

Promovendo Dados Abertos na Pesquisa

Claudia Bauzer Medeiros

Coordenação Programa eScience e Data Science – FAPESP

Instituto de Computação - UNICAMP



Main take-aways

- What is open science?
- What does OPEN mean?
- FAPESP and open research data?

Outline

Open Science

Definitions

Challenges

4 Challenges in (Research Data) Sharing

Open Data at FAPESP

Data Management Plans

Network of open research data repositories

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

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

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

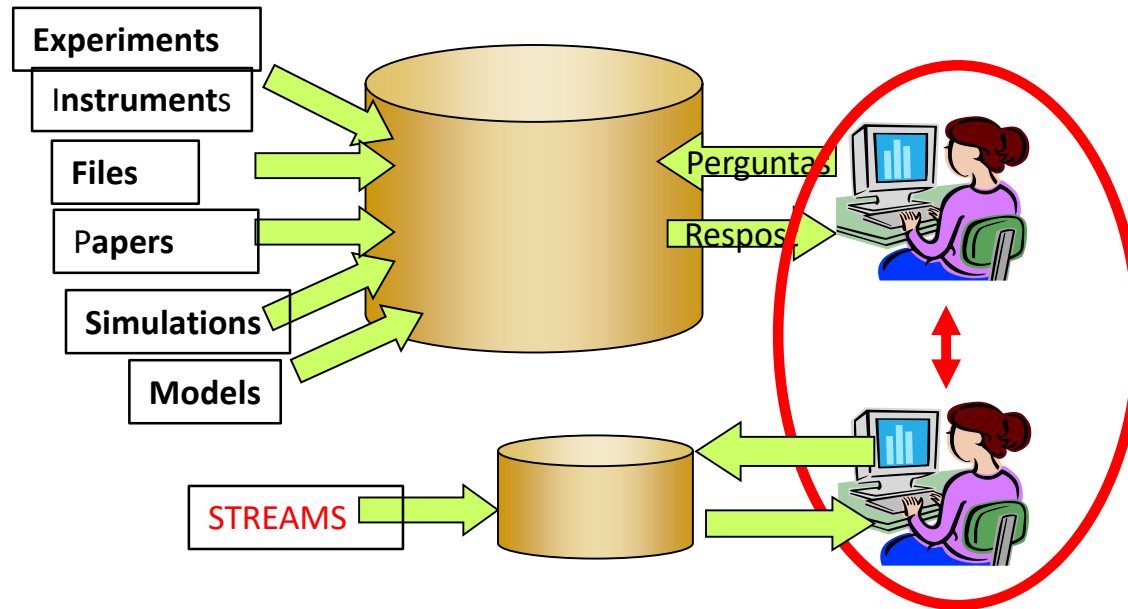
[Array Database Assessment](#)

OPEN????



**All artifacts associated with a scientific
experiment
available in public repositories**

How to Share – slide adapted from Jim Gray



Data driven-science

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

Horizon Europe: evolution not revolution

Specific objectives of the Programme

Support the creation and diffusion of high-quality knowledge

Strengthen the impact of R&I in supporting EU policies

Foster all forms of innovation and strengthen market deployment

Optimise the Programme's delivery for impact in a strengthened ERA



Pillar 1 Open Science

European Research Council

Marie Skłodowska-Curie Actions

Research Infrastructures



Pillar 2 Global Challenges and Industrial Competitiveness

Clusters

- Health
- Inclusive and Secure Society
- Digital and Industry
- Climate, Energy and Mobility
- Food and natural resources

Joint Research Centre



Pillar 3 Open Innovation

European Innovation Council

European innovation ecosystems

European Institute of Innovation and Technology

Strengthening the European Research Area

Sharing excellence

Reforming and Enhancing the European R&I system

What is Open Data?

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

Under constraints – security, confidentiality, ethics, intellectual property

OPEN = OPEN METADATA

OPEN SCIENCE

Metadata on papers – and papers...

Metadata on data

Metadata on software

**Metadata on everything associated
with experiment**

FILES FILES FILES FILES FILES

...

FILES FILES FILES FILES FILES

“The European Union could save €10.2 billion per year by using FAIR and free open access research data.”*

PwC EU Services (March 2018). *Cost-Benefit analysis for FAIR research data - Cost of not having FAIR research data*. Directorate General for Research and Innovation (European Commission).

