

# Fostering Open Data in Research

## The FAPESP Initiative

Claudia Bauzer Medeiros  
Institute of Computing – Unicamp

Member of the Coordination  
FAPESP Program on eScience and Data Science



# Research in Sao Paulo



50% of all research  
Produced in Brazil

25% in Sao Paulo

FAPESP

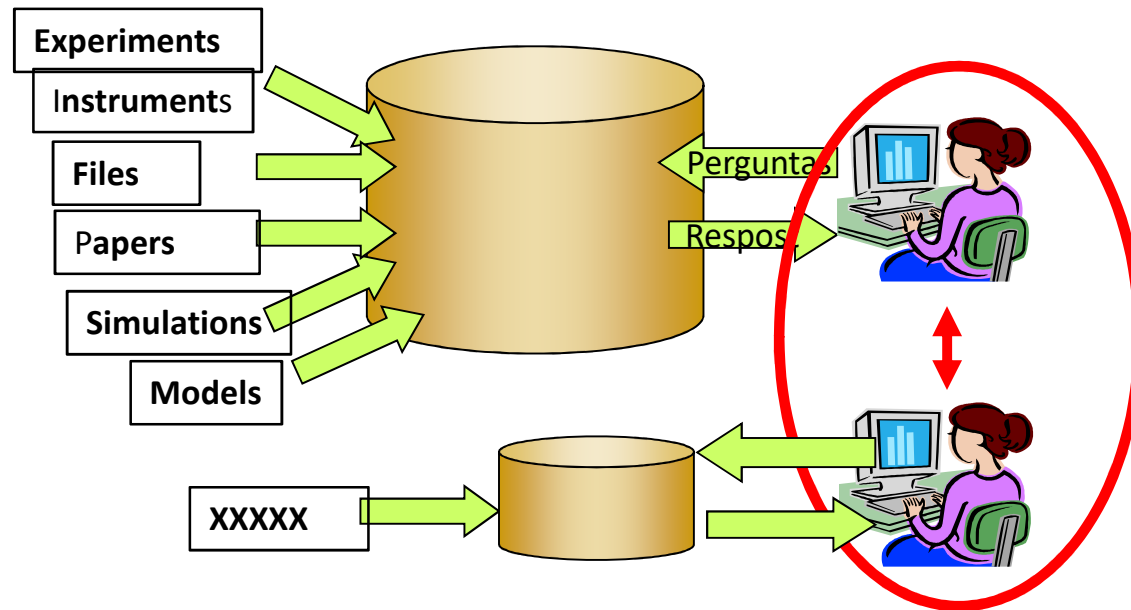
Approx 27,000 projects in 2017

Time for decision – 69 days, 40%  
success rate

Annual budget: \$PPP 534 million in  
2016

**COLLABORATION THROUGH DATA –  
OPEN SCIENCE AND OPEN DATA**

# Open Science – slide adapted from Gray

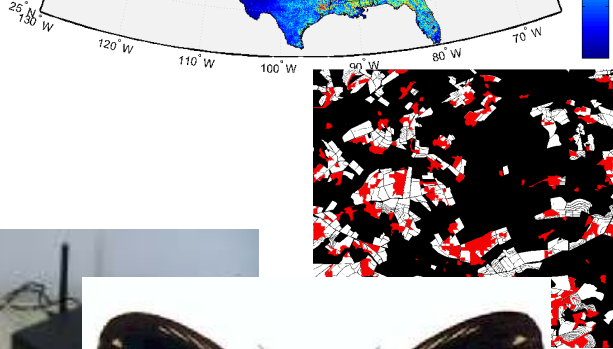
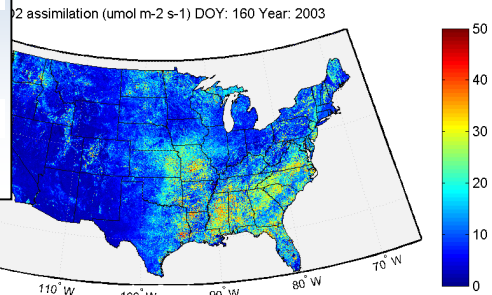
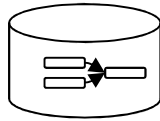


