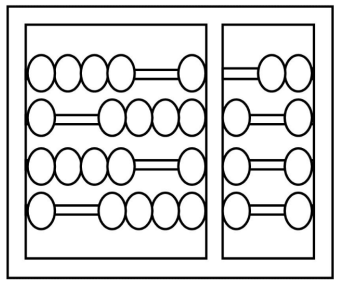


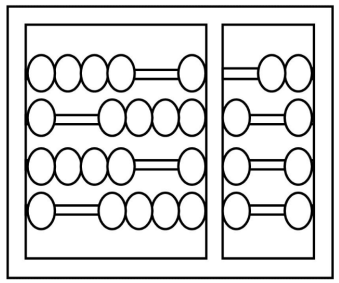
Open Science – Challenges and Opportunities

Claudia Bauzer Medeiros
Institute of Computing – Unicamp

UNICAMP

PESC
Programa de Engenharia
de Sistemas e Computação

OBRIGADA



Open Science – Challenges and Opportunities

Claudia Bauzer Medeiros
Institute of Computing – Unicamp

UNICAMP

Main take-aways

- What is open science?
- What does OPEN mean?
- What is data?
- Why should I care?

Outline

Open Science

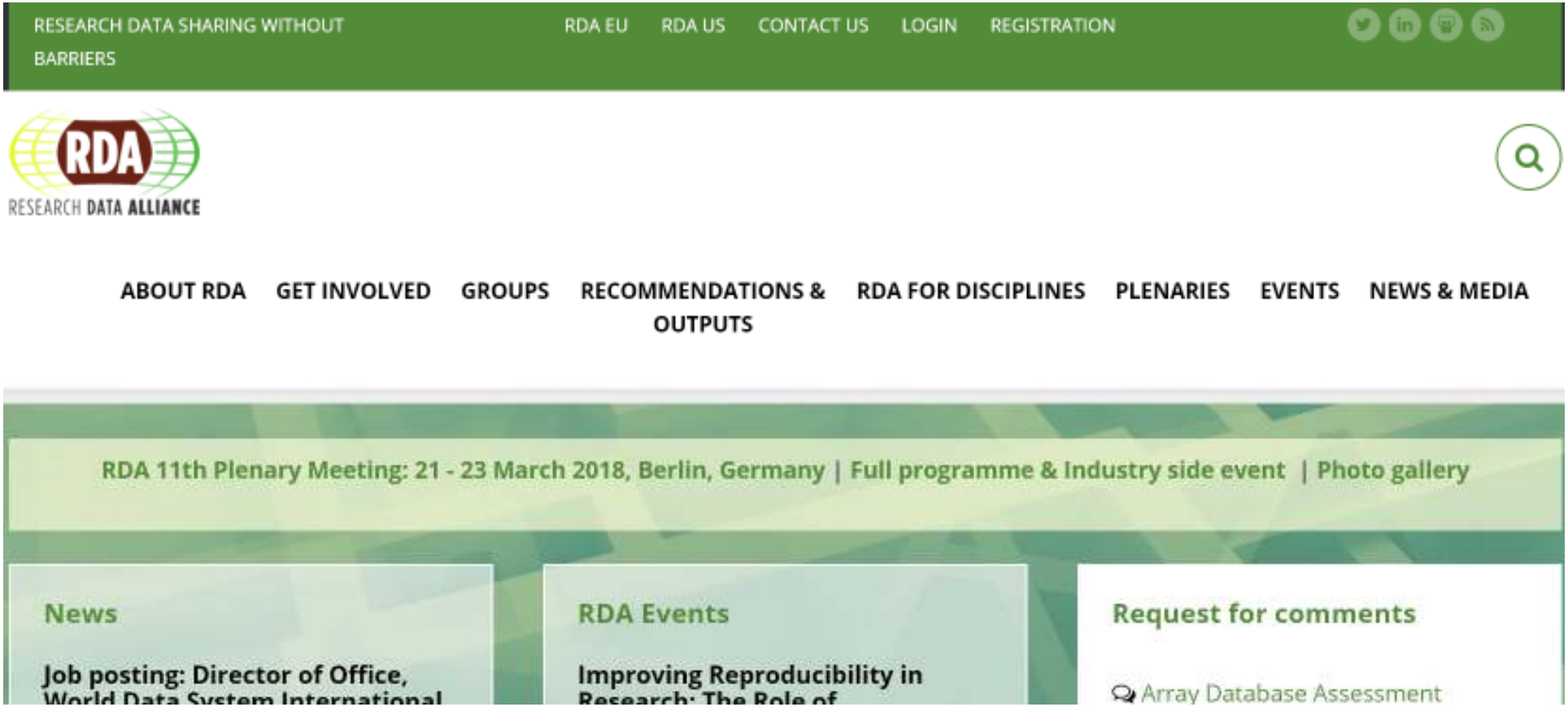
- **Definitions**
- **Challenges**

4 Challenges in (Research Data) Sharing

Open Data at FAPESP

- Data Management Plans
- Network of open research data repositories

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