*FAIR: Data meeting standards of Findability, Accessibility, Interoperability and Usability.

Challenges?

- Volume
- Analysis needs
- Preservation

⇒ NASA supercomputer -> 710 Tbyte memory

⇒ IBM sequoia -> 1Pbyte memory

Plan S

Making full and immediate Open Access a reality



FAPESP and Open Access

- All publications resulting from FAPESP-funded projects must be published as open access
- (According to the policy of each journal)
- As soon as accepted

Outline

Open Science

Definitions

Challenges

4 Challenges in (Research Data) Sharing

Open Data at FAPESP

Data Management Plans

Network of open research data repositories

(Four) Challenges in Sharing Data

(1). What is data ?

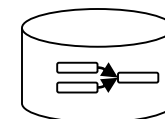
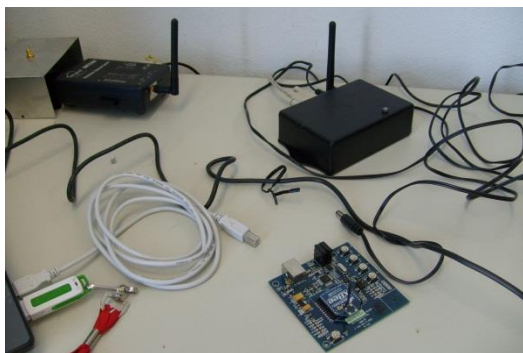
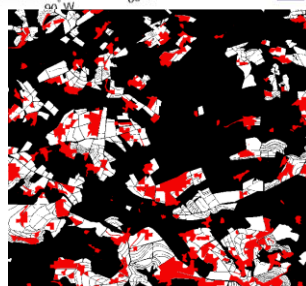
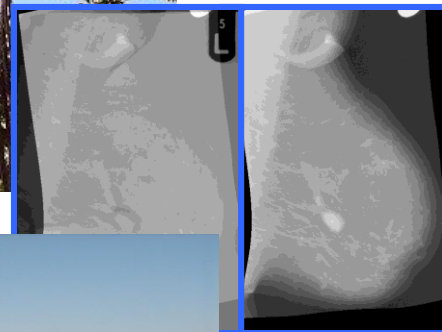
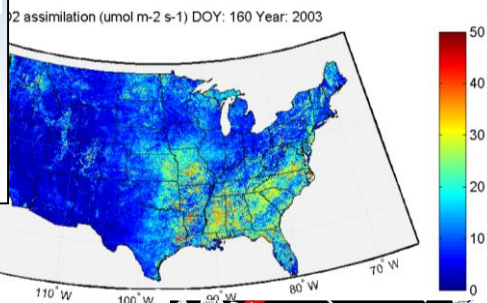
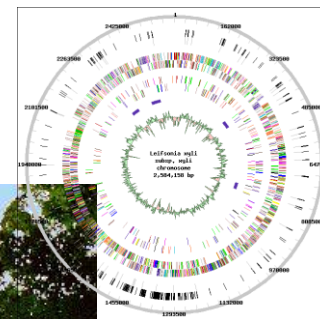
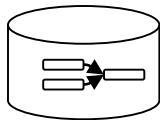
(2). Lack of common/consensual infrastructures

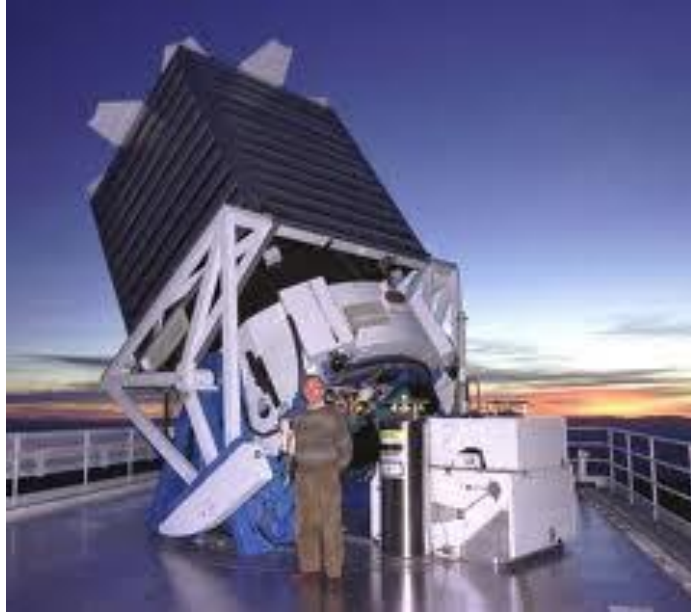
(3). Finding and identifying

(4). Understanding

What is data?

