



# PRESERVAÇÃO DE DADOS DE PESQUISA

**Miguel Ángel Márdero Arellano**

**MCTIC/IBICT**



**Rede CARINIANA**

# REPOSITÓRIOS DE DADOS E A PRESERVAÇÃO DIGITAL

Dados de pesquisa

Gestão de dados de pesquisa

Repositórios de dados de pesquisa

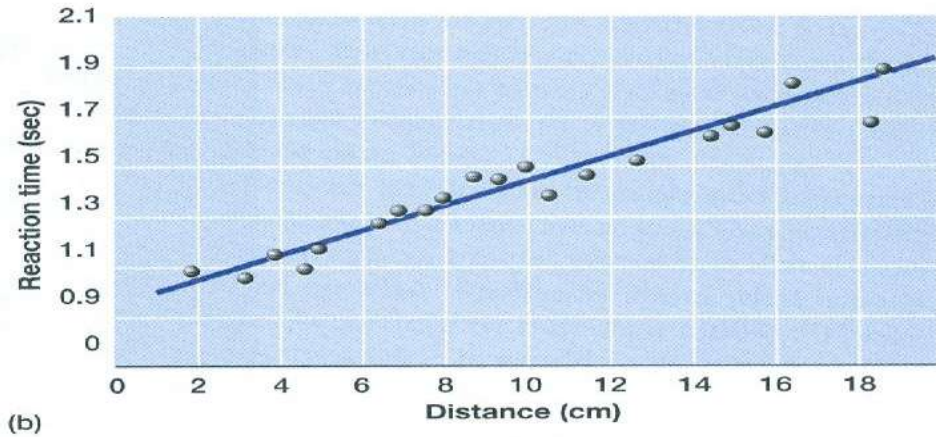
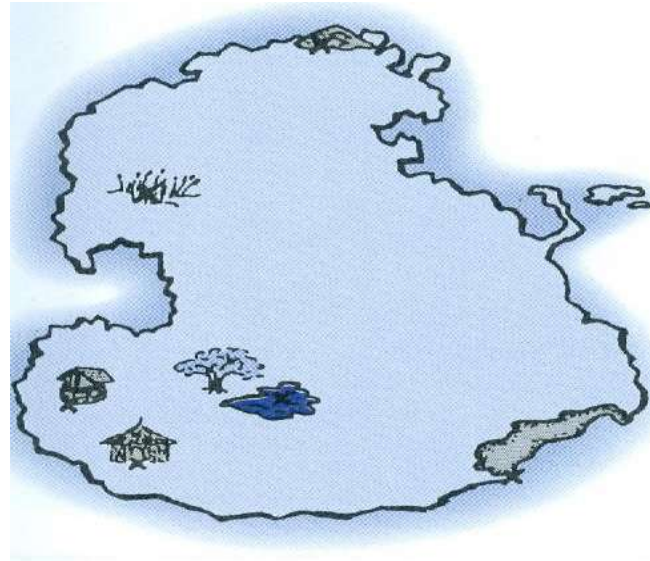
Curadoria de dados de pesquisa

Preservação de dados de pesquisa

Preservação de dados de pesquisa no Brasil

Preservação de dados de pesquisa em Rede

# DADOS DE PESQUISA



■ **Figure 9.5** (a) Island used in Kosslyn et al.'s (1978) image-scanning experiment. Participants mentally traveled between the various locations on the island. (b) Results of the island experiment. (S. M. Kosslyn, T. Ball, & B. J. Reiser, "Visual Images Preserve Metric Spatial Information: Evidence From Studies of Image Scanning." From *Journal of Experimental Psychology: Human Perception and Performance*, 4, pp. 47–60, Figs. 2 & 3. Copyright © 1978 with permission from the American Psychological Association.)