**Data driven-science**

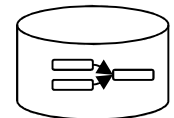
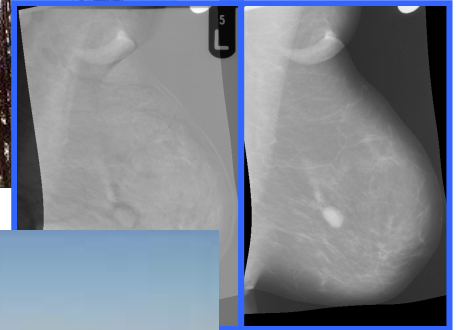
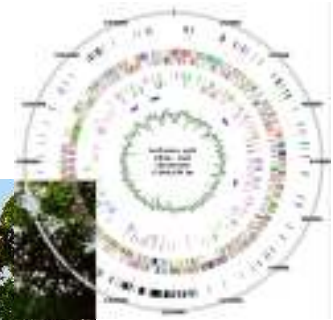
Research Data?

# Data sources?

Table of Product Characteristics		
id	Property name	Value
MilkProd	productsrep	MilkA
MilkProd	quantity	10000
MilkProd	validity date	10/06/2006
CheeseProd	productsr	Minas
CheeseProd	etquantity	2000
CheeseProd	validity date	12/02/2006
CheeseProd	shape	Circular

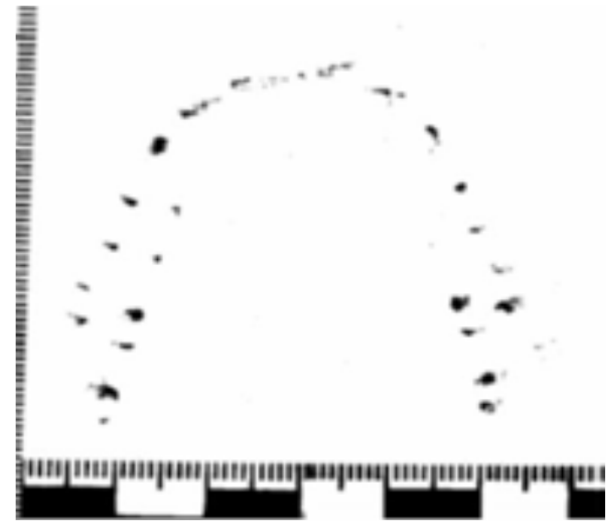
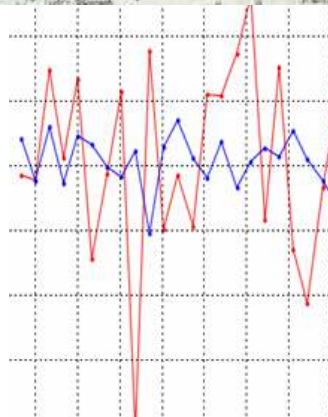


(D)  
*Dasyophthalma creusa*



# Data sources?

- Direct and indirect observations



7/10000

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE

**CONSENSUS STUDY REPORT**

# OPEN SCIENCE BY DESIGN

Realizing a Vision for 21st Century Research



**National Academies of Sciences,  
Engineering, Medicine**

**July 2018**

**Open science =**

**Open access = papers**

**Open data**

**Open methods = open source**



# Open Data

- “What is OPEN **DIGITAL** DATA”
  - Share “everything”? Not really
- Everyone can
  - Discover if data exist
  - Discover how to obtain them

Under constraints – security, confidentiality, ethics, intellectual property

# Open Data – Open **META**Data

- “What is OPEN **DIGITAL** DATA”
  - Share “everything”? Not really
- Everyone can
  - Discover if data exist
  - Discover how to obtain them

Under constraints – security, confidentiality, ethics, intellectual property

# HOW???

## Datacite.org

(Find, share, cite, connect)



Find what you're looking for by searching millions of records with extensive, reliable metadata.