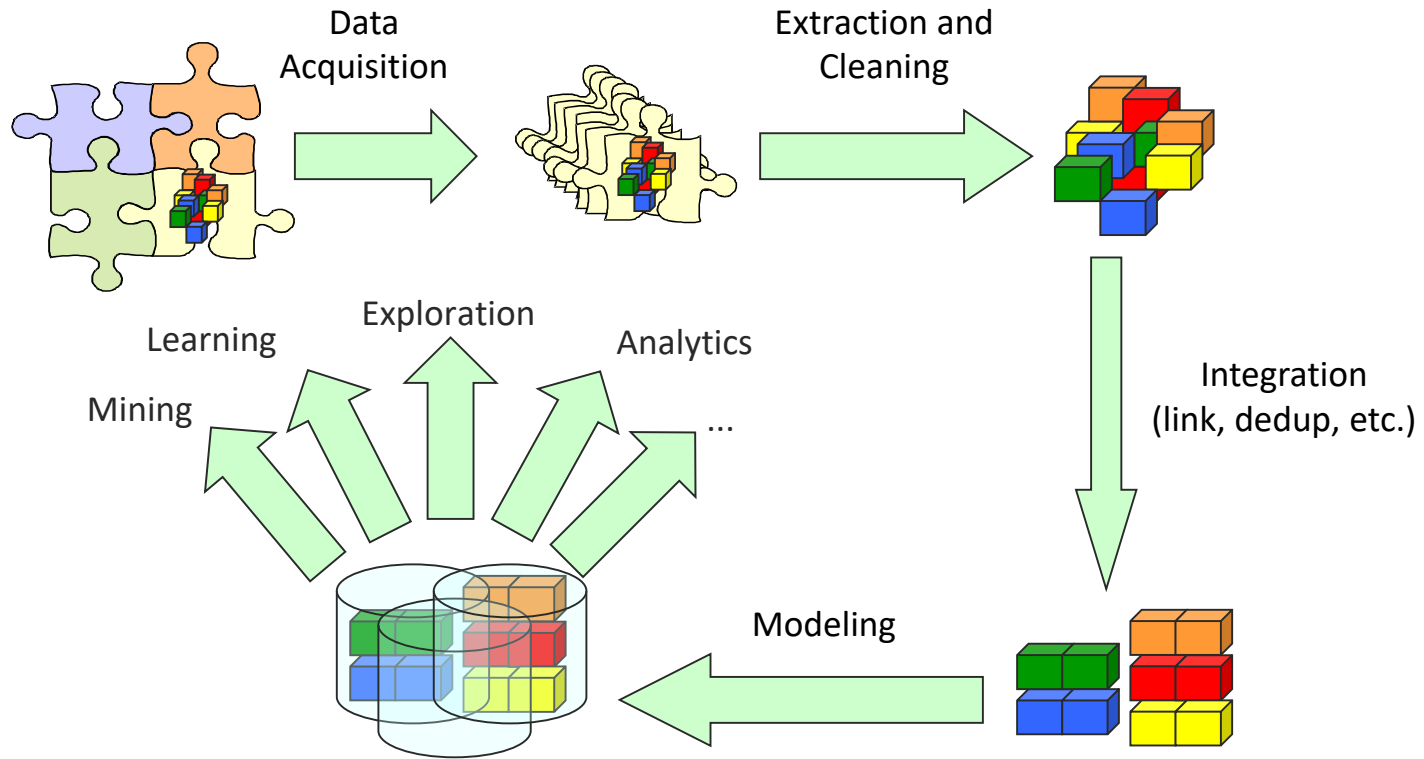
Table of Product Characteristics		
id	Property name	Value
MilkProd	productsrep	MilkA
MilkProd	quantity	10000
MilkProd	validity date	10/06/2006
CheeseProd	productsr	Minas
CheeseProd	equantity	2000
CheeseProd	validity date	12/02/2006
CheeseProd	shape	Circular



