Desafios da Ciência Invisível:

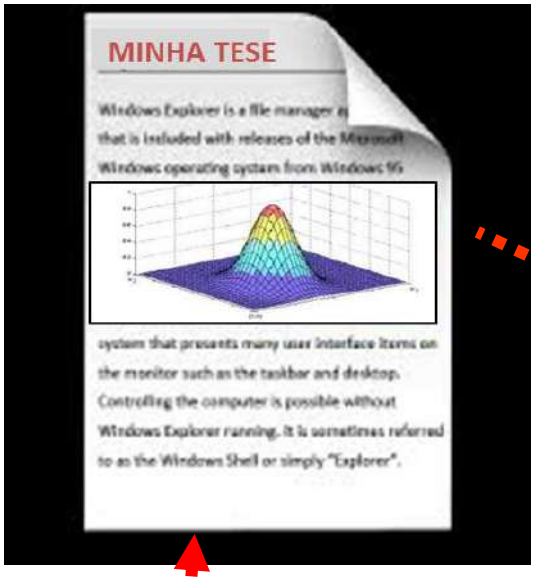
Dados da cauda longa da pesquisa

Luis Fernando Sayão
CNEN/CIN

DADOS DE PESQUISA SÃO MUITO SUSCETÍVEIS A PERDAS

PROJETO DE PESQUISA

PUBLICAÇÃO

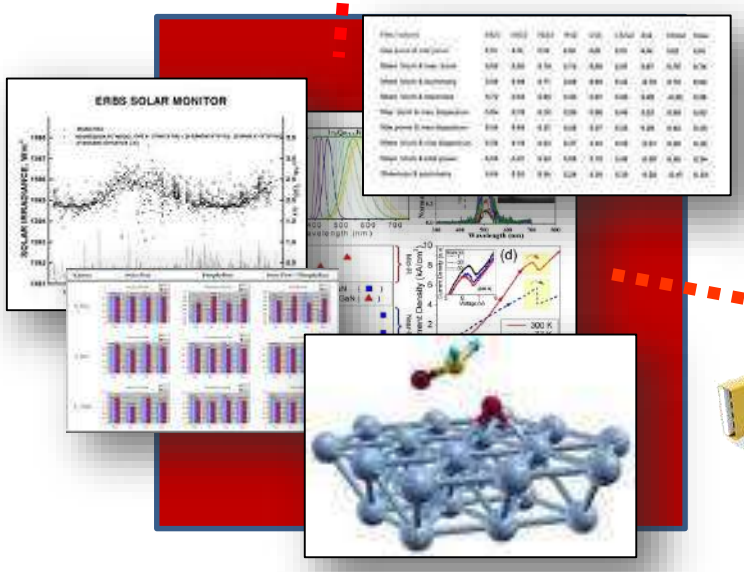


BIBLIOTECA CONVENCIONAL

BIBLIOTECA DIGITAL

REPOSITÓRIO DIGITAL

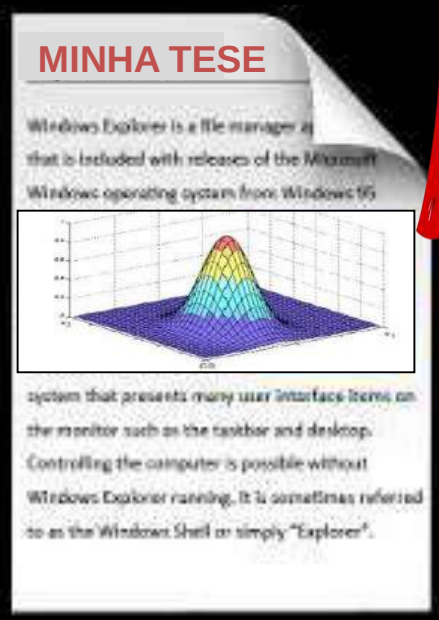
DADOS de PESQUISA

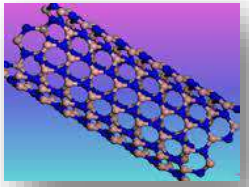


VISÍVEL

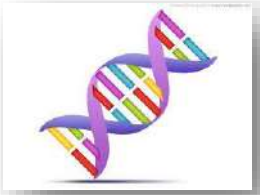
INVISÍVEL

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



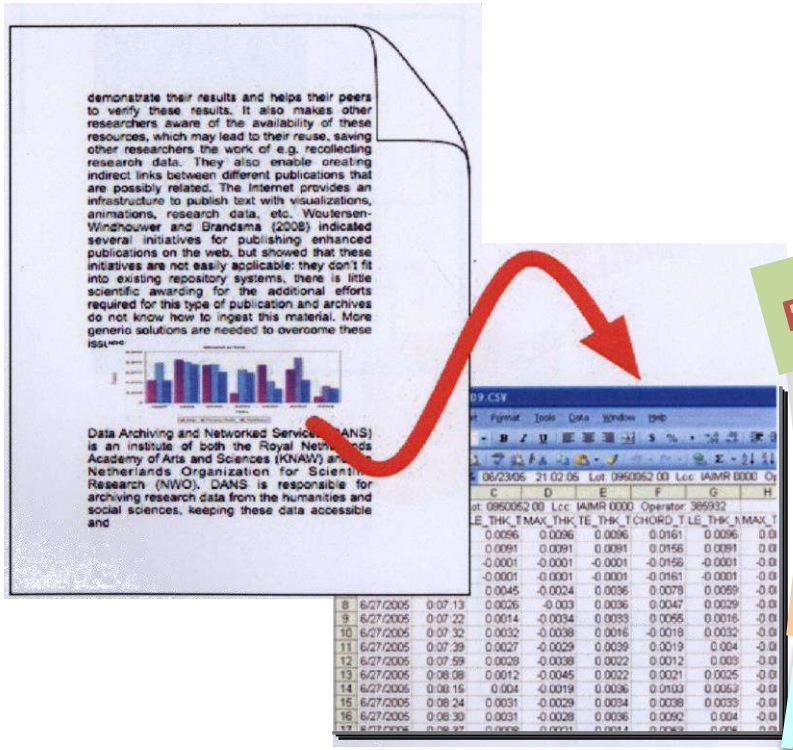


UMA VISÃO DOS DADOS !!!



[revisão por pares]
[validação da pesquisa]

OS DADOS SUBJACENTES ÀS PUBLICAÇÕES APOIAM A:



REPRODUTIBILIDADE

VALIDAÇÃO

AUTOCORREÇÃO

AVALIAÇÃO POR PARES

INTERPRETAÇÃO

BASE PARA NOVAS PESQUISAS

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

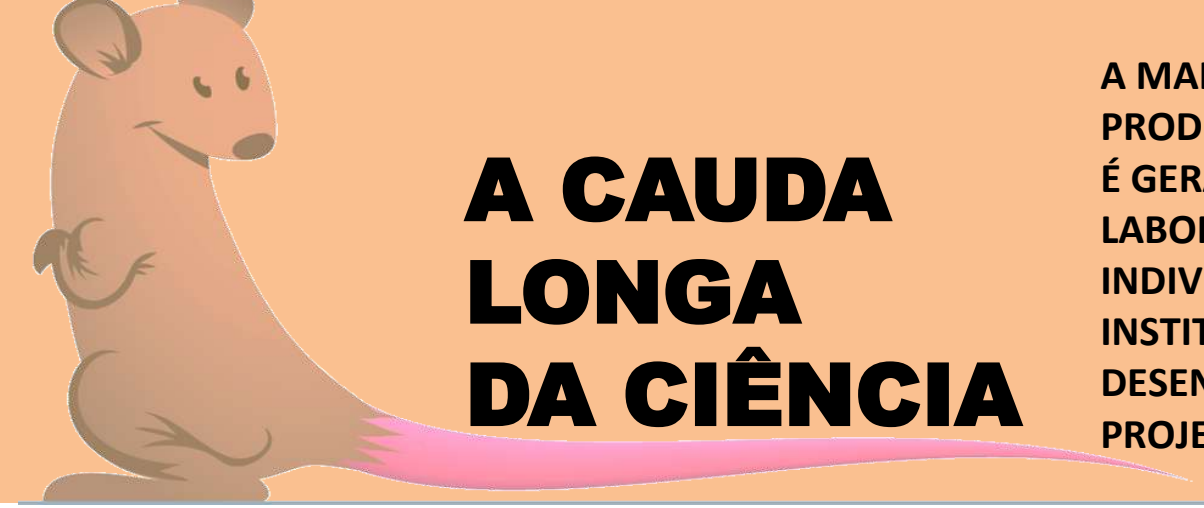
ORIENTADA POR HIPÓTESES


ORGANIZAÇÕES
UNIVERSIDADES
INSTITUTOS DE PESQUISA
AGÊNCIAS DE FOMENTO
BIBLIOTECAS, ARQUIVOS, MUSEUS
ORGANIZAÇOES VIRTUAIS;
COMUNIDADES
EXPERTISES
PESQUISADORES
CIENTISTAS DE DADOS
BIBLIOTECARIOS DE DADOS
ARQUIVISTAS
INSTRUMENTOS CIENTÍFICOS
TELECÓPIOS
SATÉLITES
COLISORES
SENSORES
CIBERINFRAESTRUTURA DE PESQUISA
DADOS
BASES DE DADOS
REPOSITÓROS
ACESSO
GESTÃO
CURADORIA
MINERAÇÃO
PRIVACIDADE
REDES
REDES DE PESQUISA/EDUCAÇÃO NACIONAIS E INTERNACIONAIS;
SEGURANÇA
RECURSOS COMPUTACIONAIS
SUPERCOMPUTADORES
NUVEM, GRID, CLUSTER;
VISUALIZAÇÃO;
CENTROS DE COMPUTAÇÃO
SOFTWARE
APLICAÇÕES;
DESENVOLVIMENTO
E SUPORTE

PIRÂMIDE DE GESTÃO DE DADOS

HIERAQUIA DE VALOR & PERMANÊNCIA

Baseado em THE ROYAL SOCIETY (June 2012)



