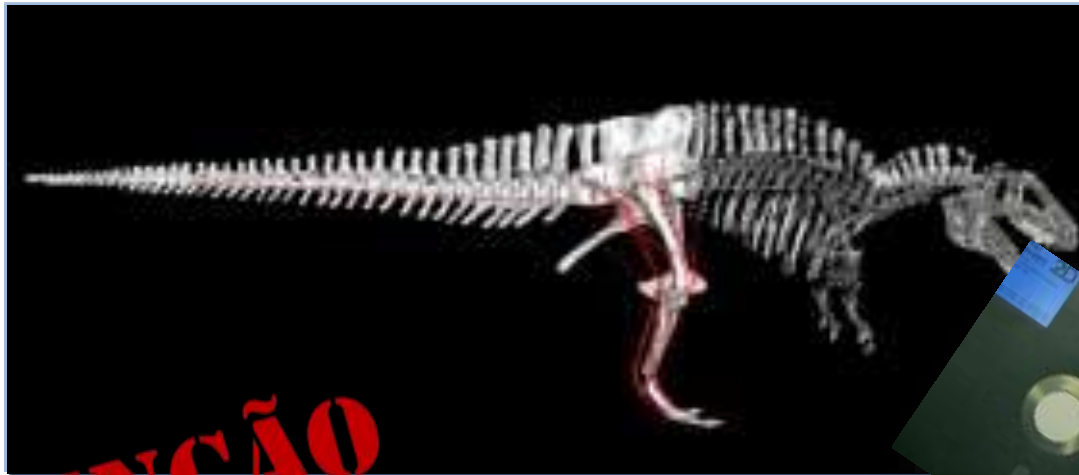
Dataverse supports DataCite DOIs. You can configure your installation following this guide: <http://guides.dataverse.org/en/latest/installation/config.html?highlight=datacite>



O **DSpace** é um software dedicado à implementação de repositórios e que vem sendo amplamente adotado no Brasil. **DSpace** <http://www.dspace.org>

DSpace is an open source software platform that enables organisations to: capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options distribute an organisation's digital assets over the web through a search and retrieval system preserve digital assets over the long term.

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

A **WEB** não tem
MEMÓRIA
PERSISTENTE,
portanto é preciso
CONSTRUI-LA
INTENCIONALMENTE



Como adicionar confiança aos repositórios digitais



信
頼

As instituições de patrimônio **já são confiáveis**. Elas são encarregadas dos materiais e dos objetos que documentam nossa herança cultural. Elas são instituições reconhecidas historicamente como confiáveis para armazenar e preservar esses valiosos materiais e para promover o acesso a eles.

Porém, como a informação digital é menos tangível que outros materiais e muito mais mutável, **confiança e credibilidade podem ser mais difícil de provar**.

A certificação é um elemento essencial para confiabilidade dos repositórios digitais
PADRÕES, BOAS PRÁTICAS, GESTÃO, TECNOLOGIA, ORGANIZAÇÃO



As organizações estão crescentemente interessadas em avaliar suas **infraestruturas de preservação digital** tomando como referência arcabouços padronizados

AUDITORIA e CERTIFICAÇÃO

AUTOAVALIAÇÃO

A IMPORTÂNCIA DOS PROCESSOS DE AVALIAÇÃO RESIDE NO FATO DELES PROMOVEREM A CONFIANÇA NA USABILIDADE, SUSTENTABILIDADE E PERSISTÊNCIA POR LONGO PRAZO DOS DADOS DISPONÍVEIS PARA COMPARTILHAMENTO.

CERTIFICAÇÃO & AUDITORIA

A IMPORTÂNCIA DOS PROCESSOS DE AVALIAÇÃO RESIDE NO FATO DELES PROMOVEREM A CONFIANÇA NA USABILIDADE, SUSTENTABILIDADE E PERSISTÊNCIA POR LONGO PRAZO DOS DADOS DISPONÍVEIS PARA COMPARTILHAMENTO.