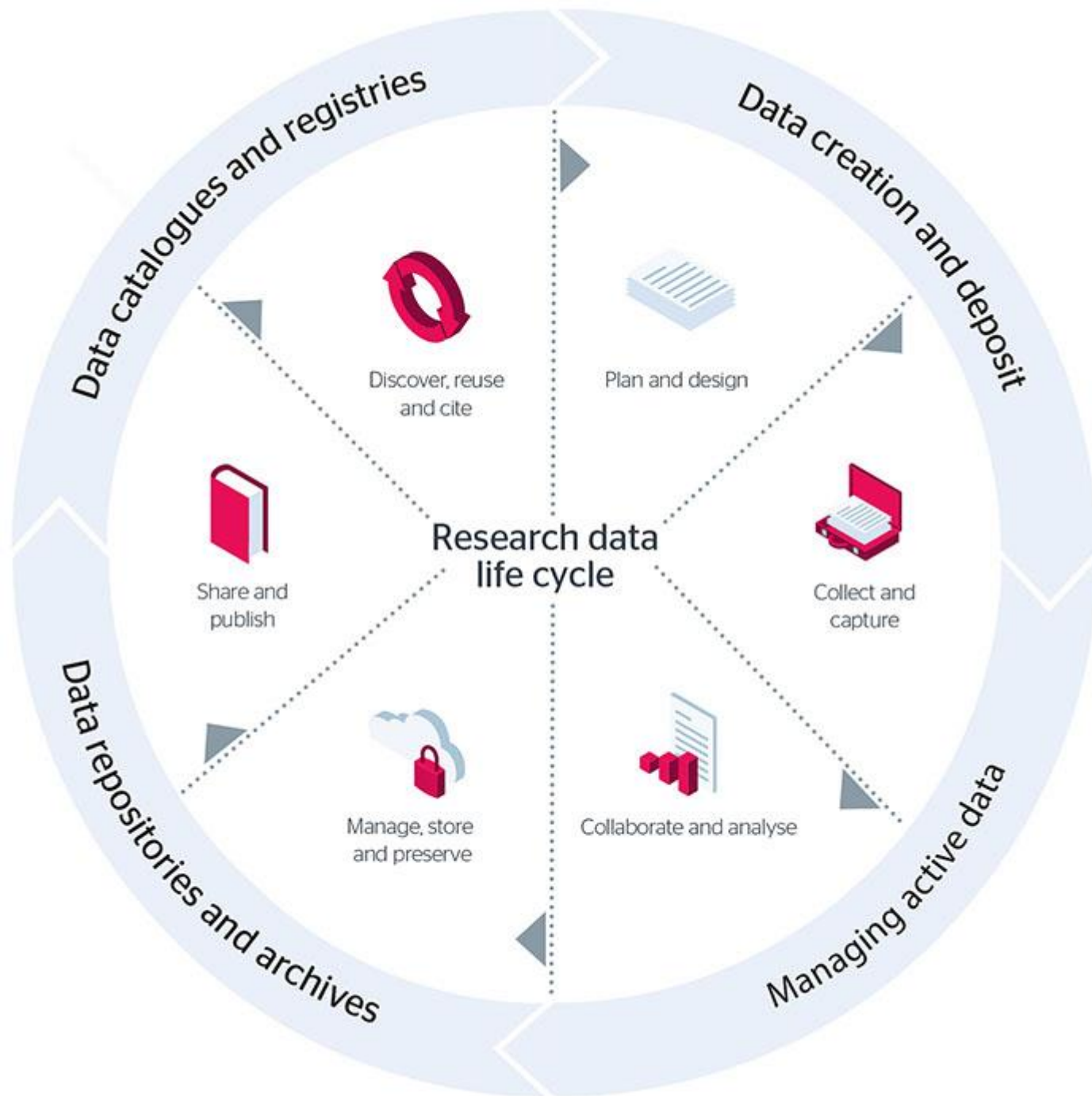




The Big Data Pipeline



H. V. Jagadish ACM SIGMOD Blog - 06/2012



Finding and Identifying

How to search/find (specification ?)

Where?

What to open (everything vs ethics/privacy)

Unique id?

HOW???

Datacite.org

(Find, share, cite, connect)



About us ▾

Services ▾

Resources ▾

Community ▾

Become a member

WELCOME TO DATACITE

Locate, identify, and cite research data with the leading global provider of DOIs for research data.