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

DOMÍNOS ESPECÍFICOS

ASTRONOMIA
FISICA NUCLEAR
GENOMA
PROTEINA
SENSORIAMENTO REMOTO

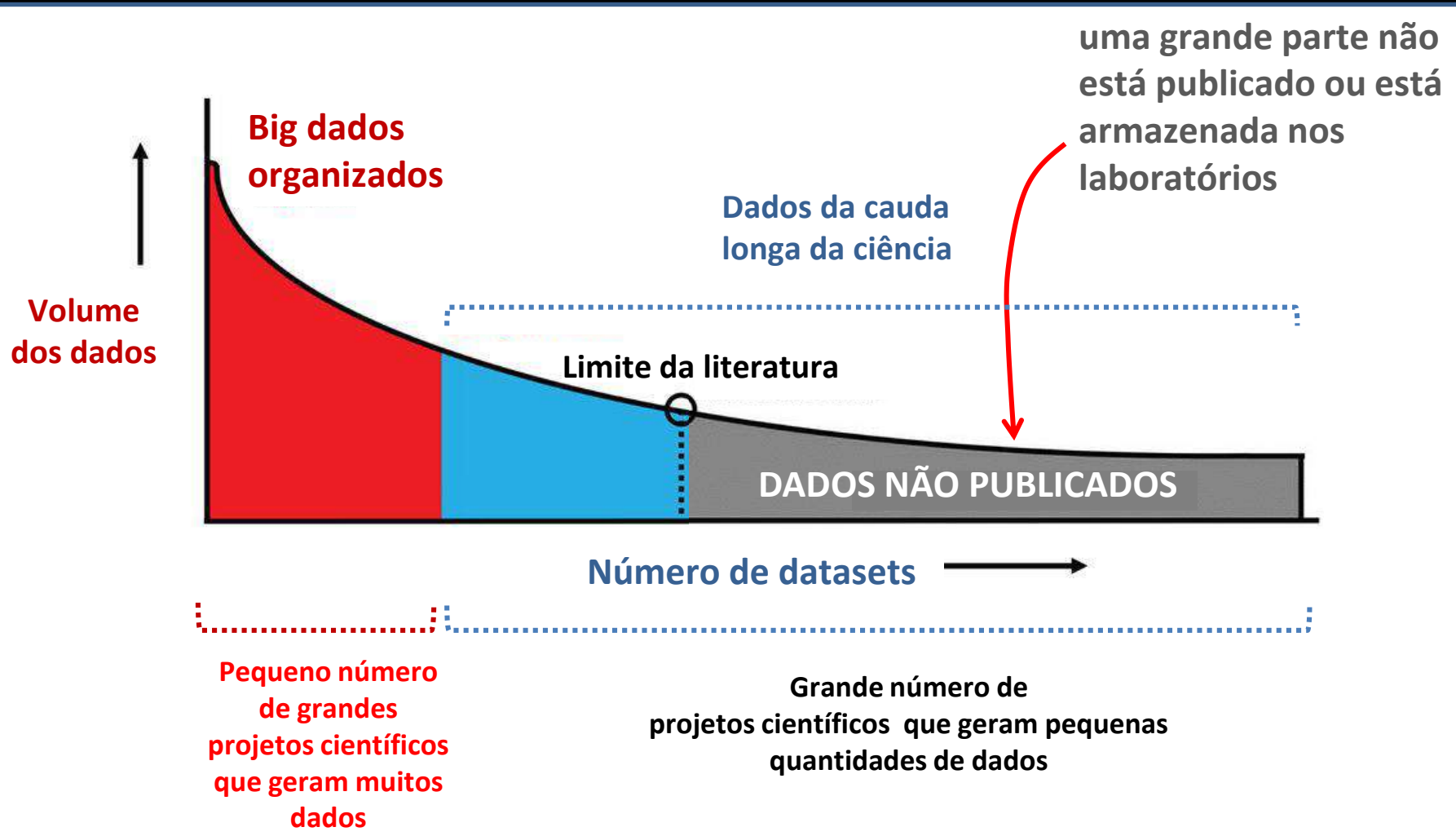
Dados da **Grande Ciência** são fáceis de **manipular, compreender e arquivar;**
A **Pequena Ciência** é terrivelmente **heterogênea** e muito mais vasta e gera **2-3 vezes mais** dados do que a Big Science (MacColl, 2010)

Volume dos dados

VÁRIOS DOMÍNIOS E INSTITUIÇÕES

Número de datasets

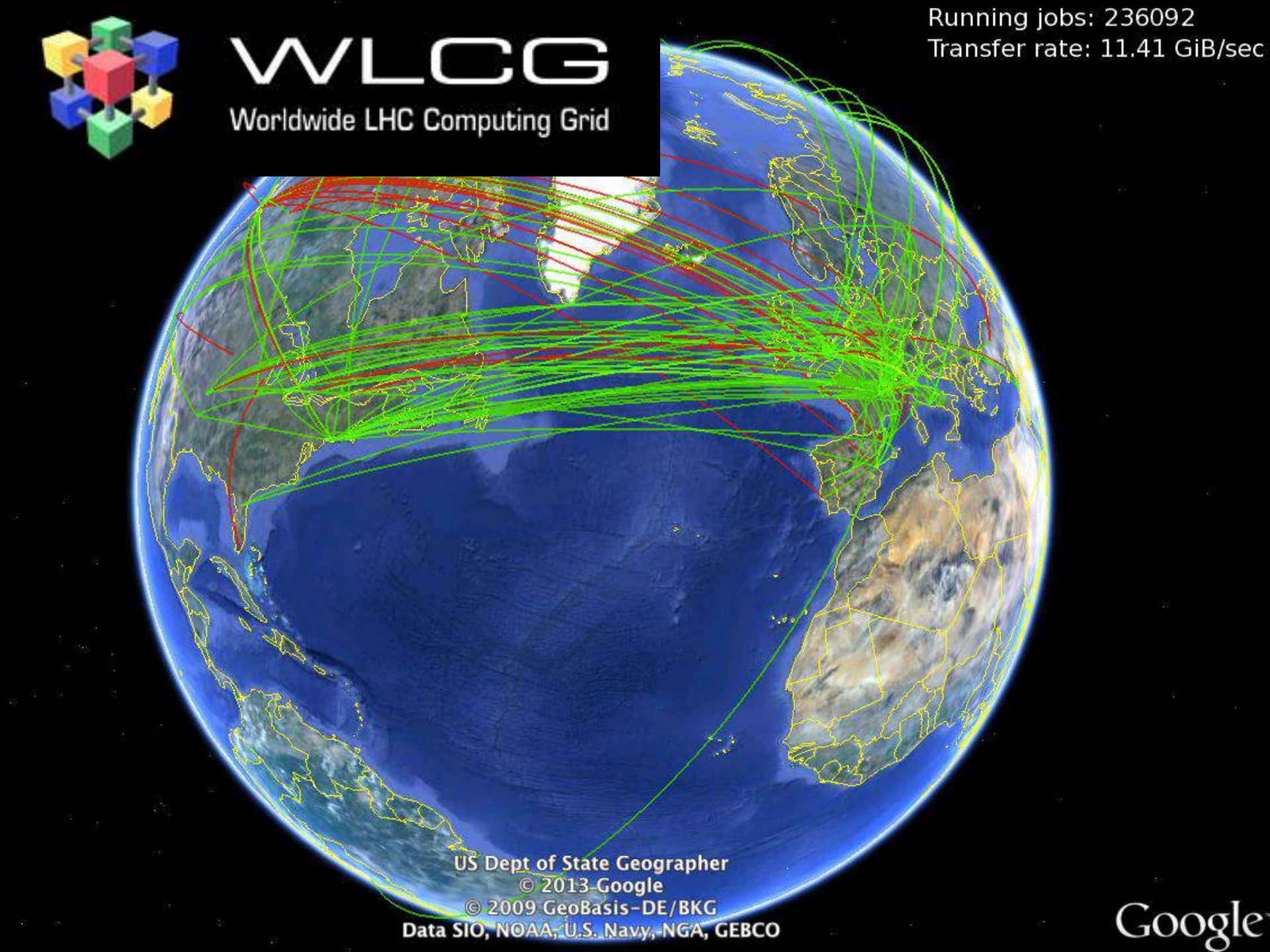
PEQUENOS LABORATÓRIOS, EQUIPES E PESQUISADORES INDIVIDUAIS



Os dados gerados ou coletados em decorrência dos pequenos projetos de pesquisa são distribuídos por todos os domínios do conhecimento, das artes e humanidades até as áreas mais identificadas como os padrões da grande ciência como física e astronomia

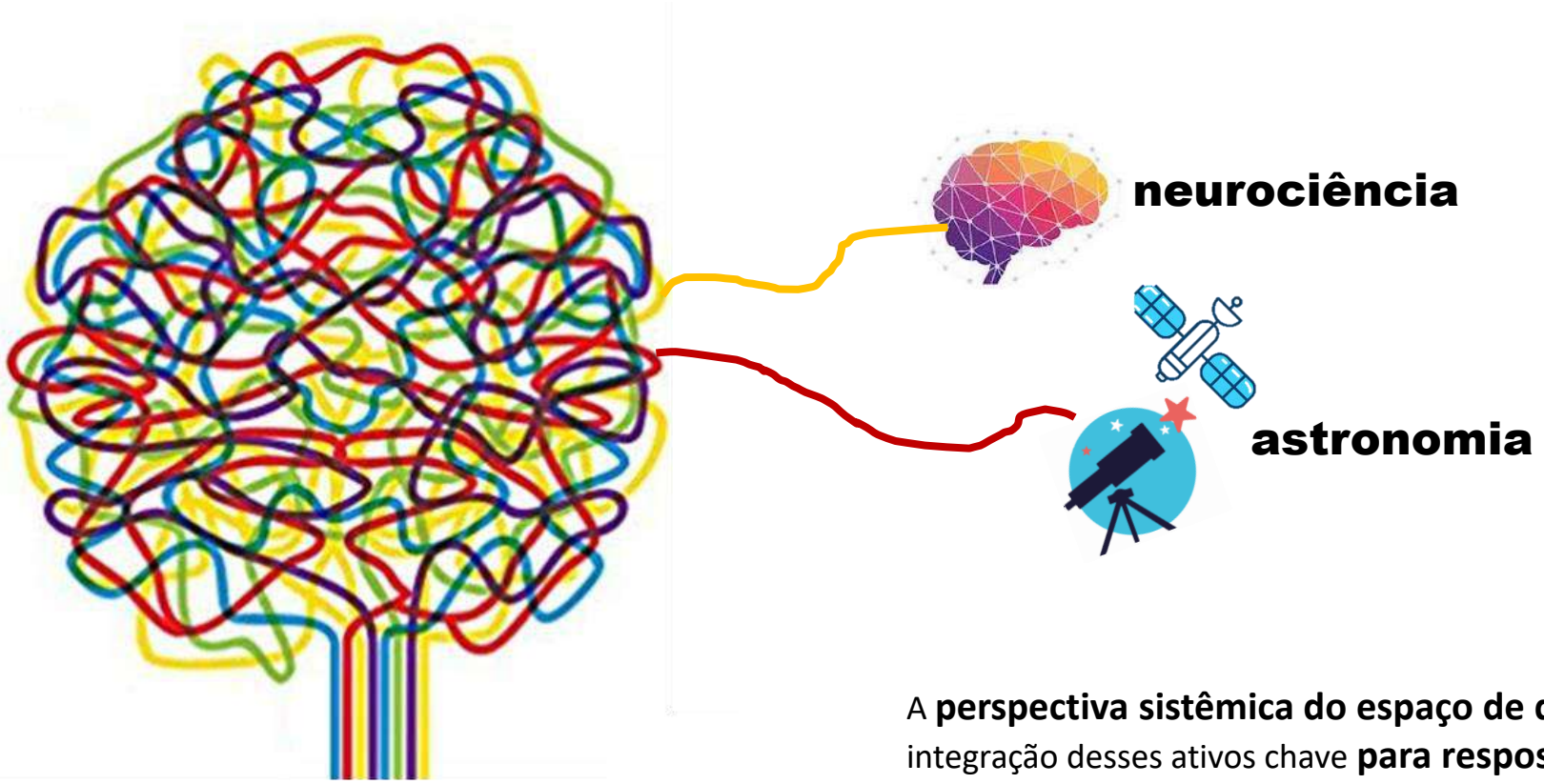
Parece provável que a ciência transformadora venha mais da cauda do que da cabeça (Heidorn, 2008)