O DATA SEAL APPROVAL (DSA) É UM PROCESSO DE AUTOAVALIAÇÃO PARA ARQUIVOS DIGITAIS QUE OBJETIVA ESPECIFICAMENTE OS ARQUIVOS QUE DETÉM DADOS



OS DADOS PODEM SER **ENCONTRADOS** NA INTERNET

OS DADOS SÃO **ACESSÍVEIS** (direitos e licenças claros)

OS DADOS ESTÃO EM **FORMATOS ADEQUADOS** PARA O USO

OS DADOS SÃO **CONFIÁVEIS**

OS DADOS SÃO **IDENTIFICADOS** DE MANEIRA ÚNICA E PERSISTENTE DE FORMA QUE POSSAM REFERENCIADOS

PRESERVING DIGITAL INFORMATION: REPORT OF THE TASK FORCE ON ARCHIVING OF DIGITAL INFORMATION - CPA/RLG

COLOCA OS DESAFIOS: EVITAR A OBSOLESCENCIA TECNOLÓGICA; MIGRAÇÃO; QUESTÕES LEGAIS E ORGANIZACIONAIS; INFRAESTRUTURA DE GESTÃO

CONTEÚDO | FIXIDADE | REFERÊNCIA | PROVENIÊNCIA | CONTEXTO

REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS) - CCSDS

TRUSTED DIGITAL REPOSITORIES: ATTRIBUTES AND RESPONSIBILITIES (TDR) - RLG-OCLC - 2002

Examina os requisitos que uma organização deve cumprir para estar em conformidade com o OAIS.

TRUSTWORTHY REPOSITORIES AUDIT & CERTIFICATION: CRITERIA AND CHECKLIST (TRAC) - RLG/NARA - 2007

Examina os requisitos que uma organização deve cumprir para estar em conformidade com o OAIS.

Preserving Digital Information

Report of the Task Force on Archiving of Digital Information

commissioned by

The Commission on Preservation and Access and The Research Libraries Group

Consultative Committee for Space Data Systems

RECOMMENDATION FOR SPACE DATA SYSTEM STANDARDS

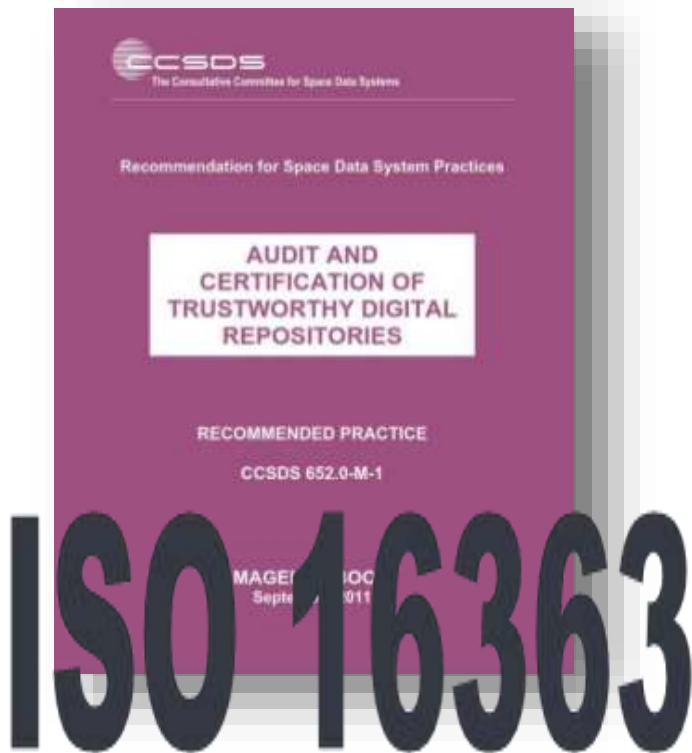
Reference Model for an Open Archival Information System (OAIS)

CCSDS 600.0-P-1
BLUE BOOK

Trusted Digital Repositories: Attributes and Responsibilities

An RLG-OCLC Report





INFRAESTRUTURA ORGANIZACIONAL

Esta dimensão abrange a estrutura governativa do repositório e a sua viabilidade organizacional, analisando as questões relacionadas com a gestão do repositório, processos e recursos humanos afetos. Inclui ainda: políticas de preservação, documentação dos processos, sustentabilidade financeira das instituições que gerem os repositórios, contratos, licenças e responsabilidades do serviço.