Share your data and reuse the data of others to create the highest impact in the research community.



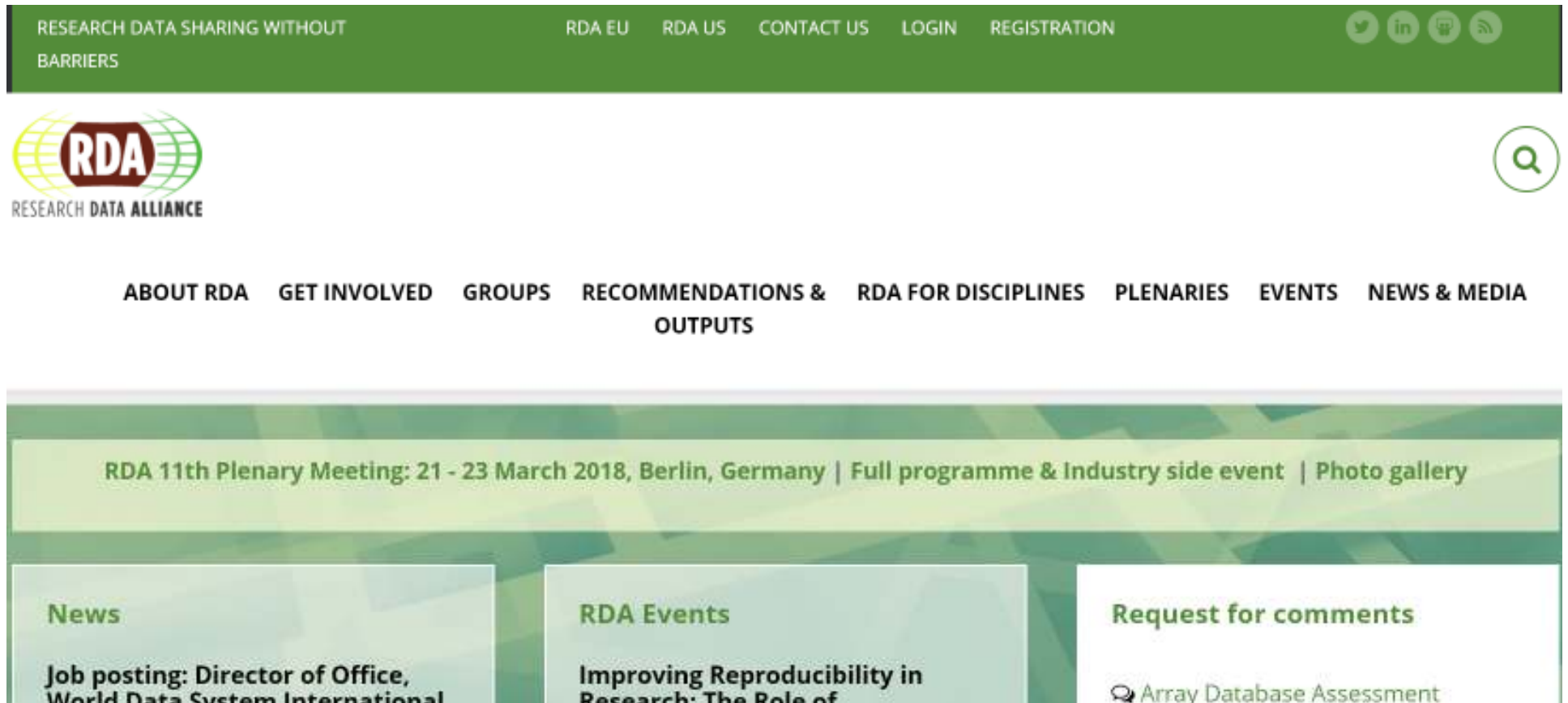
Cite your research sources with confidence, and receive proper credit when your work is reused.



Connect your research – public datasets, software, authors, institutions and funding data all in one place.