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???) Handle.Net

Corporation for National Research Initiatives

Handle.Net® Registry

[HOME](#) [SOFTWARE](#) [PREFIXES](#) [PAYMENT](#) [DOCUMENTATION](#) [SUPPORT](#)

HDL.NET® Information Services

Welcome to the web site of the Handle.Net Registry (HNR), run by [Corporation for National Research Initiatives \(CNRI\)](#). CNRI is a [Multi-Primary Administrator \(MPA\)](#) of the Global Handle Registry (GHR), authorized by the DONA Foundation to allot prefixes to users of the Handle System. The [DONA Foundation](#) is a non-profit organization based in Geneva that has taken over responsibility for the evolution of CNRI's Digital Object (DO) Architecture including outreach around the world. One of the Foundation's responsibilities is to administer and maintain the overall operation of the GHR, a task that was previously performed by CNRI.

The Handle.Net Registry will allot prefixes of the form "20.500" followed by four or more digits (i.e., 20.500.1234). Users who are allotted a prefix from the HNR will have their associated prefix handle records registered with the HNR to enable HDL.NET resolution services for their identifiers. Please click [here](#) to request CNRI to allot a prefix or renew a previously allotted prefix.

Interpretation and Understanding

How to understand the data you find?

How to reuse it?

Who are the experts?

Who will interpret?

For whom?

Outline

Open Science

Definitions

Challenges

4 Challenges in (Research Data) Sharing

Open Data at FAPESP

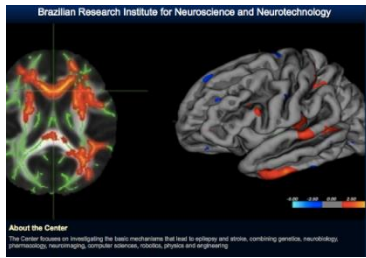
Data Management Plans

**Network of open research data
repositories**

Fapesp: São Paulo Research Foundation

- Mission: support research in all fields
- All proposals are peer reviewed (26,000 proposals in 2017)
- Funded (1962) by the State of São Paulo with 1% of all state revenues
- Annual budget: US\$ 530 M

BRAINN



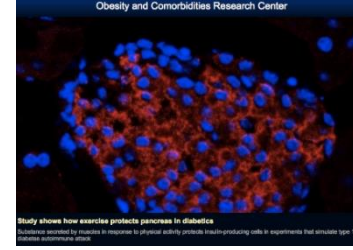
CTC



CRID



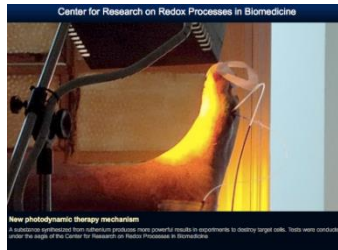
OCRC



HUG-CELL



REDOXOMA



CIBFar



CETICS



FORC



CEPIV



CDMF



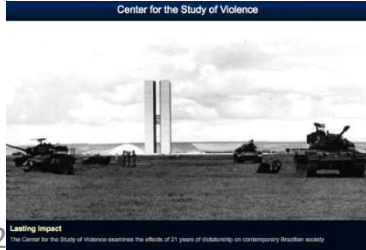
CEPOF



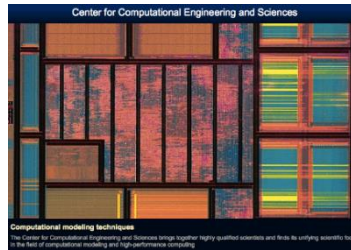
CEM



NEV



CCES



CEMEAI



NEUROMAT





2017

Data Management Policy

**Compulsory Data
Management Plans**

WG – 7 public universities

**Establish network of
Research data repositories**

Data Management Plan

- **WHICH** data will produce
- **WHERE** will store
- For how long **TIME**
- **HOW**
- Given ethical, privacy, IP aspects etc