GESTÃO DE OBJETOS DIGITAIS

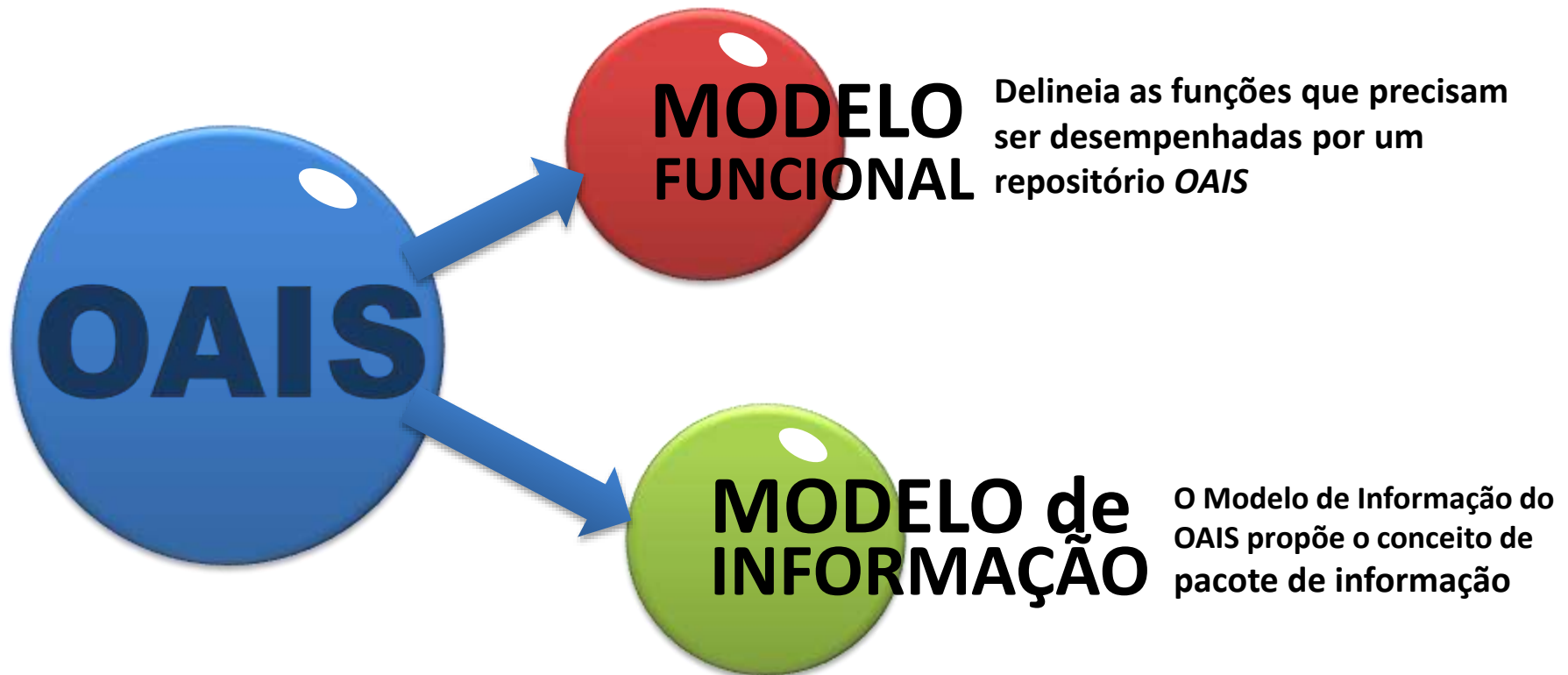
Esta dimensão analisa o processo de ingestão e gestão de objetos digitais do repositório, ou seja, a forma como incorpora a informação digital assim como a criação e gestão dos Pacotes de Informação de Arquivo (AIP). Engloba ainda o planeamento da preservação e a forma como os AIP são preservados. Finalmente, observa as componentes de gestão de informação do serviço e a gestão de acessos.