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


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

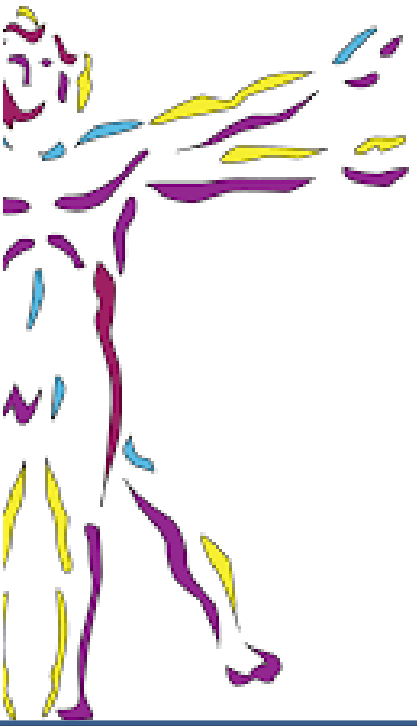
DIVERSIDADE DOS DADOS

Os dados da cauda longa, com **sua natureza heterogênea e diversificada**, devem se **integrar a homogeneidade da grande ciência** formando **uma ecologia ou diversidade de dados**. Isto por que nem sempre a grande ciência, definida por predicados homogêneos e estáveis **é o modelo mais adequado para algumas das áreas mais avançadas** e inovadoras da pesquisa científica. Na maioria das vezes, integrar dados formando uma diversidade de dados transversalmente rica, estabelece modelos eficientes de geração de conhecimento



transdisciplinaridade

A **perspectiva sistêmica do espaço de dados** torna a integração desses ativos chave **para respostas a novas indagações da ciência**. Isso acontece especialmente ao vincular a estabilidade da grande ciência ao território de alto coeficiente de autonomia e independência da cauda longa, cujas condutas desafiadoras favorecem a inovação e a geração de conhecimentos multi e interdisciplinar.



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

A ABERTURA DOS DADOS E O SEU IMPACTO NA COMUNICAÇÃO CIENTÍFICA

Novos tipos de publicação

- Publicação ampliada
- Data journal
- Periódico de resultados negativos
- Cadernos eletrônicos de laboratório

Encurtamento do ciclo de comunicação científica

- Compartilhamento
- Reuso
- Organização dos dados
- Novas formas de socialização

Novos modalidades de revisão

- Consulta aos dados
- Revisão em rede
- Revisão interativa
- Anotações
- Revisão dos dados

DILÚVIO DE DADOS NA CIÊNCIA

eScience

Ciência aberta

Cauda longa

Humanidades Digitais

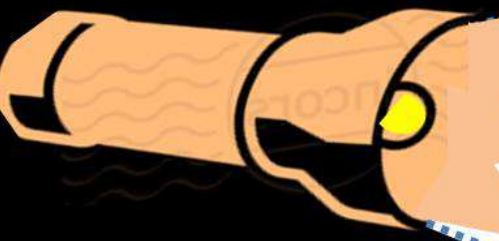
BIG DATA CIENTÍFICO
Grandes projetos
Observatórios
Instalações complexas
Dados distribuídos
Simulação por computador

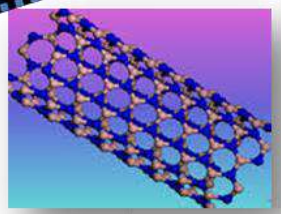
DADOS ABERTOS
Metodologias
Equipamentos
Software
Cadernos de laboratório
Roteiro de entrevistas
Resultados negativos

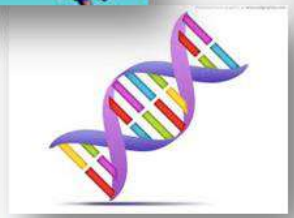
DADOS DOS DO GRANDE NÚMERO DE PEQUENOS LABORATÓRIOS
Heterogêneos
Não tratados
Invisíveis
Coletivamente é o maior volume

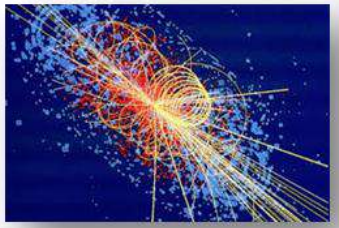
TECNOLOGIA COMPUTACIONAL APLICADAS A ESTUDOS EM HUMANIDADES.
Humanidades estudando Tecnologias digitais (Bobley)

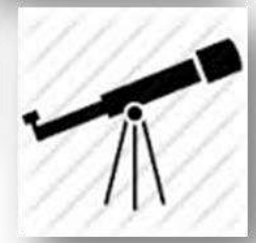
Há uma parcela dos produtos de pesquisa que necessita de infraestruturas
INFORMACIONAIS
TECNOLÓGICAS
POLÍTICAS
GERENCIAIS











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

Por que os pesquisadores não compartilham seus dados de pesquisa?

A maioria dos pesquisadores concordam em tese com os princípios de compartilhamento e reuso preconizados pela ciência aberta, mas relutam em compartilhar os seus próprios dados como parte do fluxo de pesquisa , e o fazem mais como exceção do que como regra .

AGÊNCIAS FINANCIADORAS DE PESQUISA

PLANOS DE COMPARTILHAMENTO DE DADOS

POLÍTICAS MANDATÓRIAS

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

FAPESP

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).

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.

MOTIVOS PARA O PESQUISADOR NÃO COMPARTILHAR

Restrições culturais, DISCIPLINARES e institucionais

x

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

+50%

DOS ACHADOS NÃO FORAM PUBLICADOS

LADO OCULTO DOS DADOS

DADOS NÃO PUBLICADOS

INDIVIDUAIS
Personalidade e **atitudes** do pesquisador em relação aos seus dados.

DISCIPLINARES
Cultura do compartilhamento do domínio

ORGANIZACIONAIS
Obstáculos inerentes às instituições de pesquisa em relação à gestão de dados

POLITICAS
Condicionantes políticas voltadas para **recompensa, financiamento e sustentabilidade** de longo prazo

INFRAESTRUTURAIS
Aportes tecnológicos disponíveis

O COMPRTILHAMENTO PODE REVELAR VALORES IMPORTANTES OCULTOS NESSES DADOS

BARREIRAS DISCIPLINARES

Para algumas disciplinas o compartilhamento é determinante para a geração de conhecimento, para outros é somente uma ação entre colegas.

Como a ciência é um **empreendimento humano**, argumentos a favor e contra o compartilhamento de dados frequentemente focam menos nos benefícios percebidos para a ciência como um todo e **mais nos efeitos sobre o pesquisador individualmente** (FERGUSON et al, 2014)

BARREIRAS INDIVIDUAIS

Oportunismo de outros pesquisadores
quando eles disponibilizam os seus materiais de pesquisa nas fases preliminares do processo de pesquisa, expondo-os a **abusos nos seus direitos intelectuais**;

Ser "furado"
ou seja, de ter publicações baseada nos seus dados lançadas em primeira mão por outros autores.

Explorar mais os dados
publicando o máximo possível de artigos baseados nesses dados, já que esse é o **critério academicamente mais valorizado**

Reanálises equivocadas
Dados de má qualidade e análises por não especialistas

Erros nos dados ou nas análises
temor que outros pesquisadores descubram erros nos dados ou questionam a validade das análises.

Tempo, esforço e recurso
Para que as coleções de dados possam ser reusados de ser limpas; organizadas, documentadas, anonimizadas e descritas por metadados que evidenciem os instrumentos e métodos usados para obtê-las e, finalmente, publicadas em bases de dados/repositórios

Falta de reconhecimento pelos sistemas de recompensa por organizar os dados

Falta de conhecimento das tecnologias para o compartilhamento

Resultados de experimentos que não deram certo

Problemas éticos, de privacidade e legais
"A maior barreira para o compartilhamento de dados de pesquisa são os temores dos pesquisadores **em relação às questões legais e o mau uso de seus dados**"

Baixo impacto na carreira do pesquisador.

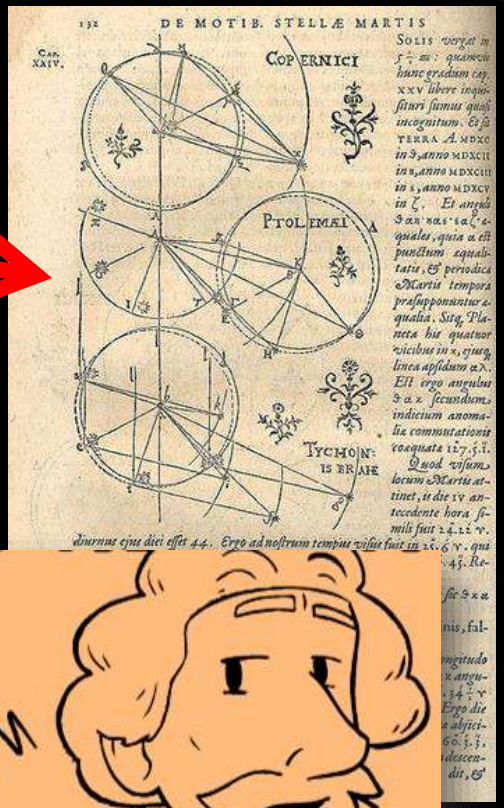
Interesse econômico

DADOS EXPERIMENTAIS

TEORIA


64 Tabularum Rudolphi

Tabula Æquationum MARTIS.


132 DE MOTIB. STELLÆ MARTIS

Cap. XXIV.

COPERNICI

PTOLEMÆI

TYCHONIS BRAHE

TICHO BRAHE

JOHANNES KEPLER

KEPLER que era assistente de TICHO BRAHE pegou o catálogo de observações astronômicas sistemáticas do TICHO e descobriu as leis do movimento planetário.