**HOW IS THE WORLD ORGANIZING  
OPEN DATA REPOSITORIES /  
INFRASTRUCTURES / POLICIES?**





# RDA - <https://www.rd-alliance.org>





The image shows a screenshot of the RDA (Research Data Alliance) website homepage. The top navigation bar is green and contains the text "RESEARCH DATA SHARING WITHOUT BARRIERS" on the left, and "RDA EU", "RDA US", "CONTACT US", "LOGIN", and "REGISTRATION" in the center. On the right side of the navigation bar are social media icons for Twitter, LinkedIn, Facebook, and RSS. Below the navigation bar is the RDA logo, which consists of a globe with "RDA" in a red circle and "RESEARCH DATA ALLIANCE" below it. To the right of the logo is a search icon. Below the logo is a horizontal menu with the following items: "ABOUT RDA", "GET INVOLVED", "GROUPS", "RECOMMENDATIONS & OUTPUTS", "RDA FOR DISCIPLINES", "PLENARIES", "EVENTS", and "NEWS & MEDIA". Below the menu is a large green banner with the text "RDA 11th Plenary Meeting: 21 - 23 March 2018, Berlin, Germany | Full programme & Industry side event | Photo gallery". Below the banner are three columns of content. The first column is titled "News" and contains the text "Job posting: Director of Office, World Data System International". The second column is titled "RDA Events" and contains the text "Improving Reproducibility in Research: The Role of". The third column is titled "Request for comments" and contains the text "Array Database Assessment".

RESEARCH DATA SHARING WITHOUT BARRIERS

RDA EU RDA US CONTACT US LOGIN REGISTRATION


 

ABOUT RDA GET INVOLVED GROUPS RECOMMENDATIONS & OUTPUTS RDA FOR DISCIPLINES PLENARIES EVENTS NEWS & MEDIA

RDA 11th Plenary Meeting: 21 - 23 March 2018, Berlin, Germany | Full programme & Industry side event | Photo gallery

**News**  
Job posting: Director of Office, World Data System International

**RDA Events**  
Improving Reproducibility in Research: The Role of

**Request for comments**  
 Array Database Assessment

**FAPESP INITIATIVES – OPEN  
DATA/SCIENCE**

# FAPESP Code of Good Scientific Practice (2011)

## 3.3. On the recording, storage and accessibility of data and information

3.3.1. Researchers must precisely and completely record the data and information collected, the procedures utilized and any partial results obtained during the course of a research study.

3.3.2. Original research data should be securely stored for a substantial period after the study's results are published. The length of this period may vary according to the study's area and characteristics but should be a minimum of five years. The researchers and their research institutions share the responsibility for storing this information.

3.3.4. After the results are published, the research records must be made available to other researchers who may want to verify the study's correctness or replicate or continue the study. Accessibility may only be limited for ethical or legal reasons.



**2017**

**Data Management Policy**

**Compulsory Data  
Management Plans**

**WG – 7 public universities**

**Establish network of  
Research data repositories**



# WG – Data repository network



Seven public universities, approx. 48 campi  
11,5 thousand faculty  
170 thousand students