INFRAESTRUTURA E GESTÃO DA SEGURANÇA

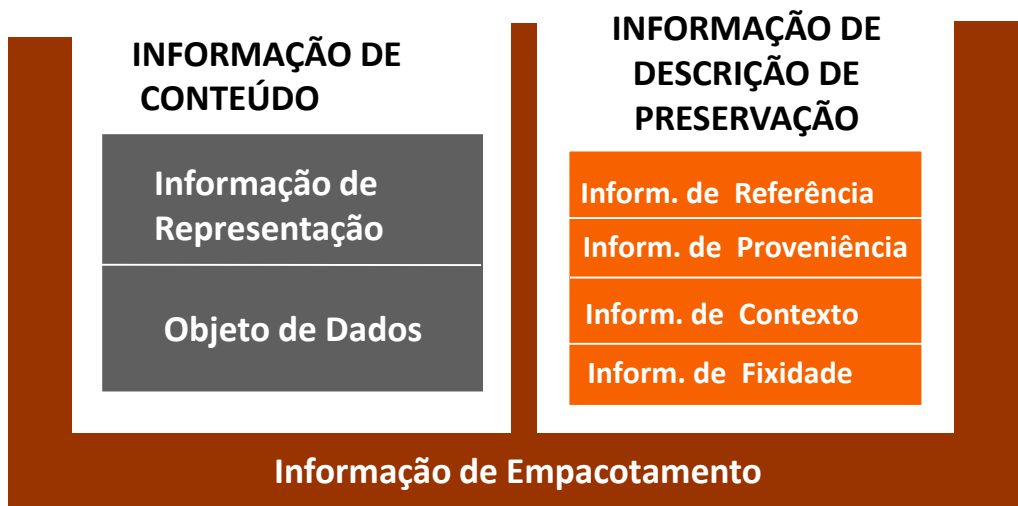
A última dimensão engloba as questões técnicas relacionadas com a gestão e controlo de riscos inerentes à infraestrutura e a gestão da segurança. Esta componente relaciona-se diretamente com as infraestruturas disponibilizadas pelo serviço SARI do projeto RCAAP.

OAIS: MODELO DE DADOS

O *OAIS* é composto por dois modelos: o modelo funcional e o modelo de informação. O modelo funcional delinea as funções que precisam ser desempenhadas por um repositório *OAIS*. A figura 1 apresenta os componentes funcionais, os pacotes de informação e as entidades externas de um repositório digital compatível com o *OAIS*.



O Modelo de Informação do *OAIS* propõe o conceito de **pacote de informação** (figura 3), que é formado pela informação de conteúdo e pela informação de descrição de preservação, encapsuladas e identificadas pela informação de empacotamento. A informação de conteúdo é o objeto de informação (objeto de dado + informação de representação) a ser preservado. A informação de descrição de preservação é a informação necessária para a adequada preservação da informação de conteúdo, e que pode ser categorizada como informação sobre proveniência, referência, fixidade e contexto.