# DADOS DE PESQUISA



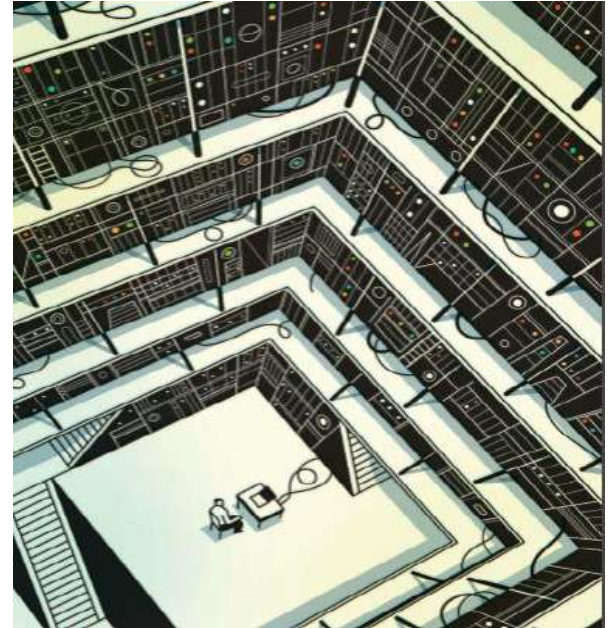
"...qualquer informação que possa ser armazenada em formato digital, incluindo texto, números, imagens, vídeos ou filmes, áudio, software, algoritmos, equações, animações, modelos, simulações, etc."

(National Science Board, Long-lived digital data collections, 2005)



# A IMPORTÂNCIA DO REGISTRO DOS DADOS DA PESQUISA

- ❑ O registro científico dos dados brutos da pesquisa precisa estar voltado para sua criação e reprodução.
- ❑ Os dados empíricos, estatísticos e computacionais são a centrais para o método científico e a busca pela ausência de erro sua motivação.



# A IMPORTÂNCIA DO REGISTRO DOS DADOS DA PESQUISA

- ▣ A simulação de dados em grande escala movimenta a ciência computacional atual, reunindo um conjunto de instruções e de dados que determinarão parte dos resultados.
- ▣ A integridade na pesquisa é garantida pela quantidade de informação necessária para o conhecimento pessoal. Um dos componentes dessa garantia de integridade são os metadados de preservação, os quais proveem informação para dar suporte ao processo permanência dos registros.

# QUANTOS DADOS DE PESQUISA JÁ PERDEU SUA INSTITUIÇÃO?



- ❑ Dados de centenas de estudos guardados em disquetes, fitas VHS, fitas cassetes, CDs e DVDs.
- ❑ O fato de alguns desses dados já estarem na Web não garante sua continuidade e integridade, é inegável que as URLs deixam de existir por vários motivos.



# DADOS DE PESQUISA EM FORMATOS PROPRIETÁRIOS

- ❑ SPSS, Excel, SAS, Word, Stata, ASCII, Access, Matlab, VHS, File Marker Pro, JPGs, etc.
- ❑ Mídias de armazenamento usadas comumente pelos pesquisadores como os hard drives pessoais.
- ❑ O problema não é só perder os arquivos pessoais, mas perder aqueles dos quais somos institucionalmente responsáveis.






# DIFICULDADES DA PRESERVAÇÃO DIGITAL DE DADOS DE PESQUISA NO BRASIL



The screenshot shows the top section of the Lattes website. On the left is the logo for 'DIRETÓRIO DOS GRUPOS DE PESQUISA NO BRASIL Lattes'. In the center is the CNPq logo and the flag of Brazil. On the right are social media icons (Facebook, Twitter, YouTube, LinkedIn, Instagram) and a search bar with the text 'Pesquisar...'. Below the header is a navigation menu with five items: 'O Diretório', 'Bases de dados', 'Censos', 'Histórico', and 'Wiki'.

O Diretório ▾      Bases de dados ▾      Censos ▾      Histórico ▾      Wiki



A photograph showing several scientists in white lab coats working in a laboratory. They are standing at a bench with various pieces of equipment, including pipettes and beakers. The scene is brightly lit, likely from large windows in the background.