@ Claudia Bauzer Medeiros

UK Research and Innovation

European Commission

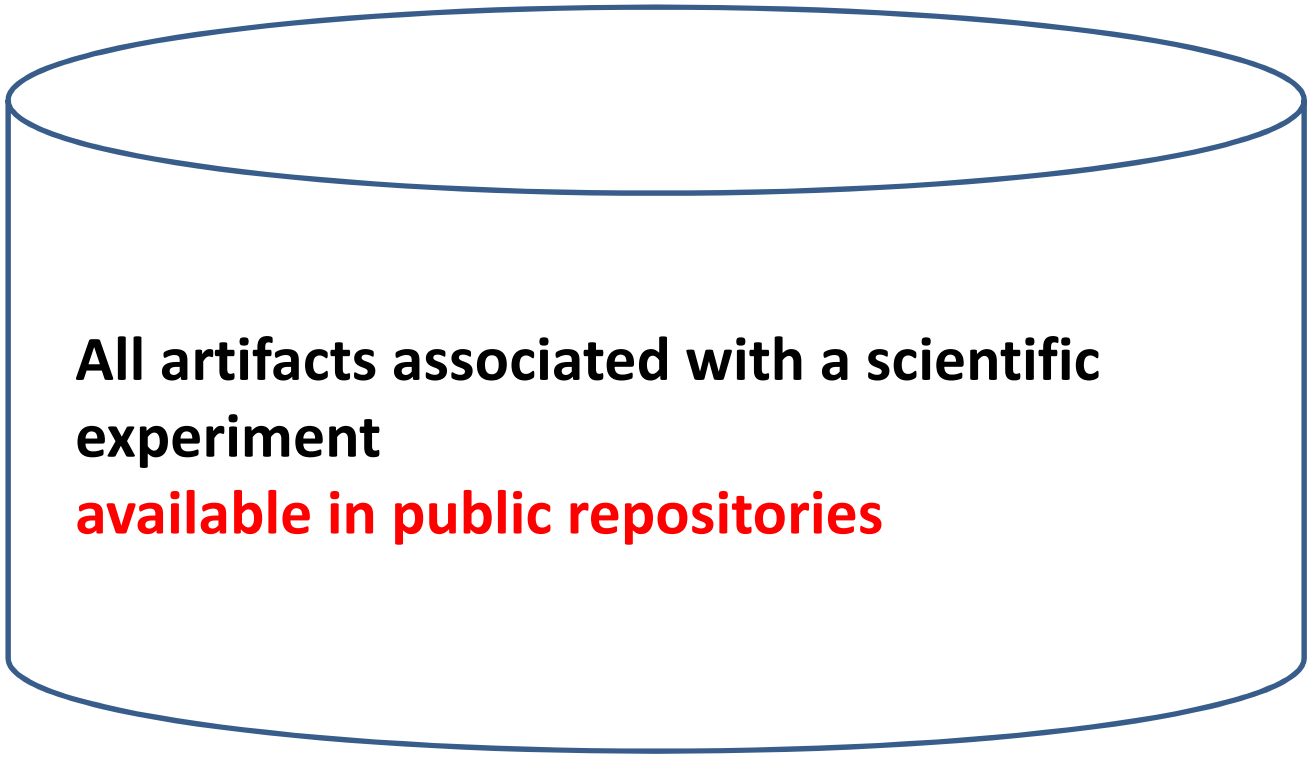
Data Archiving and Networked Services
DANS

portage
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHE
SHARED STEWARDSHIP OF RESEARCH DATA

JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE
日本学術振興会

NIH National Institutes of Health
Turning Discovery Into Health

OPEN????


All artifacts associated with a scientific experiment
available in public repositories

National Academies of Sciences, Engineering, Medicine

July 2018

Open science =
- **Open access = papers**
- **Open data**
- **Open methods = open source**

What is Open Data?

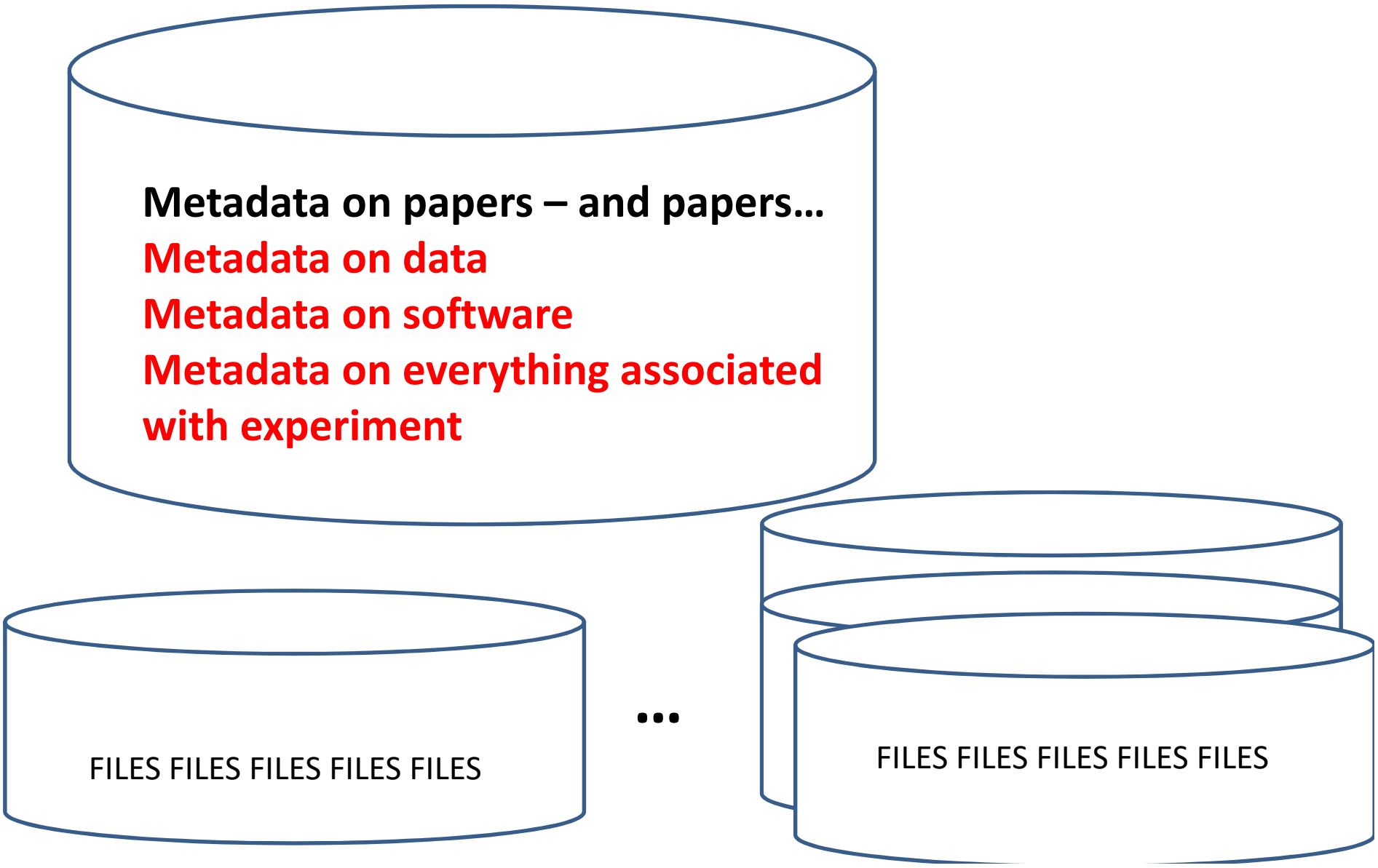
- "What is OPEN **DIGITAL** DATA"
  - Share "everything"? Not necessarily