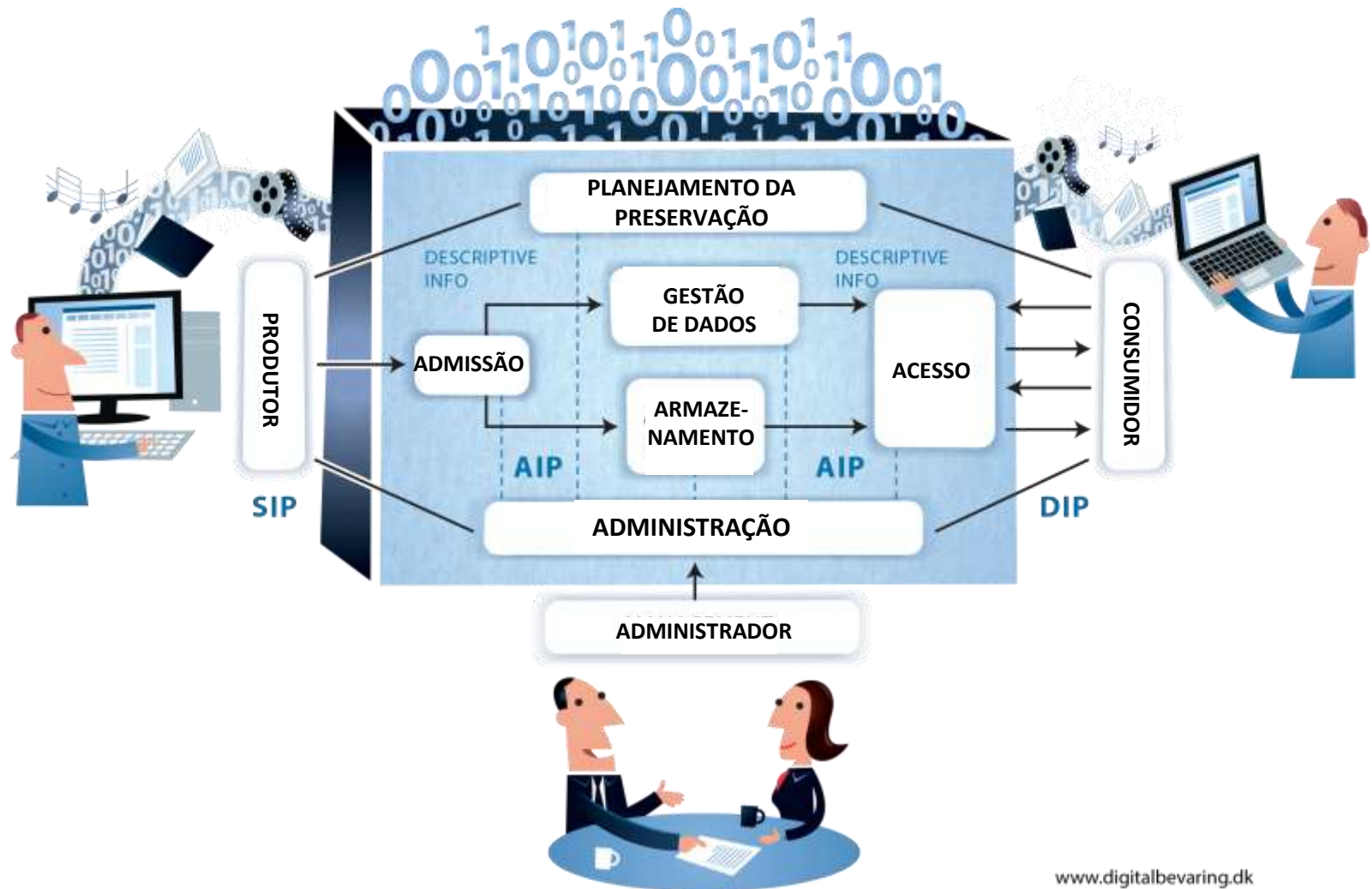
Informação
Descritiva sobre
o Pacote 1
[indexação]

O MODELO DE INFORMAÇÃO DO OAIS PROPÕE O CONCEITO DE PACOTE DE INFORMAÇÃO, QUE É FORMADO PELA INFORMAÇÃO DE CONTEÚDO E PELA INFORMAÇÃO DE DESCRIÇÃO DE PRESERVAÇÃO, ENCAPSULADAS E IDENTIFICADAS PELA INFORMAÇÃO DE EMPACOTAMENTO.