Mission – establish network

# WG – Data repository network



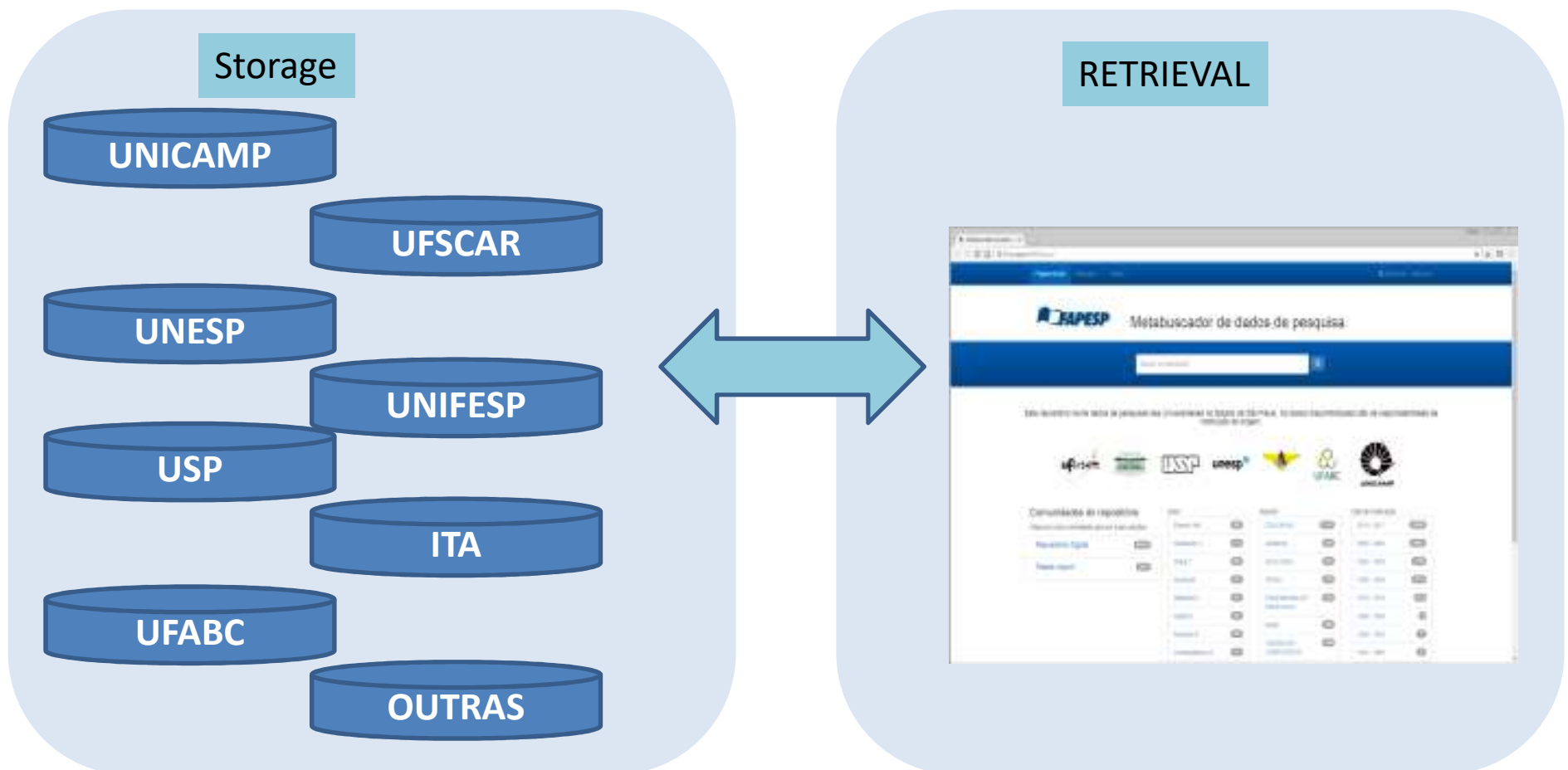
Seven public universities, approx. 48 campi  
11,5 thousand faculty  
170 thousand students  
**+ researchers in (informatics in) agriculture**

Mission – establish network

# WG – Data Repository Network



- Each university has its own system
- Single search (metadata harvester) interface



## Eight compulsory metadata fields

ID	Type	Description
1	dc.title	Project/file title
2	dc.subject	Keywords
3	dc.description	Abstract
4	dc.contributor.author	Author (ORCID)
5	dc.identifier.uri	File id
6	dc.description.sponsorship	Funding agencies
7	dc.description.sponsorshipId	Project numbers

8 dc.type

File type (software, others)

# Implementation status

Harvester prototype running

Three different data repositories online

DSPACE

DATAVERSE

CKAN

Universities at different implementation stages

**WE NEED VOLUNTEERS FOR TESTING**



**2017**

**Data Management Policy**

**Compulsory Data  
Management Plans**

**WG – 7 public universities**

**Establish network of  
Research data repositories**

# www.fapesp.br/gestaodedados



FUNDAÇÃO DE AMPARO À PESQUISA  
DO ESTADO DE SÃO PAULO

Índice ▾

Buscar 🔍

Eng

Converse com a  
FAPESP

## Plano de Gestão de Dados - FAPESP

### Gestão de Dados



A FAPESP reconhece a importância da gestão adequada dos dados de pesquisa como parte essencial das boas práticas de pesquisa. Para tanto, considera necessário que os dados resultantes de projetos financiados pela Fundação sejam gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico.