- Everyone can
  - Discover if data exist
  - Discover how to obtain them

  Under constraints – security, confidentiality, ethics, intellectual property

OPEN SCIENCE – OPEN METADATA

OPEN SCIENCE



International Scenario – Open Science

- Official policy in North America, Australia and New Zealand
- Compulsory for European financing after 2021
- Japan, South Korea
- Brazil – Federal government plans
- Brazil – FAPESP policies (Open access, open data)

Open Science – G7 Priority

1. **Human Capital Formation – research and innovation**
2. **Financing – inclusive science, research and innovation**
3. **Global Research Infrastructures**

---→ Open Science

(Canada, USA, France, Germany, Japan, Italy, UK)

+ Representative from EU

G7 SCIENCE MINISTERS' COMMUNIQUÉ

Turin, 27 – 28 September

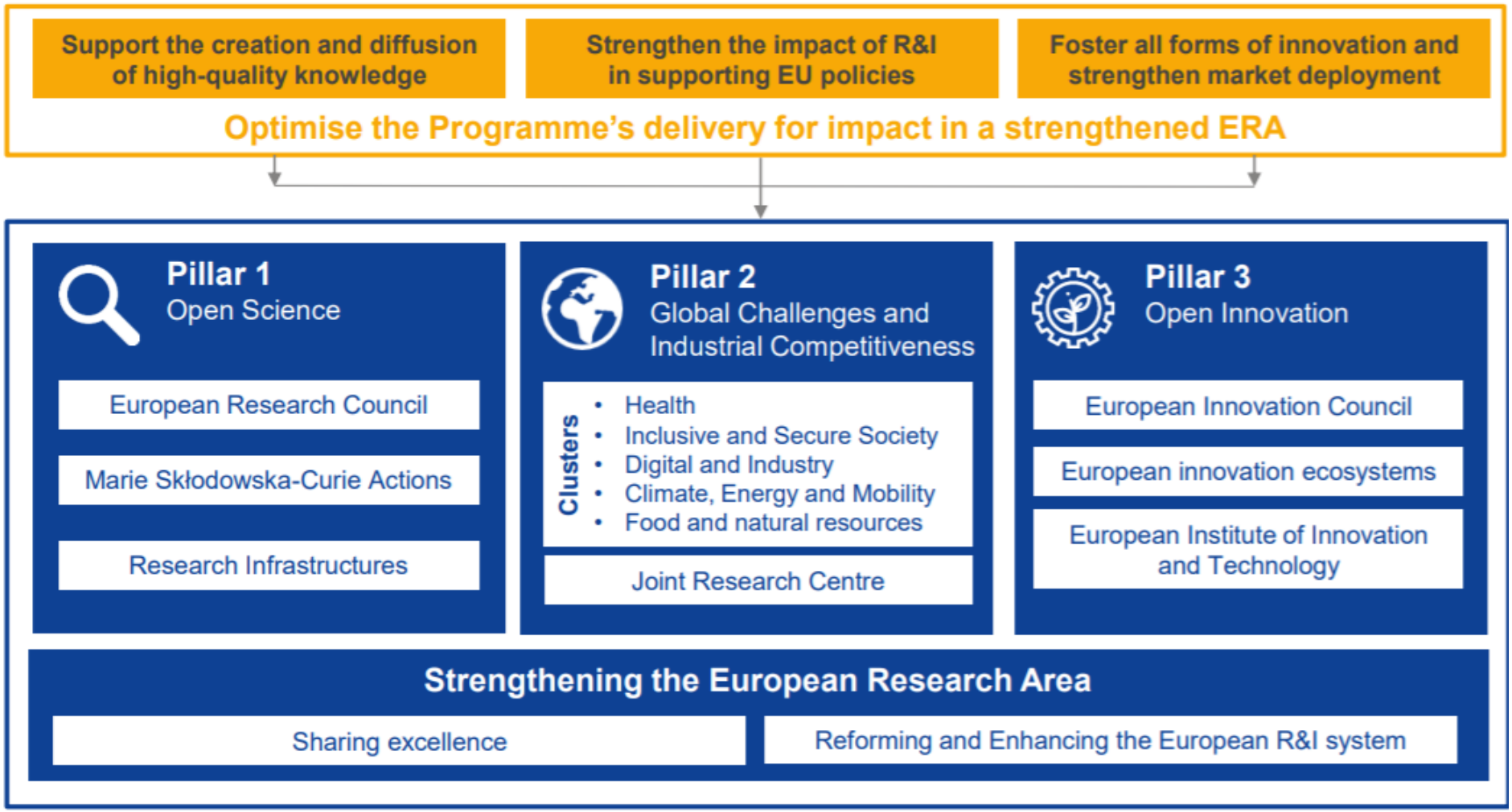
"The European Union could save €10.2 billion per year by using FAIR* and free open access research data."