Este fato estabeleceu a divisão entre a mineração e análise de dados experimentais, cuidadosamente arquivados, e a criação de teorias

[1600]

Como a ciência é um **empreendimento humano**, argumentos a favor e contra o compartilhamento de dados frequentemente focam menos nos benefícios percebidos para a ciência como um todo e **mais nos efeitos sobre o pesquisador individualmente** (FERGUSON et al, 2014)

BARREIRAS INDIVIDUAIS

Oportunismo de outros pesquisadores
quando eles disponibilizam os seus materiais de pesquisa nas fases preliminares do processo de pesquisa, expondo-os a **abusos nos seus direitos intelectuais**;

Ser "furado"
ou seja, de ter publicações baseada nos seus dados lançadas em primeira mão por outros autores.

Explorar mais os dados
publicando o máximo possível de artigos baseados nesses dados, já que esse é o **critério academicamente mais valorizado**

Reanálises equivocadas
Dados de má qualidade e análises por não especialistas

Erros nos dados ou nas análises
temor que outros pesquisadores descubram erros nos dados ou questionam a validade das análises.

Tempo, esforço e recurso
Para que as coleções de dados possam ser reusados de ser limpas; organizadas, documentadas, anonimizadas e descritas por metadados que evidenciem os instrumentos e métodos usados para obtê-las e, finalmente, publicadas em bases de dados/repositórios

Falta de reconhecimento pelos sistemas de recompensa por organizar os dados

Falta de conhecimento das tecnologias para o compartilhamento

Resultados de experimentos que não deram certo

Problemas éticos, de privacidade e legais
"A maior barreira para o compartilhamento de dados de pesquisa são os temores dos pesquisadores **em relação às questões legais e o mau uso de seus dados**"

Baixo impacto na carreira do pesquisador.

Interesse econômico

FLUXO DOS DADOS

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

INSTRUMENTO CIENTÍFICO

DADOS BRUTOS

PROCESSAMENTO DOS DADOS

Selecionar subset
Mesclar mútiplos datasets
Conversão
Normalização
Limpeza dos dados

COMPUTAÇÃO EM NUVEM
COMPUTAÇÃO EM GRADE

ANÁLISE DOS DADOS

Estatísticas
Simulação
Plotagem
Visualização
Modelos
Algoritmos
Publicações

Como a ciência é um **empreendimento humano**, argumentos a favor e contra o compartilhamento de dados frequentemente focam menos nos benefícios percebidos para a ciência como um todo e **mais nos efeitos sobre o pesquisador individualmente** (FERGUSON et al, 2014)

BARREIRAS INDIVIDUAIS

Oportunismo de outros pesquisadores
quando eles disponibilizam os seus materiais de pesquisa nas fases preliminares do processo de pesquisa, expondo-os a **abusos nos seus direitos intelectuais**;

Ser "furado"
ou seja, de ter publicações baseada nos seus dados lançadas em primeira mão por outros autores.

Explorar mais os dados
publicando o máximo possível de artigos baseados nesses dados, já que esse é o **critério academicamente mais valorizado**

Reanálises equivocadas
Dados de má qualidade e análises por não especialistas

Erros nos dados ou nas análises
temor que outros pesquisadores descubram erros nos dados ou questionam a validade das análises.

Tempo, esforço e recurso
Para que as coleções de dados possam ser reusados de ser limpas; organizadas, documentadas, anonimizadas e descritas por metadados que evidenciem os instrumentos e métodos usados para obtê-las e, finalmente, publicadas em bases de dados/repositórios

Falta de reconhecimento pelos sistemas de recompensa por organizar os dados

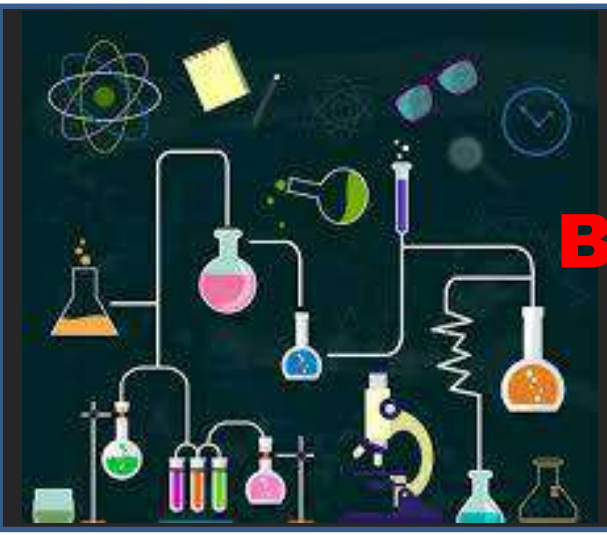
Falta de conhecimento das tecnologias para o compartilhamento

Resultados de experimentos que não deram certo

Problemas éticos, de privacidade e legais
"A maior barreira para o compartilhamento de dados de pesquisa são os temores dos pesquisadores **em relação às questões legais e o mau uso de seus dados**"

Baixo impacto na carreira do pesquisador.

Interesse econômico



BARREIRAS ORGANIZACIONAIS

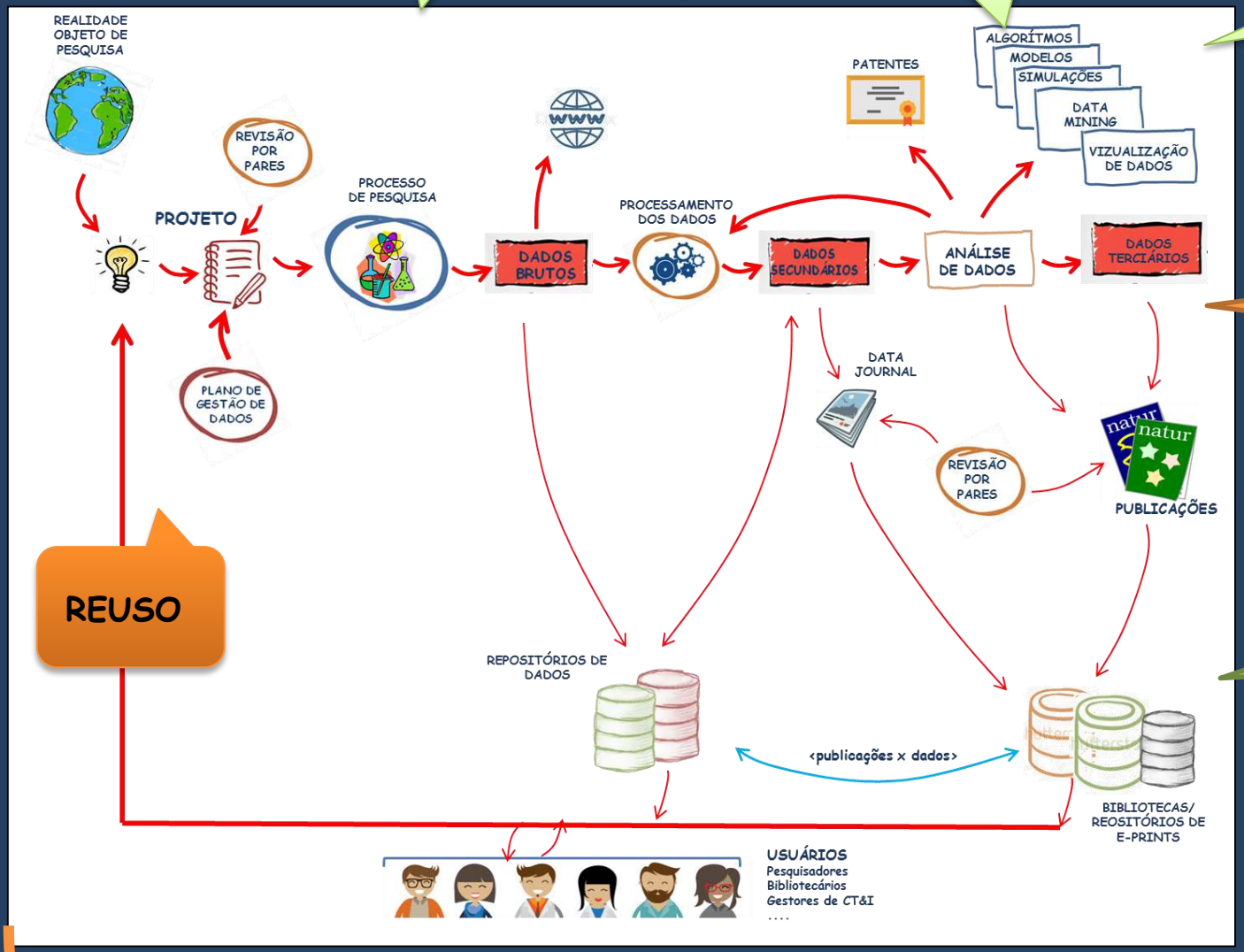
POLÍTICA DE GESTÃO DE DADOS
SERVIÇOS
BALCÃO DE REFERÊNCIA PARA DADOS
TREINAMENTO
FERRAMENTAS DE SOFTWARE
EXPERTISE EM GESTÃO DE DADOS

CAPTURA DE DADOS

LIMPEZA DOS DADOS

ANÁLISES E RESULTADOS

PROCESSOS COMPUTACIONAIS



METADADOS, DOCUMENTAÇÃO, VERSIONAMENTO

INTEGRAÇÃO COM O SISTEMA DE PUBLICAÇÃO

REUSO

ARQUIVAMENTO PRESERVAÇÃO

INTEROPERABILIDAE

DESCOBERTA & ACESSO

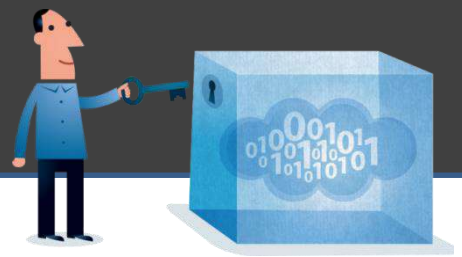
POLÍTICAS – SUSTENTABILIDADE – CONFORMIDADE LEGAL E ÉTICA

DEFINIÇÃO DE POLÍTICAS

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

Documento formal que estabelece um compromisso de como os dados serão tratados durante todo o **desenvolvimento do projeto**, e também após a sua **conclusão**.

Descreve o **ciclo de vida de gestão** para todos os dados que serão coletados,processados ou gerados por um projeto de pesquisa.

Quais as **metodologias e padrões** que serão utilizados nesses processos;

sob que condições esses dados **serão compartilhados** e/ou **tornados abertos** para a comunidade de pesquisa; e como eles serão **curados e preservados**.