# O MAIOR RISCO PARA O AVANÇO DA CIÊNCIA

Quando os dados não são arquivados de forma rápida e não há um **planejamento** inicial de guarda e proteção tornam-se um desafio sua preservação e acesso.



# A FALTA DA DOCUMENTAÇÃO E A REPRODUTIVIDADE DOS DADOS

- ❑ Um sistema de gestão de dados de pesquisa garante que esses dados, sua codificação e fluxo poderão estar disponíveis e serem citados. Sua documentação é um dos princípios que destaca a reprodutividade dos dados.
- ❑ Os pesquisadores são responsáveis por decidir quando e em quais termos os dados de pesquisa podem ser acessados.

# A FALTA DA DOCUMENTAÇÃO E A REPRODUTIVIDADE DOS DADOS

- ❑ Eles devem produzir seus planos de gestão de dados para cada projeto de pesquisa que gerará um conjunto de dados.
- ❑ Cada vez mais os projetos científicos se tornam intensivamente computacionais e por isso mais transparentes.
- ❑ As opções de customização das interfaces devem atender às demandas de cada projeto, descrevendo as especificações de cada *template*, conjunto de dados e metadados.



# A RESPONSABILIDADE DAS INSTITUIÇÕES DE PESQUISA

- ❑ **Comunicar** à comunidade científica sobre as ações que devem ser tomadas desde o início do ciclo de vida dos dados.
- ❑ Avaliar se os materiais são publicáveis, úteis e se as expensas na sua preservação a longo prazo é justificável.



# A GESTÃO DE DADOS DE PESQUISA + O CICLO DA PRESERVAÇÃO DIGITAL



Refere-se ao armazenamento, acesso e preservação de dados produzidos a partir de uma determinada pesquisa.

Toda organização de pesquisa deve ter uma política de dados de pesquisa, que inclua um esquema de gestão de dados.

# A RESPONSABILIDADE DAS INSTITUIÇÕES DE FOMENTO



Os financiadores das pesquisas também devem ter uma **política de dados de pesquisas**, estipulando as obrigações que um pesquisador deve cumprir como condição do financiamento.

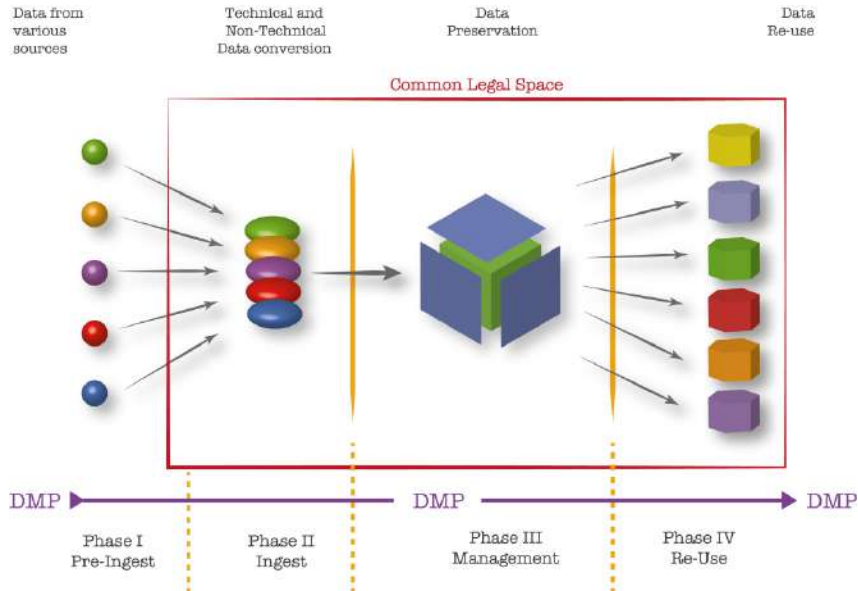
# A POLÍTICA DE RETENÇÃO DOS DADOS INSTITUCIONAL



Um dos aspectos que envolvem a preservação dos dados de pesquisa é incluir na política de gestão que existam mecanismos para garantir que quando o pesquisador sair da universidade por aposentadoria ou mudança, os dados com valor de uso a longo prazo estejam depositados num repositório de dados institucional para seu arquivamento.