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.

Horizon Europe: evolution not revolution

Specific objectives of the Programme



European Commission

JUNE 2018


Conference
Open Science
19 TO 20 MARCH 2019, BERLIN

#DontLeaveItToGoogle

How open infrastructures enable continuous innovation in the research workflow

Peter Kraker

Co-created with Maxi Schramm

Open Science Conference 2019

OPEN
KNOWLEDGE MAPS

Plan S

Making full and immediate Open Access a reality

FWF Der Wissenschaftsfonds.

ACADEMY OF FINLAND

Agence Nationale de la Recherche ANR

Science Foundation Ireland sfi For what's next

INFN Istituto Nazionale di Fisica Nucleare

Luxembourg National Research Fund

NWO Netherlands Organisation for Scientific Research

The Research Council of Norway

NATIONAL SCIENCE CENTRE POLAND

arrs SLOVENIAN RESEARCH AGENCY

FORTE: Swedish Research Council for Health, Working Life and Welfare

FORMAS

UK Research and Innovation

BILL & MELINDA GATES foundation

RIKSBANKENS JUBILEUMSFOND THE SWEDISH FOUNDATION FOR

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation

MINISTÈRE | STRATÉGIE | ENSEIGNEMENT SUPÉRIEUR | RECHERCHE | INNOVATION

Accueil > Recherche

RECHERCHE

Le Plan national pour la science ouverte : les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai, sans paiement

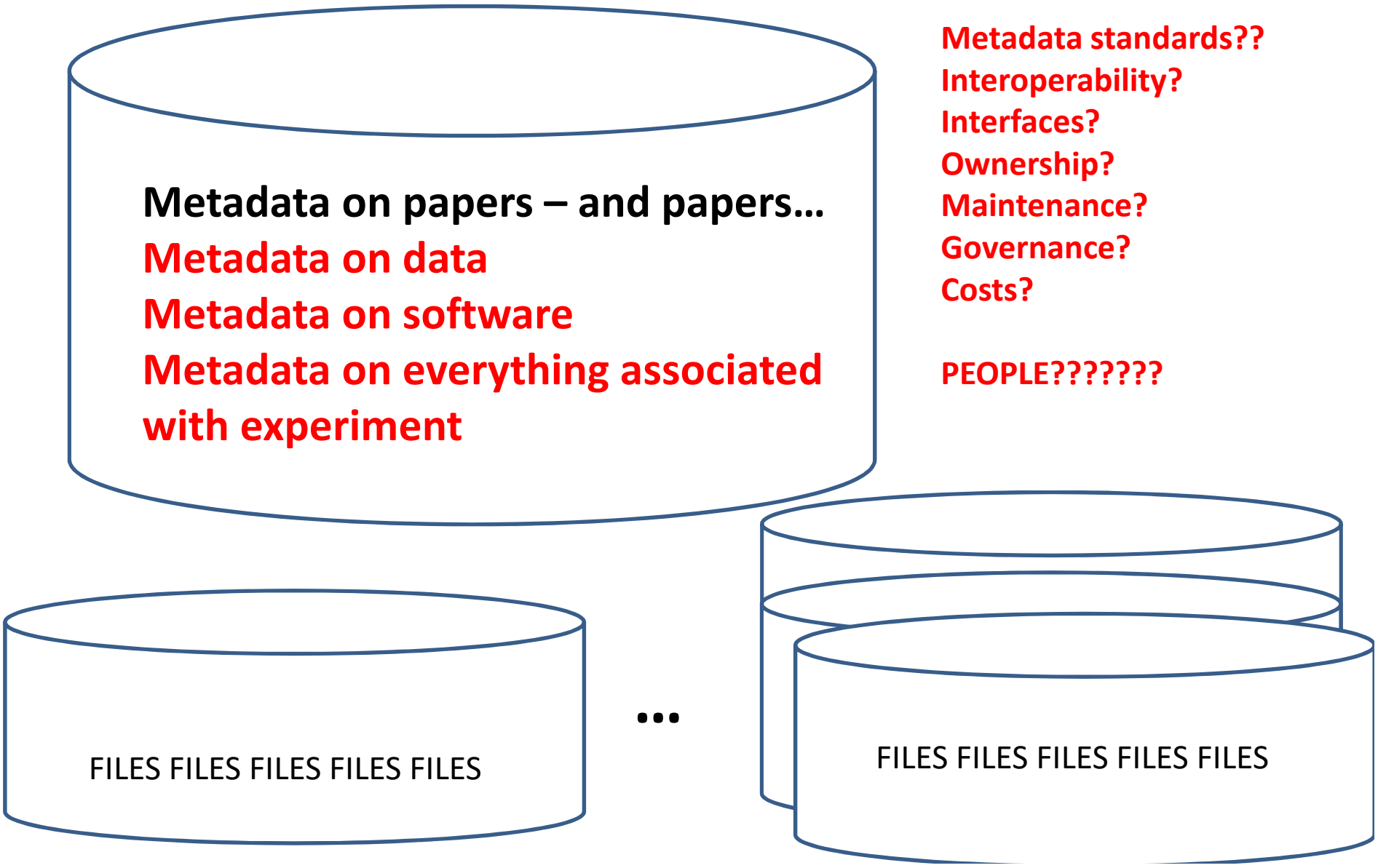
July 2018

Specifically mentions RDA and SoftwareHeritage

Why – The need for Open Science

- Validate research and advance science
- (Re)use = save resources
  - Data
  - Processes
  - People(?)
- (Re)use = improve, modify, accelerate scientific research
- Avoid fraud - transparency

OPEN SCIENCE - Challenges



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 - Big questions

WHAT IS (RESEARCH) DATA????