PGD espelha uma situação dinâmica, é necessário observar que ele **não é um documento fixo no tempo**, ao contrário, ele se desenvolve e ganha mais precisão e solidez durante o tempo de vida do projeto

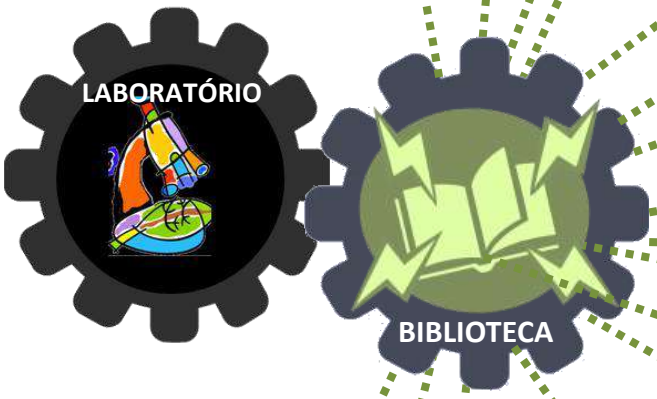
DEPOSITAR & COMPARTILHAR

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**.



TREINAMENTO EM GESTÃO DE DADOS
Formatos, padrões, licenças;
Formas de compartilhamento;
Arquivamento pré-repositório

CITAÇÃO DOS DADOS
Atribuição de DOI
Identificação das **versões**
Padrão de referência
Créditos aos pesquisadores

METADADOS
Organização, classificação, descrição dos dados.
Desenvolvimento de padrões
Metadados disciplinares
Formatos recomendados

PLANO DE GESTÃO DE DADOS
Elaboração, consultoria
Fermentas, manuais

CONTEXTUALIZAÇÃO DOS DADOS
Documentação
Cadernos de Laboratórios
Anotações
Link com artigos

SUSTENTABILIDADE
Recursos, projetos
Parcerias

CICLO DE VIDA DOS DADOS
Conhecer o fluxo de geração de dados dos laboratórios da instituição;

LICENÇAS E PRIVACIDADE
Licenças apropriadas
Proteção de informações confidencias
Tratamento de informações sensíveis

PRESERVAÇÃO DIGITAL
Assegurar que os dados estejam acessíveis agora e no futuro
PREMIS
OAIS
Certificação

ACESSO A DADOS
Escolha dos repositórios para acesso e serviços

COMPARTILHAMENTO
Apoiar o compartilhamento de dados

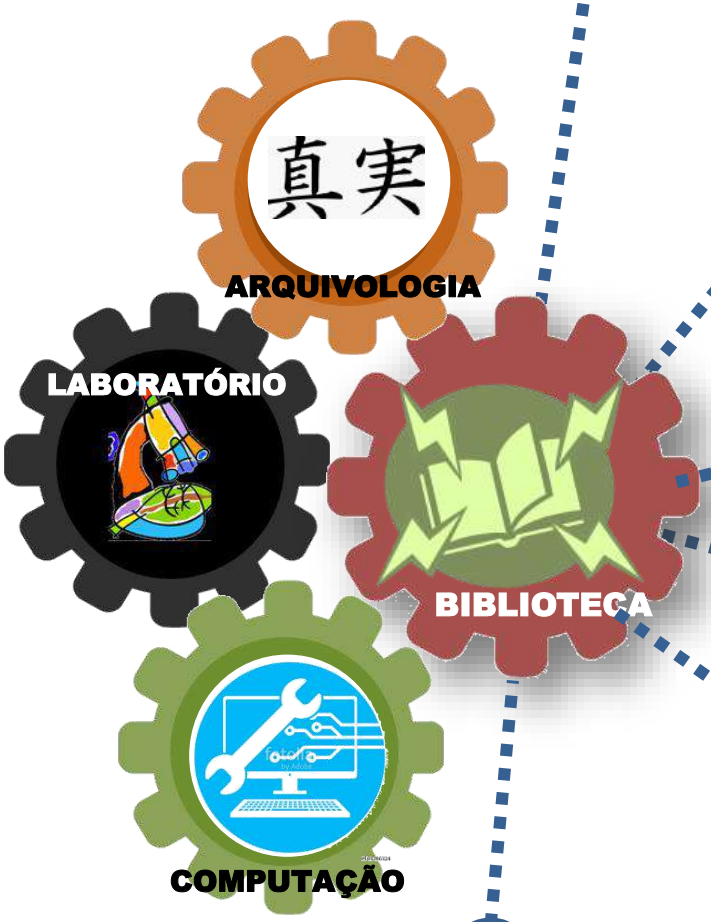
POLÍTICA DE DADOS
Contribuir na elaboração da politica da instituição

DESENVOLVIMENTO E MANUTENÇÃO DE RESPOSITÓRIOS
Política
Modelo de dados
Plataforma tecnológica
Parcerias

CURADORIA
Adicionar valor
Divulgação
Reuso

PUBLICAÇÃO DOS DADOS
Escolha dos repositórios para publicação;
Diretórios;
Apoiar a publicação dos dados

SERVIÇOS INOVADORES
Desenvolvimento de serviços que atendam as novas demandas



PROVENIÊNCIA, AUTENTICIDADE E INTEGRIDADE

Sistemas de arquivamento confiável; certificação, OAIS, eARQ

VISUALIZAÇÃO

MODELAGEM

ANÁLISE

MINERAÇÃO

PROCESSAMENTO

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

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**

COMPARTILHAMENTO RESTRITO

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

MELHORES FORMATOS E 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**.

STORAGE

Sistemas de armazenamento seguro

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 apóia 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

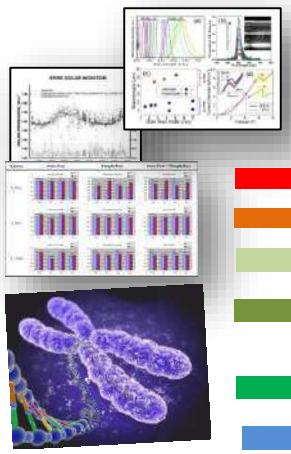
GERENTE DE DADOS – **tecnologista da informação** responsável pela manutenção e operação das bases de dados, segurança e armazenamento dos dados: backups, checagem de integridade, etc.

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

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

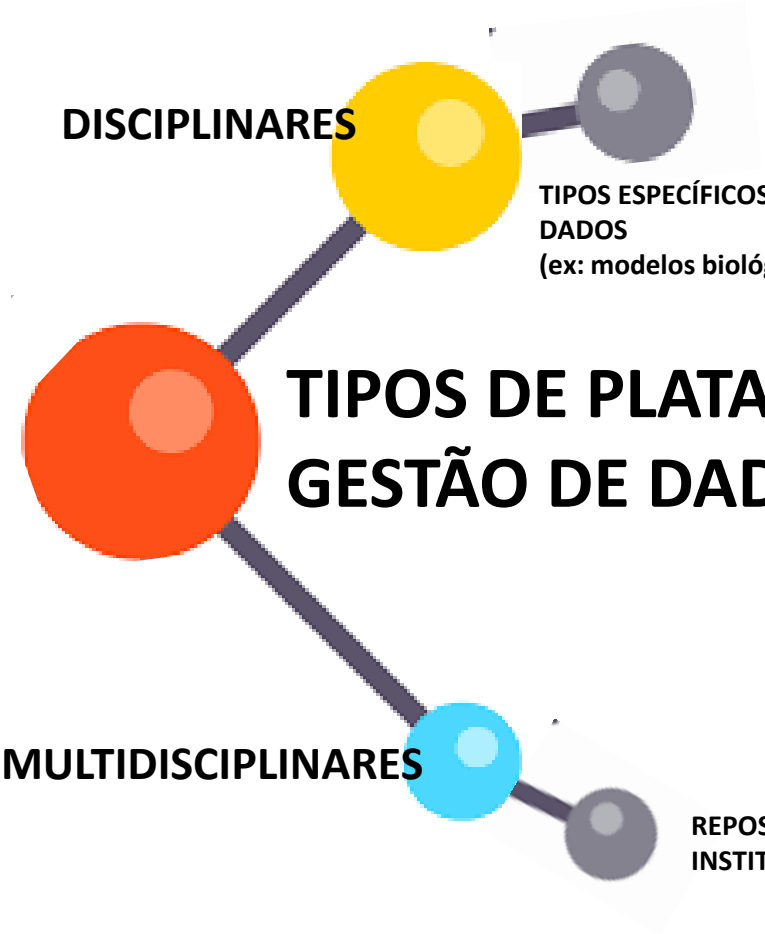
SUBMETÊ-LOS A UM **PERIÓDICO** PARA APOIAR UMA PUBLICAÇÃO

PUBLICANDO-OS EM UM **REPOSITÓRIO MULTIDICIPLINAR**

PUBLICANDO-OS NO **REPOSITÓRIO INSTITUCIONAL**

DEPOSITANDO-OS NUM REPOSITÓRIO DE DADOS DISCIPLINAR

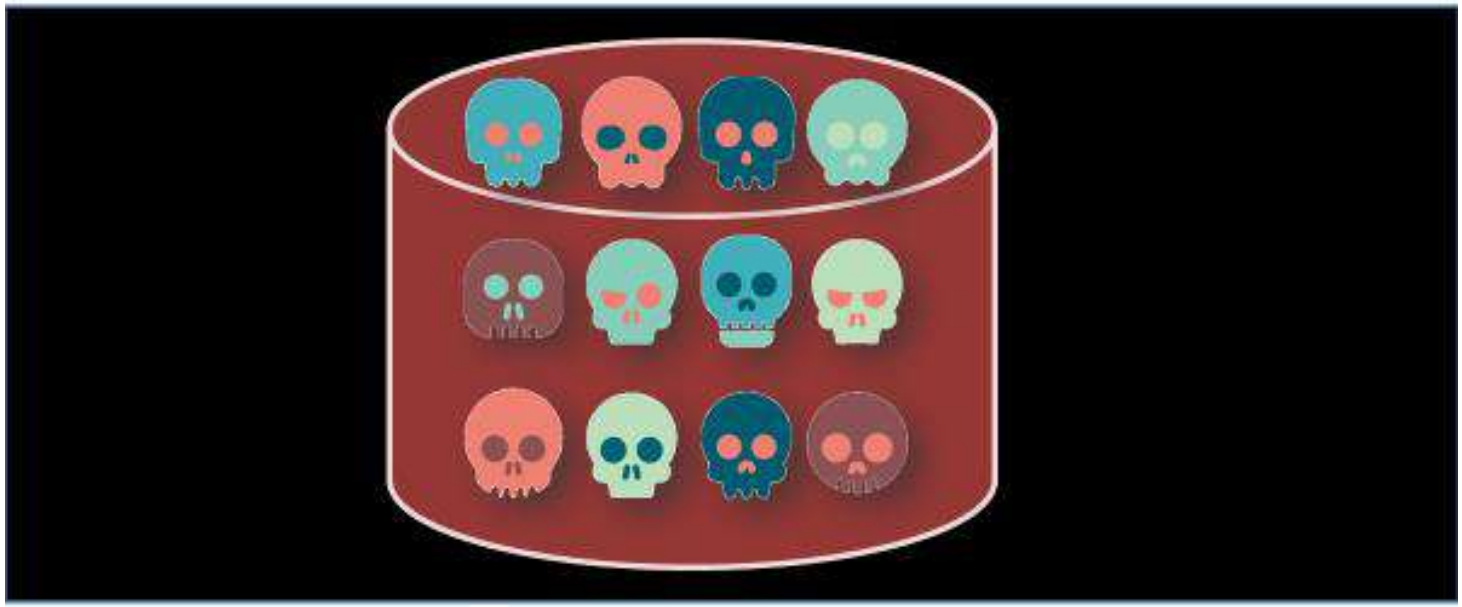
TIPOS DE PLATAFORMAS DE GESTÃO DE DADOS



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 desenvolvida 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 resultados científicos

REPOSITÓRIO DE DADOS



CEMITÉRIO DE DADOS ?