# A POLÍTICA DE RETENÇÃO DOS DADOS INSTITUCIONAL



As práticas de gerenciamento de dados cobrem todo o ciclo de vida dos dados, desde o planejamento da investigação até a realização, bem como de fazer backup de dados, uma vez que eles são utilizados para a preservação de dados a longo prazo entregues após a conclusão da investigação.

# A RESPONSABILIDADE DO PESQUISADOR



Um plano de gestão de dados é um documento formal que descreve como o pesquisador irá lidar com seus dados durante a pesquisa e após o projeto ser concluído.

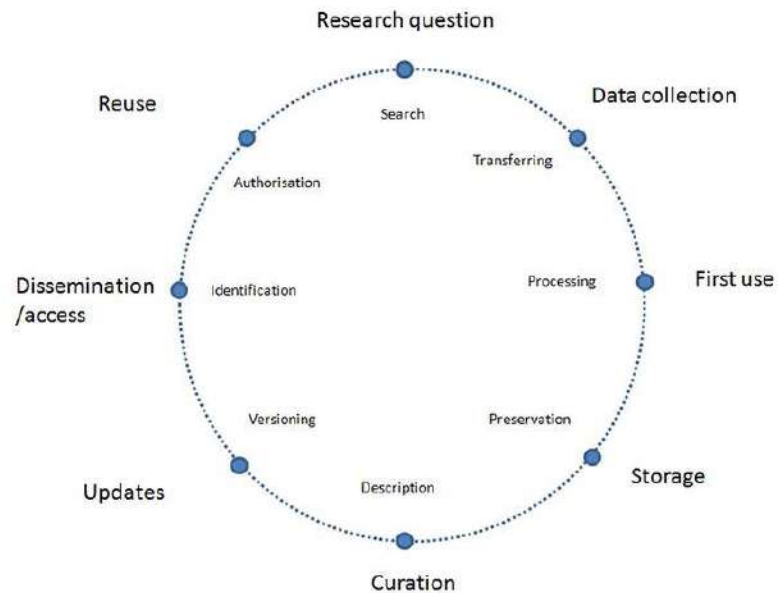
# A RESPONSABILIDADE DO PESQUISADOR



O objetivo de um plano de gestão de dados é considerar os muitos aspectos do gerenciamento de dados, geração de metadados, preservação de dados e análise antes do início do projeto, garantindo que os dados sejam bem gerenciados no presente e preparados para preservação no futuro.

# A RESPONSABILIDADE DO PESQUISADOR

- ❑ Nunca no final do ciclo, quando o projeto terminou, levando ao fracasso sua preservação.
- ❑ A mais importante mudança é estabelecer uma política que obrigue ao compartilhamento dos dados.
- ❑ O maior desafio é fazer a comunidade científica pensar no alto valor da preservação.





# PRESERVAÇÃO DE DADOS DE PESQUISA



# SERVIÇOS DE PRESERVAÇÃO DE DADOS

- ❑ ICPSR da University of Michigan em funcionamento desde 1960 e que é considerado o maior arquivo de dados das ciências sociais do mundo.
- ❑ Na University of North Carolina o Odum Institute mantem o arquivo mais antigo de dados das ciências sociais fundado em 1920.
- ❑ The National Archives and Records Administration lidera as iniciativas de preservação de acervos de dados especializados governamentais.
- ❑ Na Universidade de Harvard existe a primeira aliança nacional de preservação de dados de pesquisa, Data-PASS, especializada no desenvolvimento de soluções e políticas de gestão de dados científicos (ALTMAN *et al.* 2009).