Além de racionalizar recursos, a gestão apropriada de tais

### Índice

[Gestão de Dados](#)

[Conteúdo do Plano de Gestão de Dados - FAPESP](#)

[Ferramentas online para criação de Planos](#)

[Documentos e páginas de interesse - Planos de Gestão de Dados](#)

[Contato para dúvidas: plano-gestao-dados@fapesp.br](#)

### Documentos e páginas de interesse - Planos de Gestão de Dados

Planos de Gestão de Dados vêm sendo exigidos há mais de uma década pelas principais agências governamentais ou privadas de fomento à pesquisa na América do Norte, Europa e Austrália. Esta página apresenta alguns endereços da rede onde podem ser consultados manuais, exemplos de planos e diretivas para redigi-los, para as mais variadas disciplinas


# Data Management Plan

- **WHICH** data will manage
- **WHERE** will store
- For how long (**TIME**)
- **HOW**
- Under which **PROTECTION CONSTRAINTS** (ethics, intellectual property, etc)



**CREATING A PLAN  
(WHICH, WHERE, HOW LONG, HOW,  
PROTECTION)**

dmp-tool.org [USP, UNESP, UFABC, UNICAMP]



The image shows the homepage of the DMPTool website. The header features the DMPTool logo on the left and a navigation menu on the right with links for Home, DMP Requirements, Public DMPs, News, Help, Contact Us, and About. The main content area has a dark green background with a geometric pattern. It features a large title 'Data Management Planning Tool' and a subtitle 'Create, review, and share data management plans that meet institutional and funder requirements.' A 'Get Started' button is located in the bottom right corner.

**DMPTool**

[Home](#) [DMP Requirements](#) [Public DMPs](#) [News](#) [Help](#) [Contact Us](#) [Ab](#)

# Data Management Planning Tool

Create, review, and share data management plans that meet institutional and funder requirements.

[Get Started](#)



# Welcome

Create data management plans that meet institutional and funder requirements.

## Sign in options

Option 1: If your institution is affiliated with DMPTool.

Your institution


- or -

Option 2: If your institution is not affiliated with DMPTool.

Email address

- or -

### Look up your institution here

Go

[See the full list of participating institutions](#)

Institution not in the list? [Create an account with any email address](#)

## University of São Paulo (USP)



Sign in

Create account

\* Email

cmbm@ic.unicamp.br

\* Password

.....

Remember email

Sign in

[Forgot email?](#)  
[Forgot password?](#)



University of Campinas (UNICAMP)

Benilton Carvalho

[My dashboard](#) [Create plan](#)

## Create a new plan

Before you get started, we need some information about your research project to set you up with the appropriate template.

### What research project are you planning?

Exemplo de plano para CINFOTEC

Mock project for testing, practice, or educational purposes

### Select the primary research organization

University of Campinas (UNICAMP) x

- or -  My research organisation is not on the list or no research organisation is associated with this plan

### Select the primary funding organization

Begin typing to see a filtered list

- or -  No funder associated with this plan

[Create plan](#) [Cancel](#)

[About](#) [Terms of use & Privacy](#) [Accessibility](#) [GitHub](#) [Contact us](#)

DMPTool is a service of the University of California Curation Center of the California Digital Library  
Copyright 2010-2018 The Regents of the University of California



## Mock USP generic project

Project details

Plan overview

Descrição dos Dados e Metadados produzidos pelo projeto

Restrições legais ou éticas

Política de preservação e compartilhamento

Descrição de mecanismos, formatos e padrões para armazenamento

Share

Download

### Template USP - Mínimo

This plan is based on the "Template USP - Mínimo" template provided by University of São Paulo (USP).