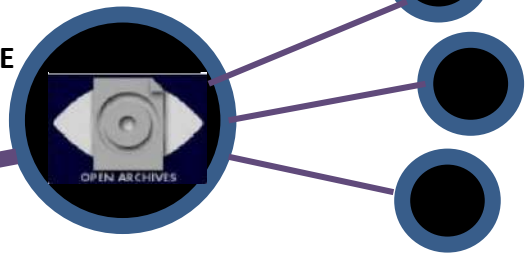
KEEP CALM

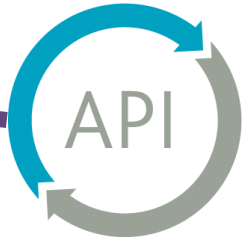
E ACREDITE

NOS REPOSITÓRIOS DIGITAIS CONFIÁVEIS

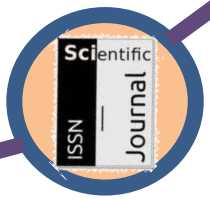
A INTEROPERABILIDADE DE SISTEMAS É COMPREENDIDA AQUI COMO A CAPACIDADE DAS PLATAFORMAS DE REPOSITÓRIOS DE DADOS INTERCAMBIAREM INFORMAÇÕES – **DADOS E METADADOS** - COM SISTEMAS EXTERNOS DE FORMA HARMÔNICA E INTEGRADA E COM PROPÓSITOS ESPECÍFICOS.

PROTOCOLO OAI-PMH PARA COLETA AUTOMÁTICA DE METADADOS





PROGRAMAS PARA EXPOR METADADOS

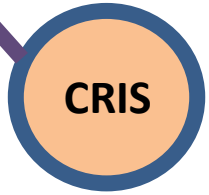


SISTEMAS DE PUBLICAÇÕES DE PERIÓDICOS

INTEGRAÇÃO COM OUTROS SISTEMAS

LINKING - USO DO LINKED DATA

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



SISTEMA DE GESTÃO DE PESQUISA

SISTEMA DE REPOSITÓRIO CONFIÁVEL

ARCHIVEMÁTICA

- VISIBILIDADE
  - disponibilidade *on-line*
  - descoberta
  - acesso
- COMPARTILHAMENTO/REUSO/INTERAÇAO
- 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


POLÍTICA INSTITUCIONAL DE GESTÃO DE DADOS DE PESQUISA
DISCIPLINA
CONTEÚDOS
FORMATO
VOLUME
DADOS SENSÍVEIS
DEPÓSITO
ACESSO
REUSO/COMPARTILHAMENTO
PERÍODO DE EMBARGO
SEGURANÇA
PLANO DE SUCESSÃO
PRESERVAÇÃO
CURADORIA
COPYRIGHT
LICENÇAS
TERMOS DE USO
PUBLICAÇÃO
CUSTOS
PUBLICAÇÃO
ACESSO
A **POLÍTICA DE UMA PALTAFORMA DE GESTÃO delineia os compromissos que a instituição se obriga em relação aos seus principais *stakeholders*** – pesquisadores, curadores, consumidores, financiadores, coletores de metadados, etc. - e com relação ao ciclo de vida das coleções de dados que estão sendo gerenciados.

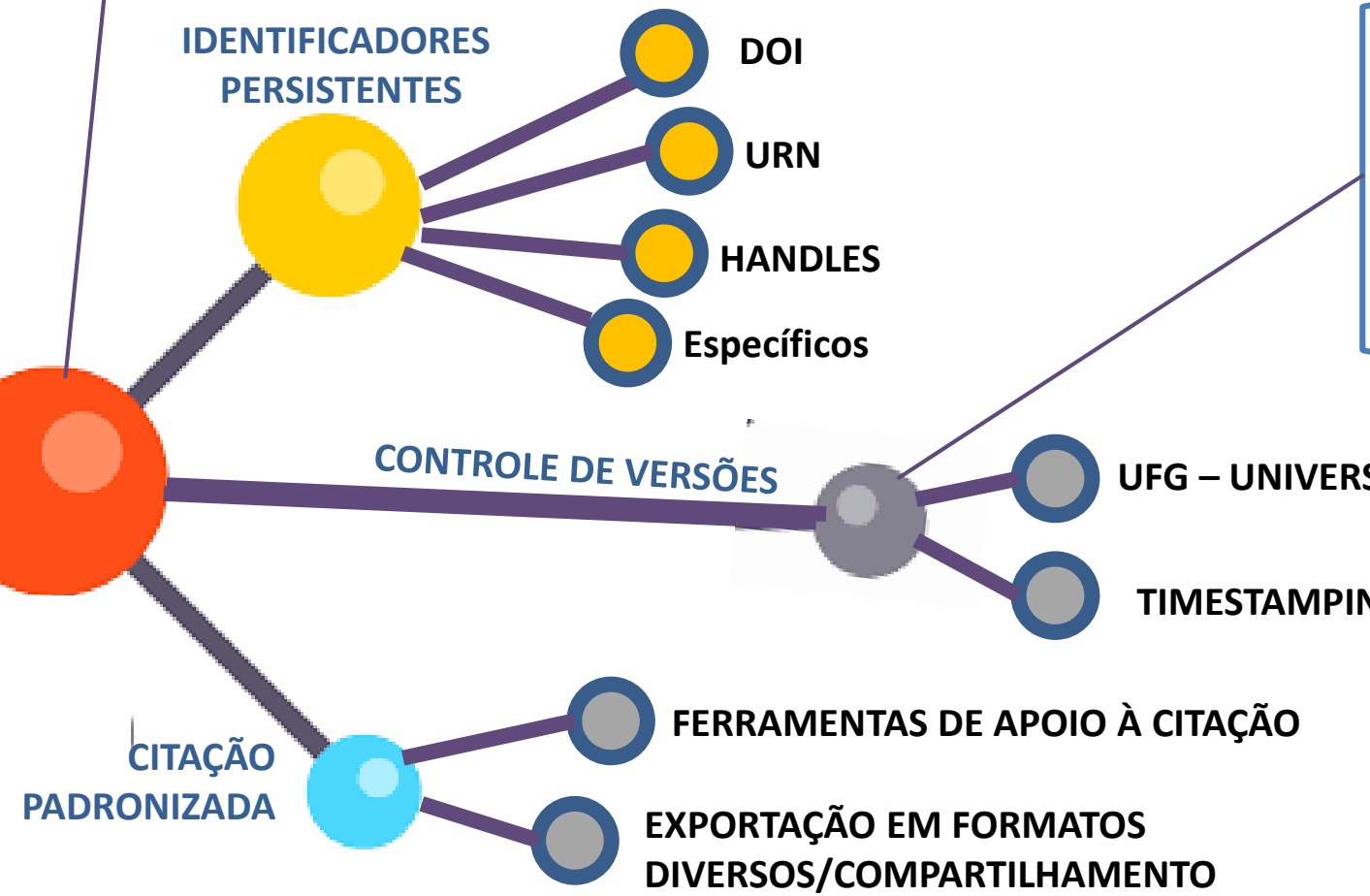
O sucesso dos novos serviços de informação para a pesquisa está relacionado à sua capacidade de dar apoio às **práticas e culturas** das comunidades científicas da instituição.

CONTORNANDO A INVISIBILIDADE DA CAUDA LONGA


RECONHECIMENTO/RECOMPENSA
CITAÇÃO/FATOR DE IMPACTO/PUBLICAÇÃO
INCENTIVOS
POLÍTICAS
EXIGÊNCIAS PARA FINANCIAMENTO DE PROJETOS
EXIGÊNCIAS DO EDITORES
PESQUISADORS COMO PRODUTORES E USUÁRIOS
CAPACITAÇÃO
BIBLIOTECARIOS DE DADOS
CIENTISTAS DE DADOS
INFRAESTRUTURA
ÊNFASE DISCIPLINAR
FINANCIAMENTO
ORIENTAÇÃO POR PROJETOS/FINANCIMENTO PERMANENTE
CUSTO-BENEFÍCIO (