# O SERVIÇO DA CANADIAN ASSOCIATION OF RESEARCH LIBRARIES

## Portage Network Research Data Management Framework

Research Data Lifecycle



Portage Network of Expertise



RDM Tools & Platforms



Portage Tools



# SERVIÇOS DE PRESERVAÇÃO DE DADOS

- ❑ Na América Latina existem instituições emergentes que requerem dos seus pesquisadores que seus dados sejam publicados em repositórios abertos, caso esses projetos recebam algum tipo de financiamento.
- ❑ No Brasil a agência de financiamento do estado de São Paulo FAPESP pede aos pesquisadores que solicitam auxílio financeiro pelo fundo do programa E-Science que apresentem seus planos de gestão de dados de pesquisa.
- ❑ Mas ainda faltam em iniciativas como essa a integração do processo de preservação em todo o ciclo de vida dos dados.
- ❑ Qualquer custo de armazenamento e gestão deve ser explicitado nos pedidos de financiamento, incluindo o local onde os dados serão disponibilizados e aplicadas as técnicas de curadoria, por anos após do termino do projeto.



# O SERVIÇO DE PRESERVAÇÃO DE DADOS DO IBICT

O Ibict é uma instituição de pesquisa da América Latina que realiza esforços para contribuir com o compartilhamento de dados e seu reuso. Essas atividades se iniciaram no Instituto com a implementação da Biblioteca Digital de Teses e Dissertações, a segunda maior biblioteca digital de seu tipo, e com a promoção da plataforma Open Journal Systems (OJS), sendo atualmente o Brasil o país com maior número de instalações dessa plataforma.



# O SERVIÇO DE PRESERVAÇÃO DE DADOS DO IBICT

- ❑ No serviço de preservação e dados de pesquisa a adoção de um sistema consorciado e distribuído de arquivamento parte da ideia de que nenhuma instituição pode manter e preservar isoladamente seus materiais por longos períodos de tempo.
- ❑ As parcerias da Rede são de colaboração técnica e científica, que seguem normas e padrões internacionais de seleção, aquisição, arquivamento e acesso compartilhados.



# A REDE CARINIANA

- ❑ O projeto do serviço de preservação digital de dados de pesquisa da Cariniana está direcionado à replicação de cópias dos arquivos dos dados armazenados nos repositórios das instituições parceiras.
- ❑ Como uma atividade de preservação distribuída, a guarda segura de pelo menos quatro cópias dos documentos digitais em instituições geograficamente distantes faz o sistema de arquivamento com o LOCKSS (Lots of Copies Keep Stuff Safe) funcionar em rede.



# A REDE CARINIANA

- ❑ A rede Cariniana adotou o sistema LOCKSS como um processo ativo de preservação onde cópias são validadas automaticamente em diferentes locais.
- ❑ Um primeiro desafio enfrentado foi a assimetria do tamanho dos arquivos, sendo que algumas instituições guardaram acervos de dados maiores dos que elas podem chegar a produzir.



Cariniana



# A REDE CARINIANA

- ❑ O segundo desafio foi o fato de lidar com material confidencial ou sigiloso, que pode ser crucial para a pesquisa, mas não podem ser compartilhados na Internet.
- ❑ Esse tipo de replicação deve ser regulado para proteger a confidencialidade dos dados.
- ❑ A preservação digital distribuída de dados de pesquisa conta com instrumentos de segurança no armazenamento compartilhado que permita a replicação segura dos arquivos.



# A REDE CARINIANA

- ❑ Em 2015 iniciaram-se os trabalhos para a criação de um serviço de preservação de dados na rede Cariniana.
- ❑ Os estudos realizados confirmaram que o Dataverse prioriza e reforça as propriedades de preservação que devem ser verificadas no funcionamento de um sistema de informação que inclui ações de preservação digital, são provas de autenticidade e de integridade dos dados nos processos de comunicação.
- ❑ A partir dessa comprovação as atividades de suporte à preservação de dados de pesquisa nas universidades parceiras da Rede começaram a ser estruturadas.