HOW TO SHARE????

FOR WHAT????

Data are the "glue"

Computer Science

Social sciences

Biological sciences

Agricultural sciences

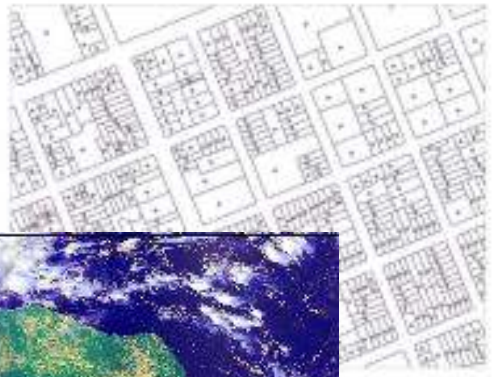
eAgriculture

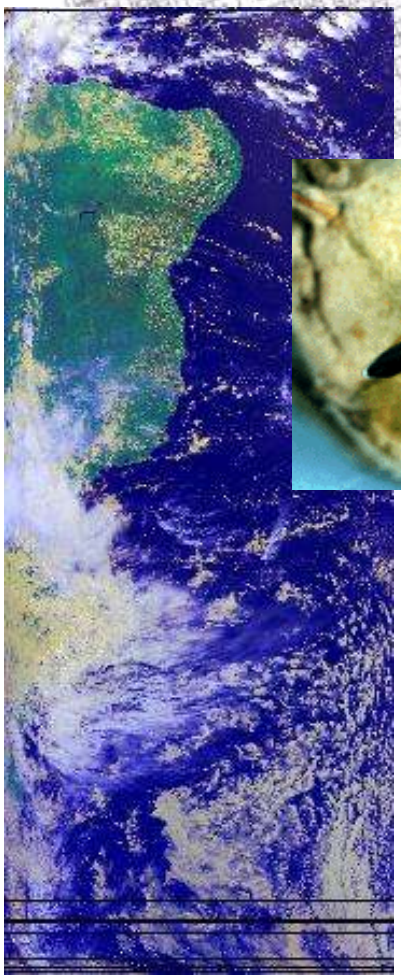
Environmental sciences

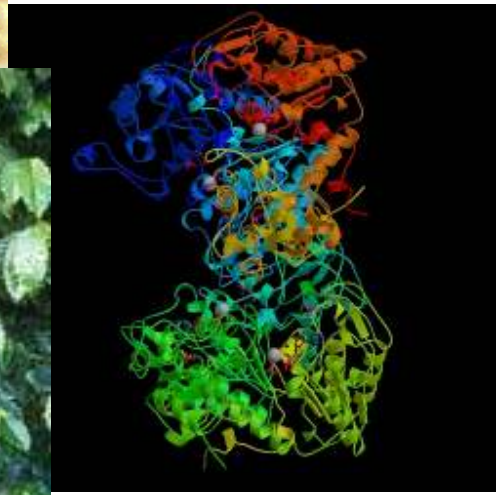
IOT

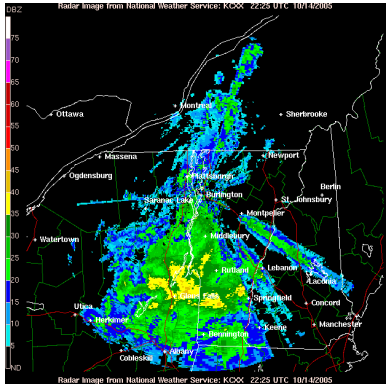
Food sciences

Data sources?



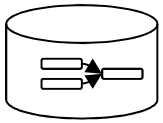


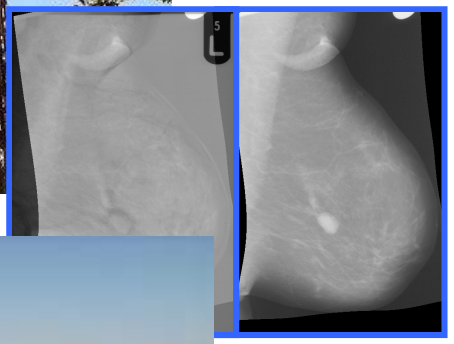


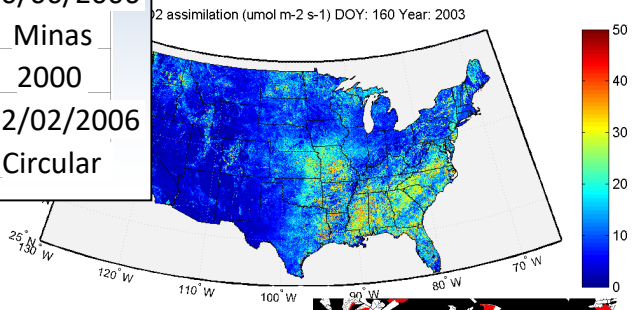


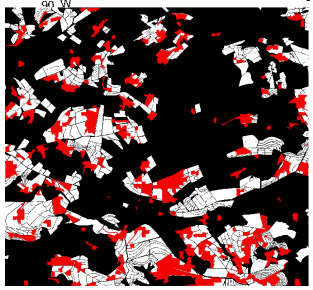
Data sources?