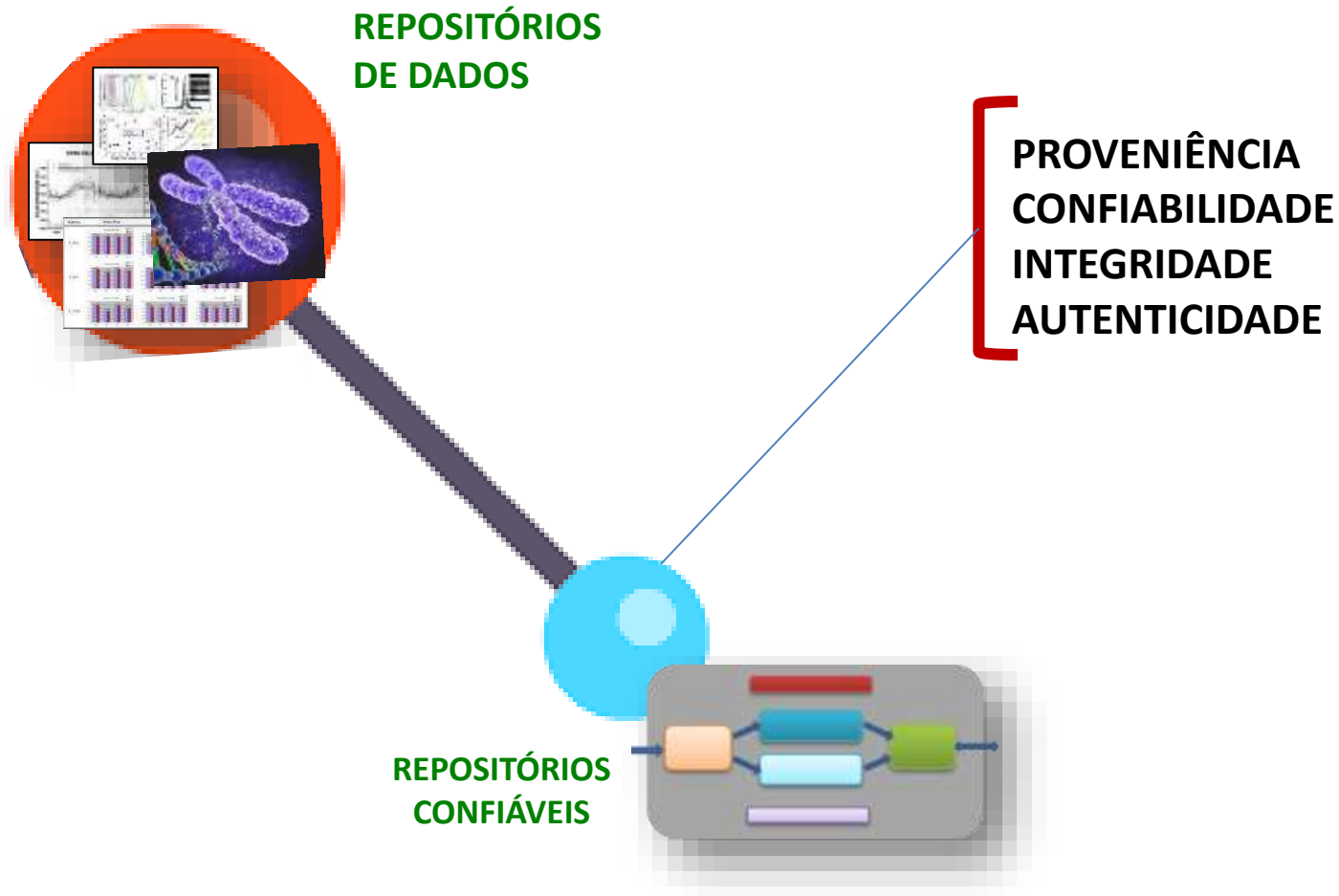
001100
100110
010010
010100
100001
010000
010111
001011
001010
101000
100001



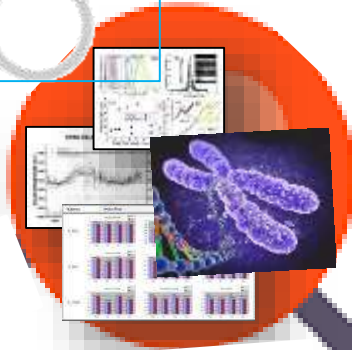
OAIS: MODELO FUNCIONAL



INTEGRAÇÃO REPOSITÓRIO DE DADOS x REPOSITÓRIO CONFIÁVEL



INTEGRAÇÃO DATAVERSE X ARCHIVEMATICA



PROVENIÊNCIA
CONFIABILIDADE
INTEGRIDADE
AUTENTICIDADE

REPOSITÓRIOS
CONFIÁVEIS



ARCHIVEMATICA é um conjunto integrado de ferramentas abertas de software que permite aos usuários processar objetos digitais da entrada até o acesso segundo os princípios do **modelo funcional do OAIS**.

Usa o **METS**, **PREMIS** e o **BAGIT** para gerar o **PACOTES DE INFORMAÇÃO DE ARQUIVAMENTO (AIP)** Confiáveis, autênticos e independentes de sistemas.



INTEGRAÇÃO DATAVERSE X ARCHIVEMATICA



API ou SWORD