Cariniana

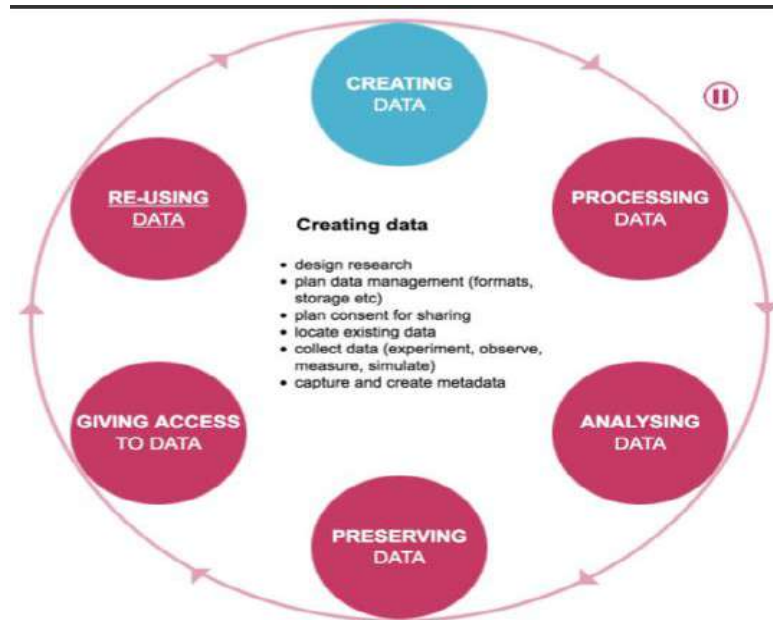
# A REDE CARINIANA

- ❑ A tradução do software para o português foi uma iniciativa do Ibtict para facilitar o acompanhamento das ações de curadoria e do cumprimento da política de gestão da instituição.
- ❑ Fico evidente que o fluxo da curadoria de dados inicia-se na observância dos padrões de depósito estabelecidos pela equipe responsável.
- ❑ Como aponta o modelo de ciclo de vida da curadoria do DCC, o rigor na submissão permite o controle das informações sobre os conjuntos de dados, seria uma auto curadoria previa ao uso dos repositórios de dados.



# A REDE CARINIANA

- ❑ Foi criado um grupo de estudos para copilar os elementos da política de gestão do repositório, que incluiu os da preservação digital. Entre eles estavam também, o reuso, o acesso, a aquisição, a retenção, o direito autoral e as licenças.
- ❑ O elemento da preservação digital na política de gestão envolve as condições de continuidade do conteúdo, a regularidade de realização de cópias de segurança, retenção de *bitstreams*, remoção de itens, autorizações de atualizações, atribuição de *checksum*, arquivamento e transferências de bancos de dados.



Research Data Life Cycle

Source: © Copyright 2002-2017 University of Essex. All rights reserved.



# REPOSITÓRIOS DE DADOS DE PESQUISA

## re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

These are the most important improvements introduced with version 3.0 for the Metadata Schema for the Description of Research Data Repositories.



### IMPLEMENTATION OF OPEN VOCABULARIES

We introduced open vocabularies for 'software', 'database-LicenseName', 'dataLicenseName', 'APItype', 'pidSystem', 'aidSystem' and 'certificate'. Now, all information will be visible to reflect details captured in re3data.org



### CITATION AND METRICS PROPERTIES

Two new properties have been included to incentivise data sharing and reuse, 'citationReference' and 'metrics'.



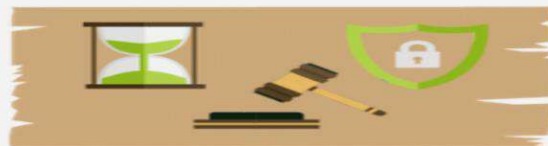
### ENGLISH LANGUAGE REQUIREMENT REMOVED

We have changed our policy to index non-English Research Data Repositories. If our editorial team has the necessary language skills we will welcome new languages! This is an important step to be a global registry.