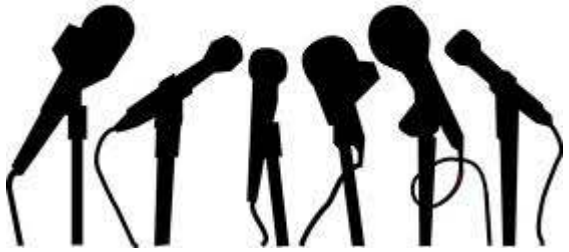
Table of Product Characteristics		
id	**Property name**	**Value**
MilkProd	productsrep	MilkA
MilkProd	quantity	10000
MilkProd	validity date	10/06/2006
CheeseProd	productsr	Minas
CheeseProd	quantity	2000
CheeseProd	validity date	12/02/2006
CheeseProd	shape	Circular

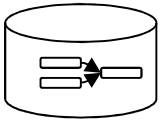








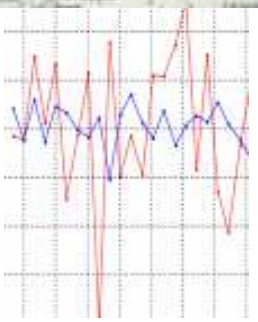




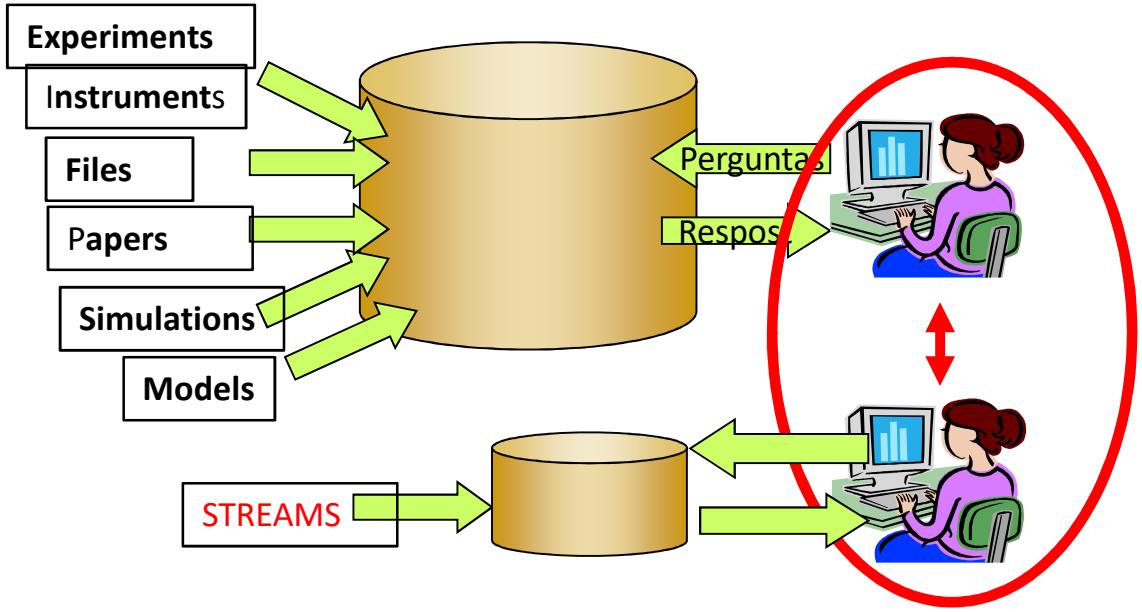
Data sources?

Research Data?

- Direct and indirect observations



How to Share – slide adapted from Jim Gray



Data driven-science

Big questions

WHAT IS (RESEARCH) DATA????

- [Digital only]
- **Any file associated with research**

HOW TO SHARE????

- **Via research data repositories**

FOR WHAT????

- **Advance research -> scientific, economic, social, cultural benefits**

Open, Sharing, Challenges few Mention

Updating and versioning

Curation and long term preservation

Visualization!!!!!!!!!!!!!!

Uses, reuses and mis-uses/ethics

Doors that are open and closed via choice of data to collect

- For xxx science to work, interpretation is needed (who are the “appropriate” experts?)

(Four) Challenges in Open Science

(1). What is data ?

(2). Lack of common/consensual infrastructures

(3). Finding and identifying

(4). Understanding

Infrastructure(s) = 260M + 12B euros/yr

Home > Research and Innovation > Strategy > Goals of research and innovation policy > Open Science >

European Open Science Cloud (EOSC)

This is a cloud for research data in Europe. Background, policy information, events and publications related to the EOSC

Home | Open Access | European Open Science Cloud | Open Science Policy Platform | Groups | Open Science Monitor

The reports "Prompting an EOSC in practice" and "Turning FAIR into reality" have been published

20 November 2018

In the perspective of the launch of the European Open Science Cloud (EOSC) implementation phase 2018-2020, two important reports are being published by the Commission that constitute major sources of strategic orientations and concrete actions for the new EOSC governance structure:

- Prompting an EOSC in practice
  Report of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC 2nd HLEG)

Events

29 January 2019, Brussels, Belgium - 2019 Infoday, Ensuring Excellent Research by Investing in Researchers Talents, Skills and Career Development

5 February 2019, Brussels, Belgium - Open access, Scientific Publishing and Plan S

See all Events

@ Claudia Bauzer Medeiros

- Initiated 2016
- Already 173 repositories from African countries
- “Gateway” to scientific resources