Template construído para responder às perguntas básicas indicadas pela FAPESP (<http://www.fapesp.br/gestaodedados/>) para um Plano de Gestão de Dados:

1. Quais dados serão gerados pelo projeto;
2. Como serão preservados e disponibilizados, considerando questões éticas, legais, de confidencialidade e outras.

O texto de um Plano varia conforme a disciplina, os tipos de dados considerados e como os responsáveis pelo projeto pretendem disponibilizá-los. Algumas chamadas FAPESP poderão especificar o formato desejado do Plano. Para todos os demais casos, o Plano submetido como anexo de uma proposta à FAPESP poderá seguir o apresentado neste template.

Descrição dos Dados e Metadados produzidos pelo projeto (1 section, 2 questions)

+

Restrições legais ou éticas (1 section, 2 questions)

+

Política de preservação e compartilhamento (1 section, 2 questions)

+

Descrição de mecanismos, formatos e padrões para armazenamento (1 section, 2 questions)

+

## Mock USP generic project

Project details

Plan overview

Descrição dos Dados e Metadados produzidos pelo projeto

Restrições legais ou éticas

Política de preservação e compartilhamento

Descrição de mecanismos, formatos e padrões para armazenamento

Share




Download

expand all | collapse all

1/2 answered

### — Descrição dos dados e metadados produzidos (1 / 2)

#### Que dados serão coletados ou criados?

**B** *I*    

1 planilha Excel com 1 linha e 2 colunas, nunca sera versionada

Save

Answered just now by [cmbm@ic.unicamp.br](mailto:cmbm@ic.unicamp.br)

Guidance

Comments

USP

#### Guidance

Aqui devemos considerar questões como:

-Que tipo, formato e volume de dados?

- Os formatos e softwares escolhidos permitem o compartilhamento e o acesso de longo prazo aos dados?

## Mock USP generic project

Project details	Plan overview	Descrição dos Dados e Metadados produzidos pelo projeto			
Restrições legais ou éticas		Política de preservação e compartilhamento			
Descrição de mecanismos, formatos e padrões para armazenamento			Share	Download	

Download settings

**Select phase to download**

Descrição dos Dados e Metadados produzidos pelo projeto ▼

Optional plan components

- project details coversheet
- question text and section headings
- unanswered questions

Format

ndf ▼

PDF formatting

Font

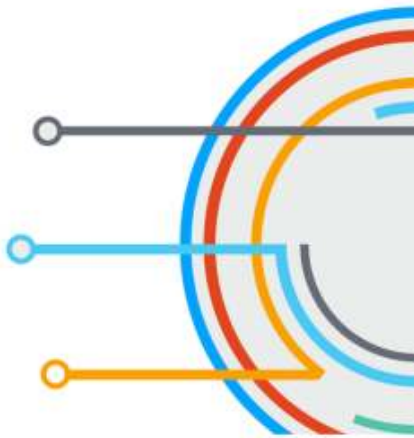
Face	Size (pt)	Margin (mm)			
		Top	Bottom	Left	Right
Arial Helvetica Sans-Sei ▼	10 ▼	▼	▼	▼	▼

[Download plan](#)



## OPEN SCIENCE BY DESIGN

Realizing a Vision for 21st Century Research



Código



### Código de boas práticas científicas

A FAPESP busca reforçar na comunidade científica paulista, uma cultura sólida e bem arraigada de integridade ética da pesquisa mediante um conjunto de estratégias assentado sobre três pilares



## Welcome

Create data management plans that meet institutional and funder requirements.

### Plano de Gestão de Dados - FAPESP

#### Gestão de Dados



#### Índice

- Gestão de Dados
- Conteúdo do Plano de Gestão de Dados - FAPESP
- Ferramentas online para criação de Planos
- Documentos e páginas de interesse - Planos de Gestão
- Contato para dúvidas: [plano-gestao-dados@fapesp.br](mailto:plano-gestao-dados@fapesp.br)

#### Publicações



Código de



Obrigada!!!!

CMEDEIROS@FAPESP.BR