### NEW REPOSITORY TYPES

Designed to make it easier to select and filter listed Research Data Repositories, since the number of them grows bigger and bigger. In addition to 'disciplinary', 'institutional' and 'other', we have added 'governmental', 'multidisciplinary' and 'project-related'.



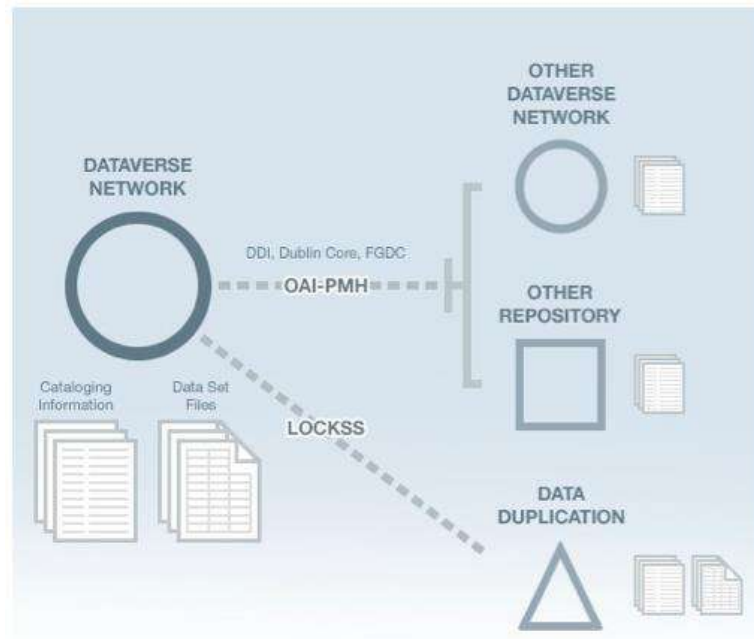
### ADDED A POLICY TYPE FIELD

This new 'policyType' field has been included to provide more detailed information about the policies used by the Research Data Repositories. It covers access, metadata, preservation, quality and other data policies.

# O REPOSITÓRIO DATAVERSE

As instituições usuárias do Dataverse podem criar suas nuvens de armazenamento, com Harvard que usa a nuvem aberta de Massachusetts da Universidade de Boston.

A integração com sistemas de preservação é uma das funcionalidades que o Dataverse permite para garantir a preservação dos conjuntos de dados.



# O REPOSITÓRIO DATAVERSE

A política de gestão de dados do repositório garante a disponibilidade de uma infraestrutura mínima para o repositório de pesquisa, propondo também que sejam aplicadas medidas de segurança e de proteção da privacidade.

The screenshot displays the DataVerse website interface. At the top, the logo 'DataVerse' is visible on the left, and navigation links for 'Sobre', 'Guias', 'Suporte', 'Acesso', and 'Entrar' are on the right. The main header features the 'Cariniana' logo and the text 'Rede Brasileira de Serviços de Preservação Digital'. Below this, it identifies the repository as 'IBICT DataVerse' and 'Cariniana'. A 'Metrics' bar shows '476 Downloads'. A descriptive paragraph states that IBICT provides a long-term data preservation repository with good archiving practices, supporting data sharing and citation. Below the text are four featured repository tiles: 'Repositório de Competências em Informação DataVerse', 'PROJETO BRASILIA 2060 DataVerse', 'Jardim Botânico de Brasília DataVerse', and 'Projeto Saberes do Cerrado DataVerse'. A search bar contains the text 'Pesquise este dataverse...'. The main content area shows search results for 'DataVerse (23)', 'Conjunto de dados (157)', and 'Arquivos (438)'. The 'Categoria' section lists 'Research Project (15)', 'Organization or institution (2)', 'Researcher (2)', 'Research Group (1)', and 'Teaching Course (1)'. The 'Busca por metadados' section lists 'IBICT DataVerse (134)' and 'Root DataVerse (46)'. The 'Data de publicação' section lists '2017 (12)' and '2015 (9)'. The search results list '1 a 10 de 180 Resultados' and includes a detailed entry for 'ESTUDO TEÓRICO-PRÁTICO SOBRE A COMPETÊNCIA EM INFORMAÇÃO COMO FATOR DE INTERAÇÃO ENTRE A CIÊNCIA DA INFORMAÇÃO E HISTÓRIA DA CIÊNCIA COM O USO DE ABORDAGENS QUALITATIVAS' by Rosseto, Marcos; Belluzzo, Regina Célia Baptista, 2017.