Finding and Identifying

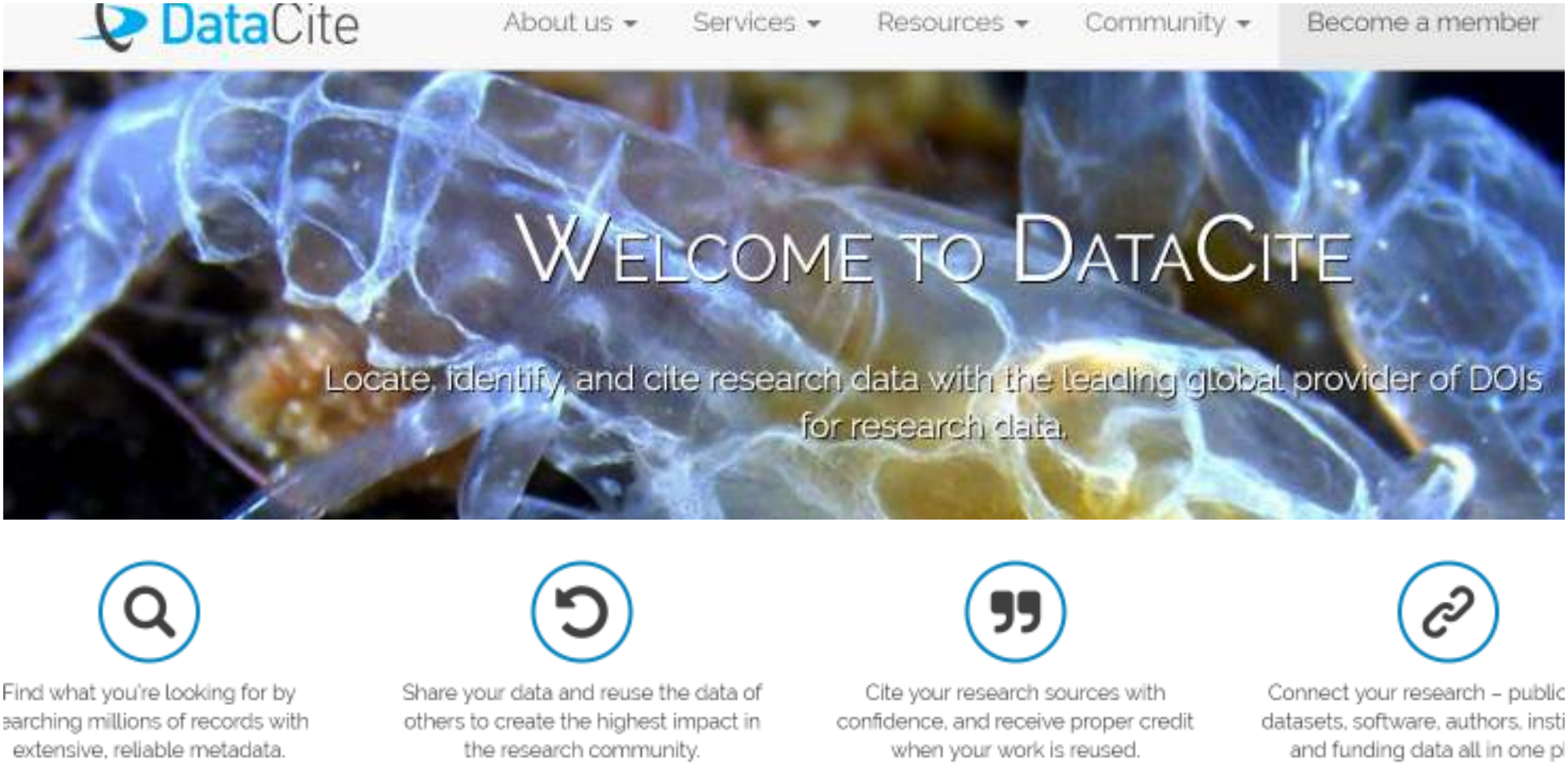
How to search/find (specification ?)

Where?

What to publish (everything vs ethics/privacy)

Unique id?

HOW??? Datacite.org
(Find, share, cite, connect)



@ Claudia Bauzer Medeiros

(HOW???) Handle.Net

Understanding

How to understand what you find?

How to reuse it?

?Metadata? Ontologies?

Everything is domain dependent

project dependent

Open Science requires FAIR Data

- Findable
- Accessible
- Interoperable
- Reusable

- **??? Have you fairicized your data???**

Open Science requires FAIR data

@ Claudia Bauzer Medeiros

Outline

Open Science

- Definitions
- Challenges

4 Challenges in (Research Data) Sharing

Open Data at FAPESP

- **Data Management Plans**
- **Network of open research data repositories**

Springer Nature – 5 essential factors for data sharing = Apr 2019

1. **Clear policy**: from funders, institutions, journals/publishers, and research communities themselves.
2. **Better credit**: to make data sharing worth a researcher's time.
3. **Explicit funding**: for data management and data sharing, as well as data publishing.
4. **Practical help**: for organising data, finding appropriate repositories, and provision of faster, easier routes to share data.
5. **Training and education**: to answer common questions from researchers on data sharing and to help build skills and knowledge.

@ Claudia Bauzer Medeiros

What are DMP

- “LIVE” text that answers 2 questions
  - Which data will my research produce/handle
  - How do I intend to share/publish them