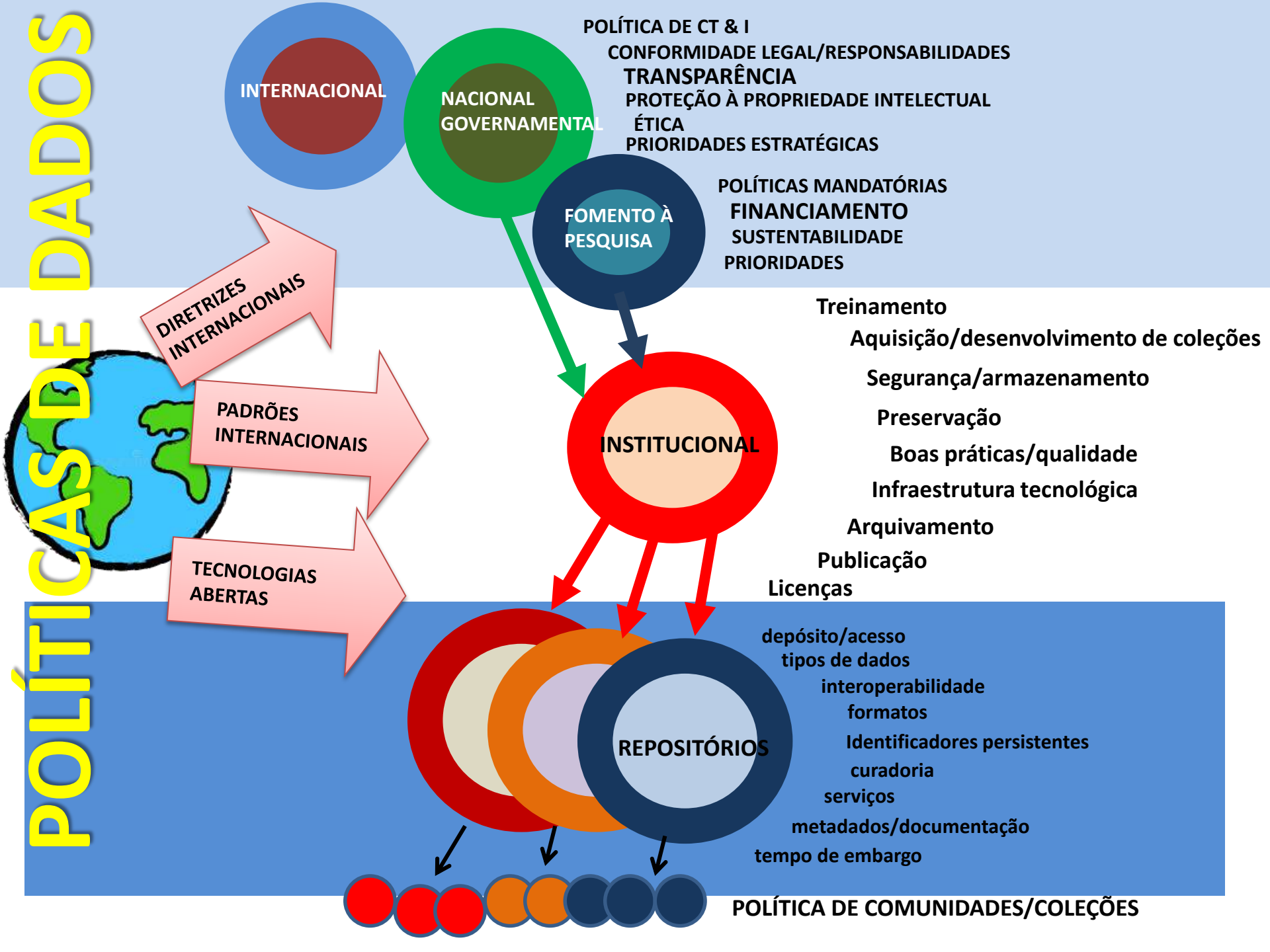
REFERÊNCIA

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*.



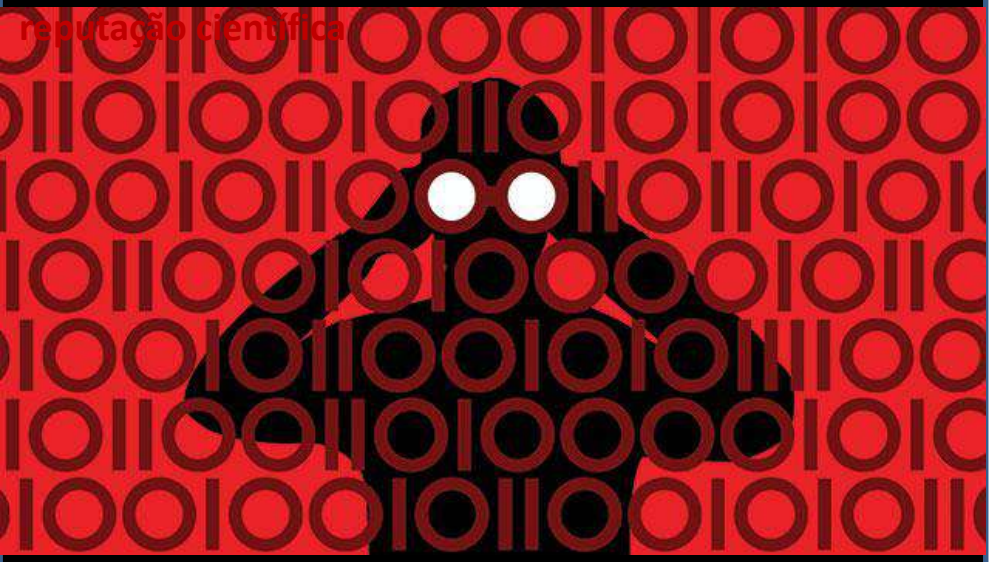
O controle de versões é um processo importante para o fundamento da reprodutibilidade da pesquisa, para a integridade da referência às coleções de dados e para proveniência dos seus conteúdos. Isto por que as coleções de dados podem evoluir no tempo por vários motivos


RECONHECIMENTO/RECOMPENSA
CITAÇÃO/FATOR DE IMPACTO/PUBLICAÇÃO
INCENTIVOS
POLÍTICAS
EXIGÊNCIAS PARA FINANCIAMENTO DE PROJETOS
EXIGÊNCIAS DO EDITORES
PESQUISADORS COMO PRODUTORES E USUÁRIOS
CAPACITAÇÃO
BIBLIOTECARIOS DE DADOS
CIENTISTAS DE DADOS
INFRAESTRUTURA
ÊNFASE DISCIPLINAR
FINANCIAMENTO
ORIENTAÇÃO POR PROJETOS/FINANCIMENTO PERMANENTE
CUSTO-BENEFÍCIO


POLÍTICAS DE DADOS
INTERNACIONAL
NACIONAL GOVERNAMENTAL
FOMENTO À PESQUISA
POLÍTICA DE CT & I
CONFORMIDADE LEGAL/RESPONSABILIDADES
TRANSPARÊNCIA
PROTEÇÃO À PROPRIEDADE INTELECTUAL
ÉTICA
PRIORIDADES ESTRATÉGICAS
POLÍTICAS MANDATÓRIAS
FINANCIAMENTO
SUSTENTABILIDADE
PRIORIDADES
DIRETRIZES INTERNACIONAIS
PADRÕES INTERNACIONAIS
TECNOLOGIAS ABERTAS
INSTITUCIONAL
Treinamento
Aquisição/desenvolvimento de coleções
Segurança/armazenamento
Preservação
Boas práticas/qualidade
Infraestrutura tecnológica
Arquivamento
Publicação
Licenças
REPOSITÓRIOS
depósito/acesso
tipos de dados
interoperabilidade
formatos
Identificadores persistentes
curadoria
serviços
metadados/documentação
tempo de embargo
POLÍTICA DE COMUNIDADES/COLEÇÕES


RECONHECIMENTO/RECOMPENSA
CITAÇÃO/FATOR DE IMPACTO/PUBLICAÇÃO
INCENTIVOS
POLÍTICAS
EXIGÊNCIAS PARA FINANCIAMENTO DE PROJETOS
EXIGÊNCIAS DO EDITORES
PESQUISADORS COMO PRODUTORES E USUÁRIOS
CAPACITAÇÃO
BIBLIOTECARIOS DE DADOS
CIENTISTAS DE DADOS
INFRAESTRUTURA
ÊNFASE DISCIPLINAR
FINANCIAMENTO
ORIENTAÇÃO POR PROJETOS/FINANCIMENTO PERMANENTE
CUSTO-BENEFÍCIO (

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



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

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

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

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

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

DATA**paper** **journal**

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

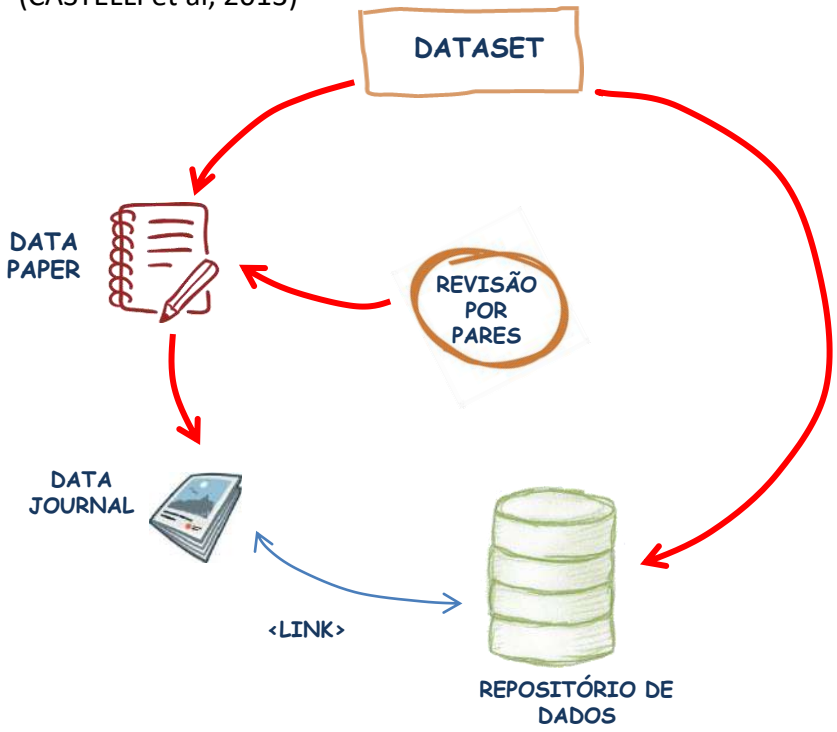
Oferecer uma publicação que **pode ser citada** e que dá **credito ao autor** e o outros envolvidos no processo;

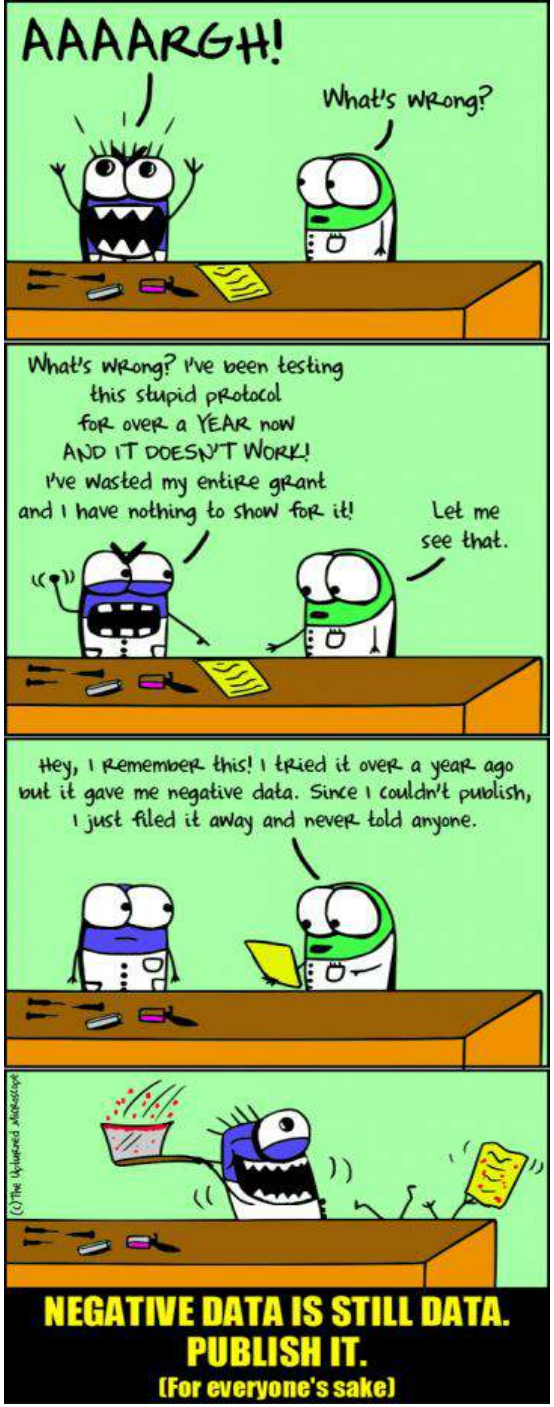
Assegura que os dados estejam **documentados para o reuso**;