# A PRESERVAÇÃO DE DADOS DE PESQUISA DA REDE CARINIANA





# A PRESERVAÇÃO DE DADOS DE PESQUISA DA REDE CARINIANA

PKP  
PUBLIC  
KNOWLEDGE  
PROJECT

## OJS-Dataverse Integration



### OJS Journal



Citation  
to  
Data



Citation  
to  
Article

### Journal Dataverse



# OS METADADOS DE CITAÇÃO

Um dos aspectos mais relevantes do uso dos repositórios de dados de pesquisa é a sua capacidade de incluir metadados de citação para cada conjunto de dados submetido. Os metadados de citação do Dataverse para todos os conjuntos de dados são compatíveis com DataCite, Dublin Core, DDI (Data Documentation Initiative).

# A REDE CARINIANA

Uma das atividades da coordenação da Rede Cariniana é também a produção de guias de usuário, desenvolvimento de tutoriais com dicas sobre o preparo do material a ser submetido, fluxo da informação e melhores práticas.



# A REDE CARINIANA

As atividades da equipe da coordenação e do seu comitê técnico científico estão centradas na motivação de pesquisadores e instituições para o uso de repositórios digitais confiáveis, para manter sua flexibilidade de integração das funções e serviços que transcendem as funções do repositório de dados de pesquisa.





# A REDE CARINIANA

<http://cariniana.ibict.br>

BRASIL Serviços Participe Acesso à informação Legislação Canais

Ir para o conteúdo Ir para o menu Ir para a busca Ir para o rodapé

IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia

## Rede Cariniana

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

Buscar no portal

Agenda Contatos Eventos Notícias Perguntas frequentes Estrutura do portal

REDE DE PESQUISA DRIADEREPOSITÓRIO DATAVERSE

### SOBRE

IBICT

Cariniana

Linha do Tempo

Mapa da Rede Cariniana

Preservação Digital

### Parcerias da Rede

Parceiros Integrais

Parceiros Institucionais

Parceiros Internacionais

Logos para Participantes

### SERVIÇOS

Rede de Pesquisa DRIADE

Lista de Discussão

## Repositório Dataverse

Publicado: Terça, 22 de Agosto de 2017, 13h35



Acesse:

[Repositório Dataverse](#)



Acesse:

[Dataverse Demo](#)

Cariniana Rede Brasileira de Serviços de Preservação Digital

IBICT Dataverse Cariniana

438 Downloads

O IBICT está disponibilizando um repositório de dados de pesquisa que cuida da preservação a longo prazo e de boas práticas de arquivamento, para que os pesquisadores possam compartilhar, manter o controle e obter o reconhecimento de seus dados. O repositório suporta o compartilhamento de dados de pesquisa com citação de dados persistentes, permitindo que eles sejam reproduzidos. O Dataverse é um grande repositório aberto de dados de todas as disciplinas, criado pelo Institute for Quantitative Sciences da Universidade de Harvard. O repositório Dataverse do IBICT fornece um meio disponível gratuitamente para depositar e descobrir conjuntos de dados específicos arquivados por colaboradores das instituições participantes da rede Cariniana.

Faculdade de Tecnologia Dataverse

Repositório de Competência em Informação Dataverse

PROJETO BRASILIA 2060 Dataverse

Jardim Botânico de Brasília Dataverse

Pesquise este dataverse ...

Encontrar Busca Avançada

Adicionar dados



**ibict**

Instituto Brasileiro de Informação  
em Ciência e Tecnologia



**OBRIGADO**