WG – Data repository network



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

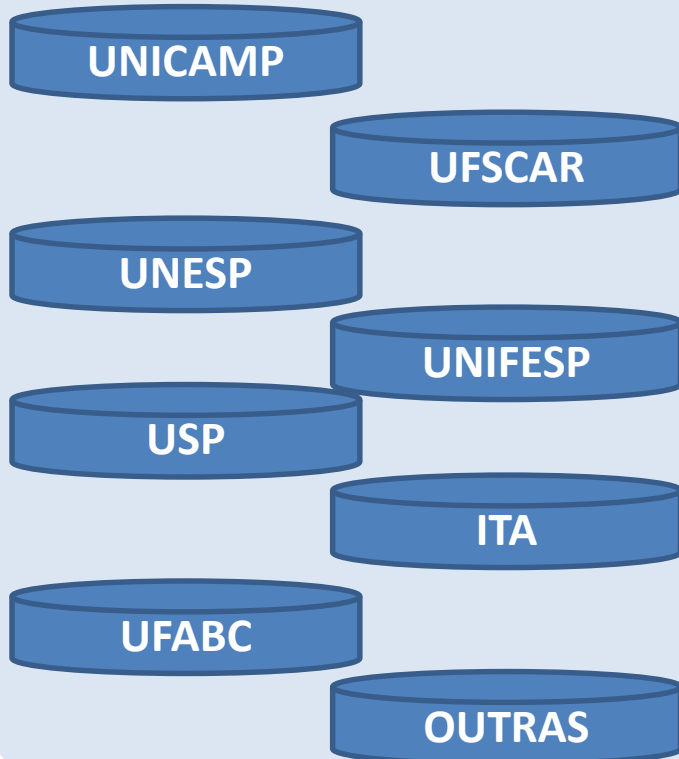
Mission – establish network

WG – Data Repository Network

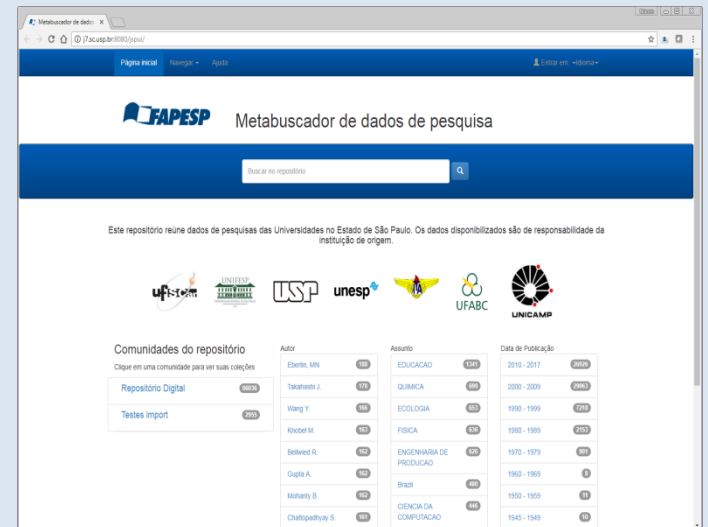


- Each participant has its own system
- Single search (metadata harvester) interface

Storage



RETRIEVAL



NINE compulsory metadata fields

ID	Type	Description
1	dc.title	Project 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)

9 dc.identifier

File id (handle)

Preparing DMP - dmptool.org



[Home](#)

[DMP Requirements](#)

[Public DMPs](#)

[News](#)

[Help](#)

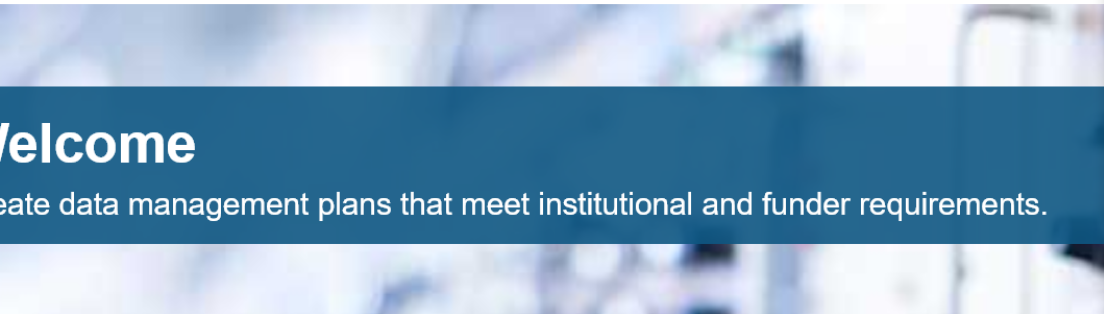
[Contact Us](#)

[About](#)

Data Management Planning Tool

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

[Get Started](#)



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](#)

What is the meaning of OPEN?

OPEN metadata in public repository

Outline

Open Science (Open metadata)

Definitions (Papers, data, software)

**Challenges – Volume, Processing,
Preservation**

4 Challenges in (Research Data) Sharing

Open Data at FAPESP

Data Management Plans

Network of open research data repositories

OBRIGADA