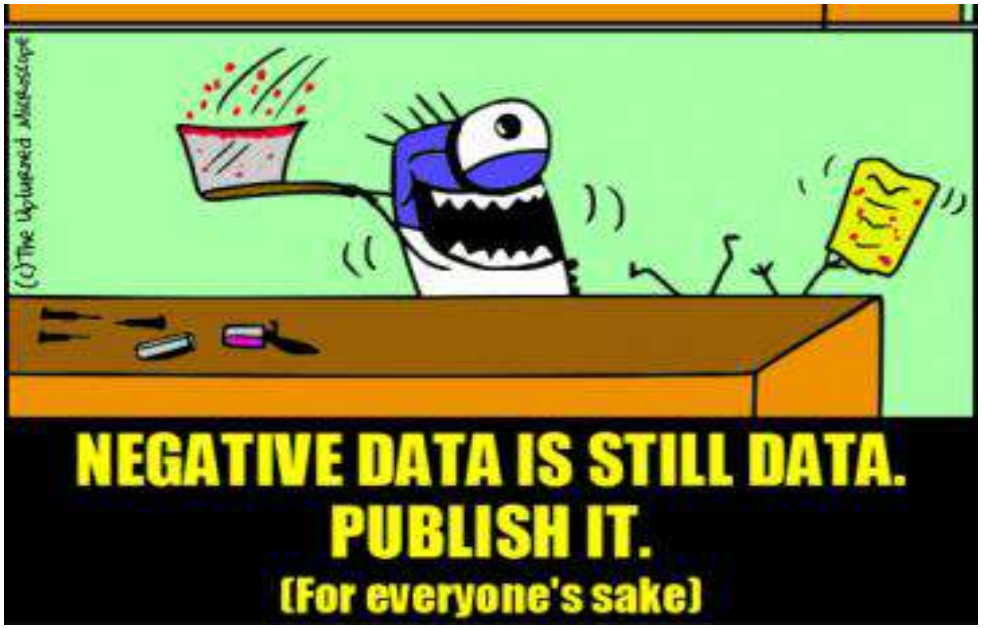
Aumenta a visibilidade dos dados na comunidade científica

“

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)







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/

JOURNAL OF NEGATIVE RESULTS
IN BIOMEDICINE

Journal
of Negative & No Positive Results

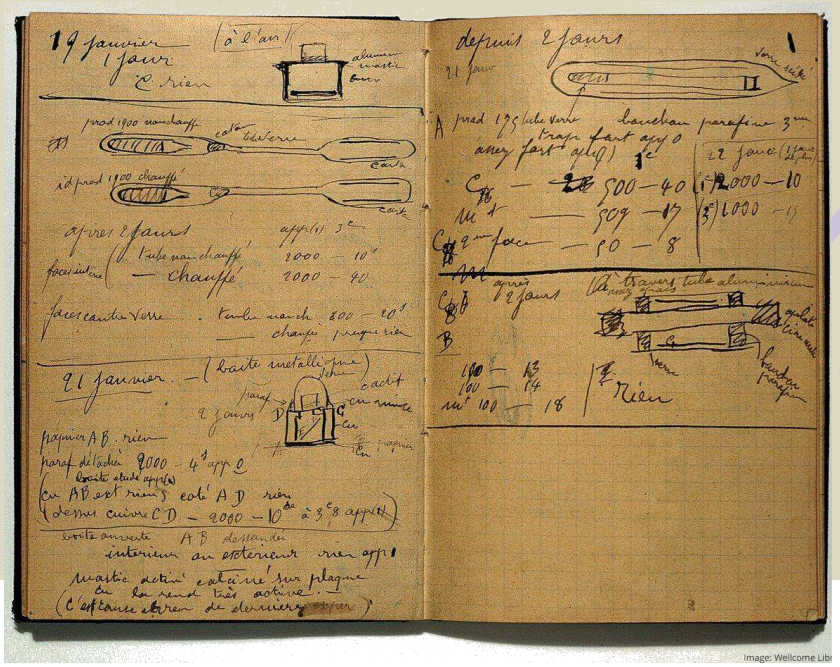
Journal of
Pharmaceutical
Negative Results

NEGATIVE RESULTS
SCIENTIFIC JOURNAL

PUBLICAÇÕES AMPLIADAS

Caderno de laboratório

O caderno de laboratório é uma ferramenta de organização e de memória que serve de registro primário da pesquisa científica e das atividades relacionadas. O caderno de pesquisa **registra as hipóteses, experimentos e análises iniciais ou interpretações dos experimentos;** serve também como o **registro legal da propriedade intelectual das ideias e dos resultados obtidos pela pesquisa** (SCHNELL, 2015).



Cadernos abertos
disponibilização dos dados acontece em tempo real, à medida que a pesquisa vai sendo feita

Sistemas complexos
integração com os equipamentos do lab

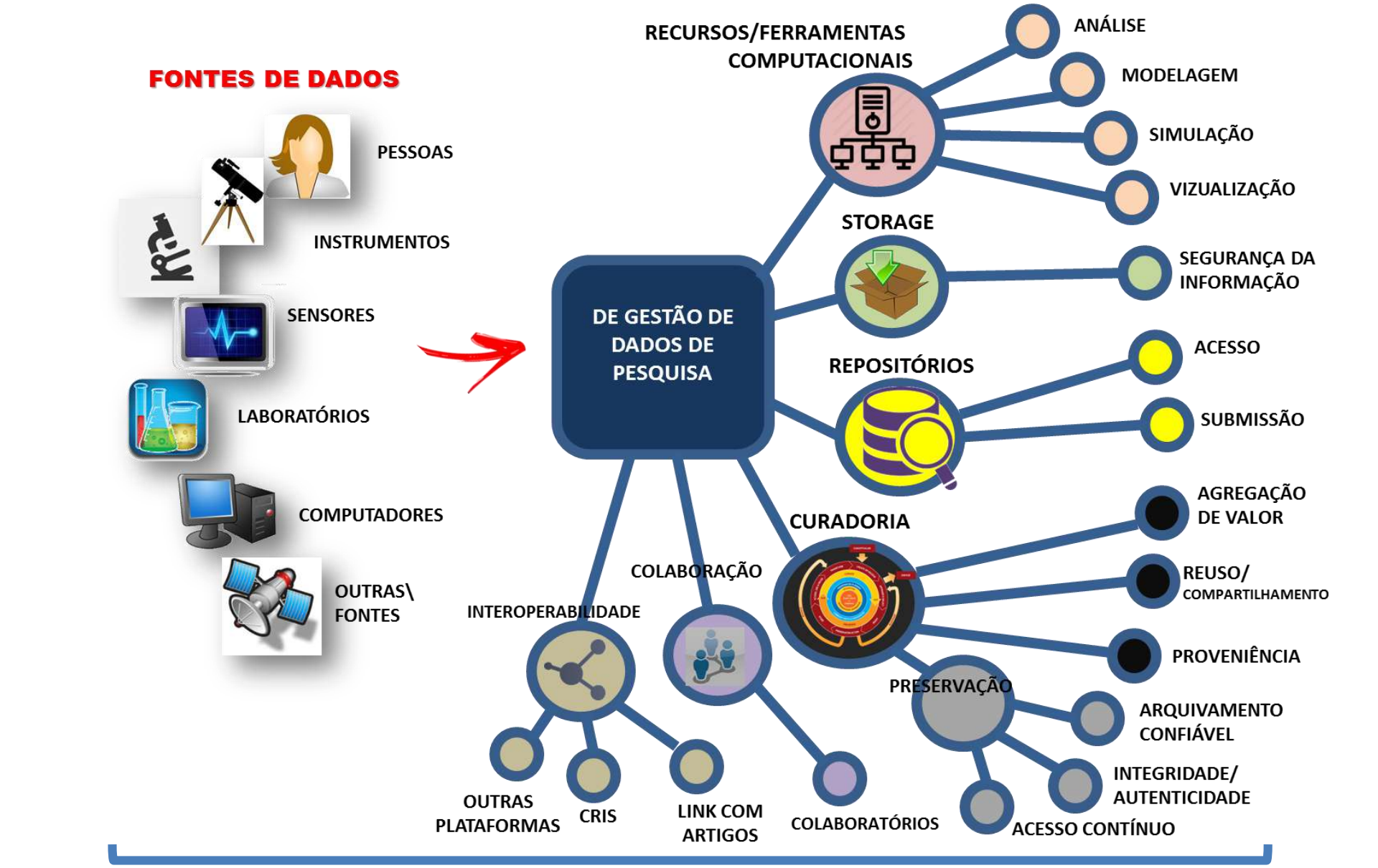
Cadernos Eletrônicos
auditoria | certificação

Cadernos convencionais

EXISTEM CÓDIGOS INTERNACIONAIS, NACIONAIS E INSTITUCIONAIS QUE DETALHAM AS ESPECIFICAÇÕES E GUARDA DESTES CADERNOS


RECONHECIMENTO/RECOMPENSA
CITAÇÃO/FATOR DE IMPACTO/PUBLICAÇÃO
INCENTIVOS
POLÍTICAS
EXIGÊNCIAS PARA FINANCIAMENTO DE PROJETOS
EXIGÊNCIAS DO EDITORES
PESQUISADORS COMO PRODUTORES E USUÁRIOS
CAPACITAÇÃO
BIBLIOTECARIOS DE DADOS
CIENTISTAS DE DADOS
INFRAESTRUTURA
ÊNFASE DISCIPLINAR
FINANCIAMENTO
ORIENTAÇÃO POR PROJETOS/FINANCIMENTO PERMANENTE
CUSTO-BENEFÍCIO (inovação, eficiência, efetividade)

CIBERINFRAESTRUTURA DE DADOS DE PESQUISA



A GUISA DE CONCLUSÃO

O COMPARTILHAMENTO DE DADOS DA COMO PARTE DA CULTURA ACADÊMICA E A GESTÃO COMO PARTE DA PROFISSÃO DE PESQUISADOR

A BIBLIOTECA UNIVERSITÁRIA COMO PROTAGONISTA NA GESTÃO E INTEGRAÇÃO DE DADOS DA CAUDA LONGA

PRETEXTO PARA APROXIMAR A BIBLIOTECA DOS LABORATÓRIOS E DOS INFORMÁTICOS

NOVOS TEMAS DE PESQUISA