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

UNICAMP

UFSCAR

UNESP

UNIFESP

USP

ITA

UFABC

OUTRAS

RETRIEVAL

NINE compulsory metadata fields

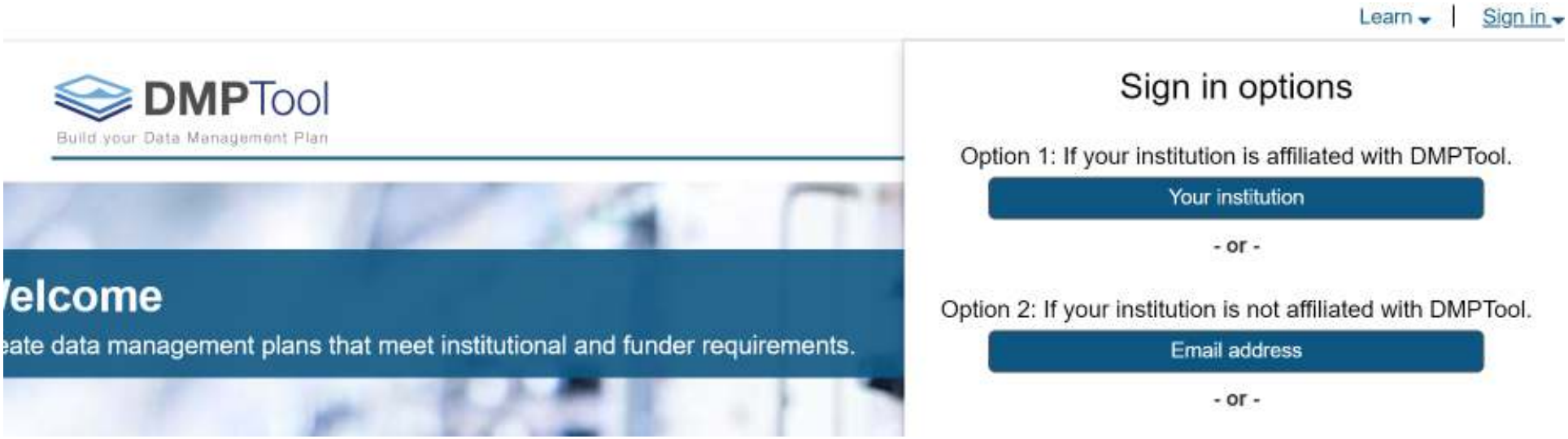
ID	Type	Description
1	dc.title	Project title
2	dc.subject	Keywords
3	dc.description	Abstract
4	dc.contributor.author	Author (ORCID)
5	dc.identifier.uri	File id
6	dc.description.sponsorship	Funding agencies
7	dc.description.sponsorshipId	Project numbers

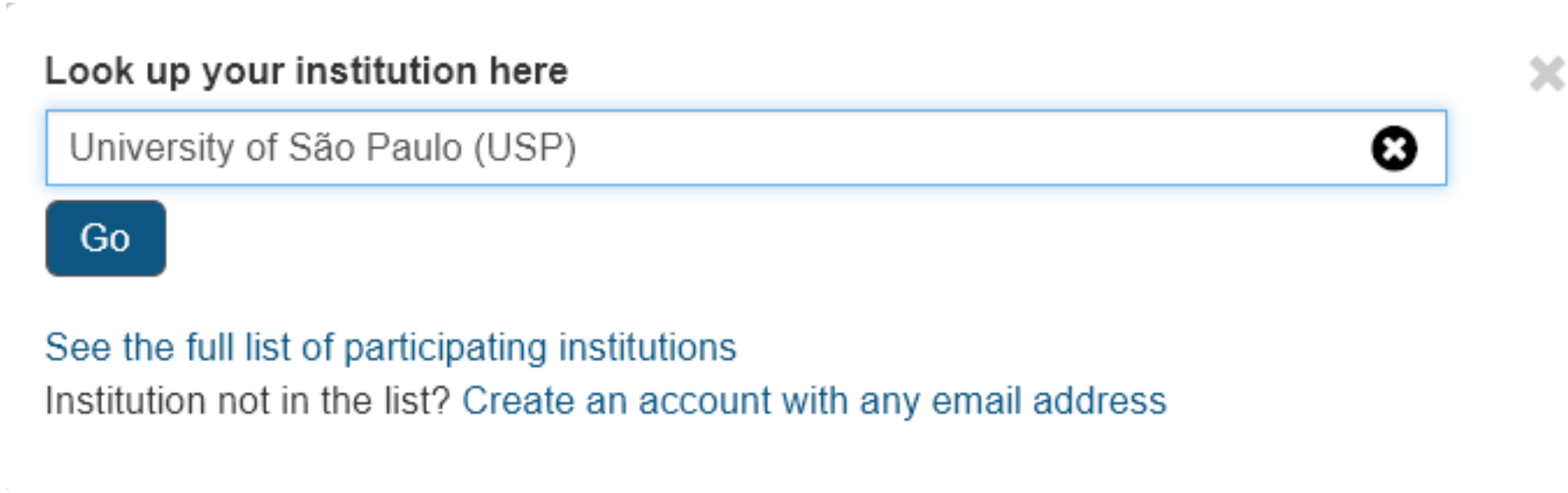
8 dc.type File type (software, others

9 dc.identifier File id (handle)

HOW TO UNDERSTAND DATA MANAGEMENT REQUIREMENTS !!!!!!!!!!!!!!!!!

Preparing DMP - dmptool.org


Learn | Sign in
DMPTool
Build your Data Management Plan
elcome
eate 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
University of São Paulo (USP)
Go
See the full list of participating institutions
Institution not in the list? Create an account with any email address

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

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?

Some remarks

- Nothing is for free
- Hardware, software, “dataware”, PEOPLEWARE
- International collaboration will increasingly require open science considerations
- OPEN does not mean FAIR

Open, Sharing, Challenges few Mention

Updating and versioning

Curation and long term preservation

Visualization!!!!!!!!!!!!!!

Uses, reuses and mis-uses/ethics

Doors that are open and closed via choice of data to collect

- For xxx science to work, interpretation is needed (who are the “appropriate” experts?)

Main take-aways

- What is open science?
- What does OPEN mean?
- What is data?
- Why should I care?

What is data?

Any object in digital format, static or dynamic

What is the meaning of OPEN?

OPEN metadata in public repository

Why should I care?

Worldwide collaboration
Financing opportunities
Visibility of my work
Accelerating my research

Outline

(4) Challenges in (Research Data) Sharing

Open Science -> Open MetaData

- Definitions (Open Access, Open Data)
- Challenges (Find, Identify, Understand)

Open Data at FAPESP

- Data Management Plans
- Network of open research data repositories

OBRIGADA

Open Source Data Repository Sw

At a glance – CIHR Health Research and Health-Related Data Framework and Action Plan

News

New data re-use prizes help unlock the value of research

A new method for ethical data science

by Danil Mikhailov

Open access and Plan S: how Wellcome is tackling four key concerns

by Robert Kiley, David Carr

http://www.orfg.org/incentivization-blueprint

OBRIGADA