TÉCNICO LISBOA

TÉCNICO LISBOA

Data sharing and management

Pedro S. F. Mendes

Alessandra Soro, Yasemin Türkyilmaz-van der Velden, Connie Claire, Paula Martinez Lavanchy

TÉCNICO LISBOA

Open Science

The motivation

Periodic table: a data sharing tale

Reihen	Gruppo I. — R^1O	Gruppo II. — RO	Gruppo III. — R^2O^3	Gruppo IV. RH^4 RO^2	Gruppo V. RH^3 R^2O^5	Gruppo VI. RH^2 RO^3	Gruppo VII. RH R^2O^7	Gruppo VIII. — RO^4
1	H=1							
2	Li=7	Be=9,4	B=11	C=12	N=14	O=16	F=19	
3	Na=23	Mg=24	Al=27,3	Si=28	P=31	S=32	Cl=35,5	
4	K=39	Ca=40	—=44	Ti=48	V=51	Cr=52	Mn=55	Fe=56, Co=59, Ni=59, Cu=63.
5	(Cu=63)	Zn=65	—=68	—=72	As=75	Se=78	Br=80	
6	Rb=86	Sr=87	?Yt=88	Zr=90	Nb=94	Mo=96	—=100	Ru=104, Rh=104, Pd=106, Ag=108.
7	(Ag=108)	Cd=112	In=113	Sn=118	Sb=122	Te=125	J=127	
8	Cs=133	Ba=137	?Di=138	?Ce=140	—	—	—	— — — —
9	(—)	—	—	—	—	—	—	
10	—	—	?Er=178	?La=180	Ta=182	W=184	—	Os=195, Ir=197, Pt=198, Au=199.
11	(Au=199)	Hg=200	Tl=204	Pb=207	Bi=208	—	—	
12	—	—	—	Th=231	—	U=240	—	— — — —

The Crisis of Science

The New York Times *18/12/2019*

There's No Winter Break From 'Publish or Perish'

"Welcome to the co-author's party! You're number twenty-one!"

Published in: Mohamed Gad-el-Hak; *Physics Today* **57,** 61-62 (2004)
DOI: 10.1063/1.1712503

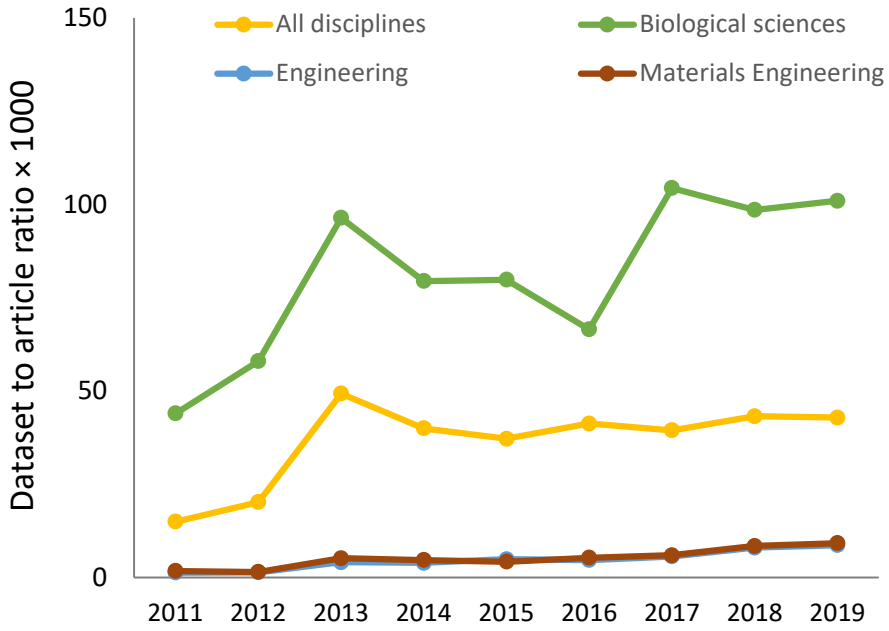
The Economist

Unreliable research

18/10/2013

Trouble at the lab

Today's data sharing



Dimensions.ai, 16/04/2020
doi.org/10.5281/zenodo.3939624

"The catalyst is not working, Mr. Watson!"

Is there a revolution on the way?

Scientists also want a change

Let's change what we value in research.

Sign DORA

LERU PUSHING THE FRONTIERS OF INNOVATIVE RESEARCH

ROYAL SOCIETY OF CHEMISTRY

Governments and funders are pushing for it

RRI Tools

zenodo

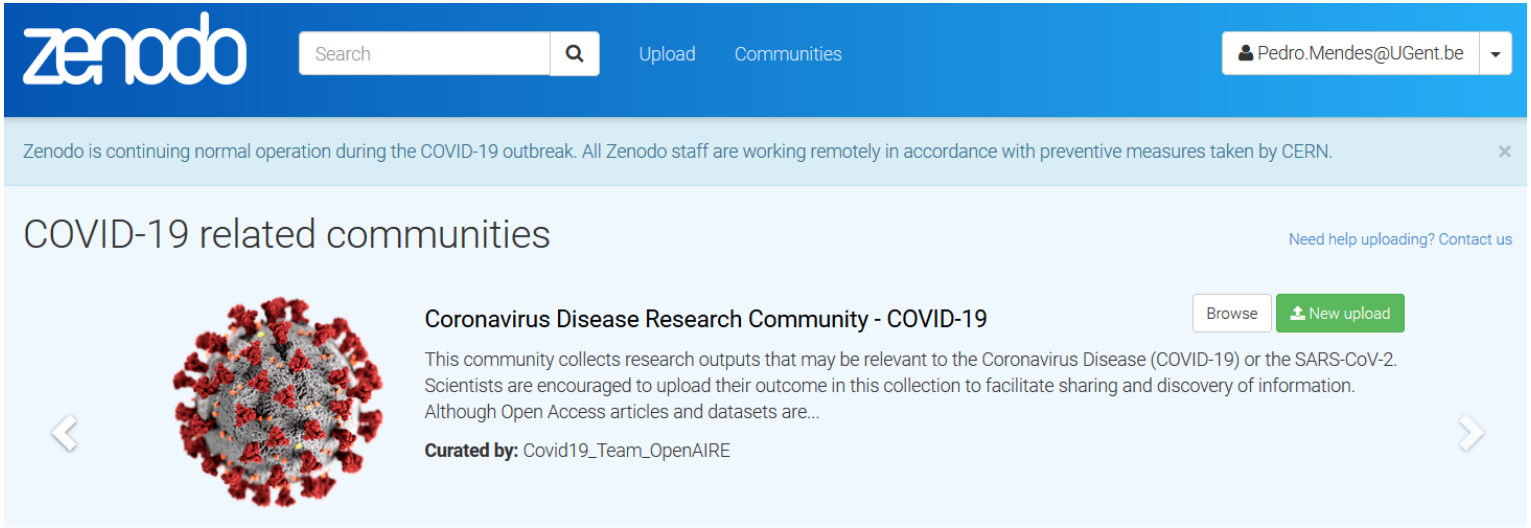
fwo

Data Management Plan (DMP)

What is Open Science about?

break barriers to knowledge

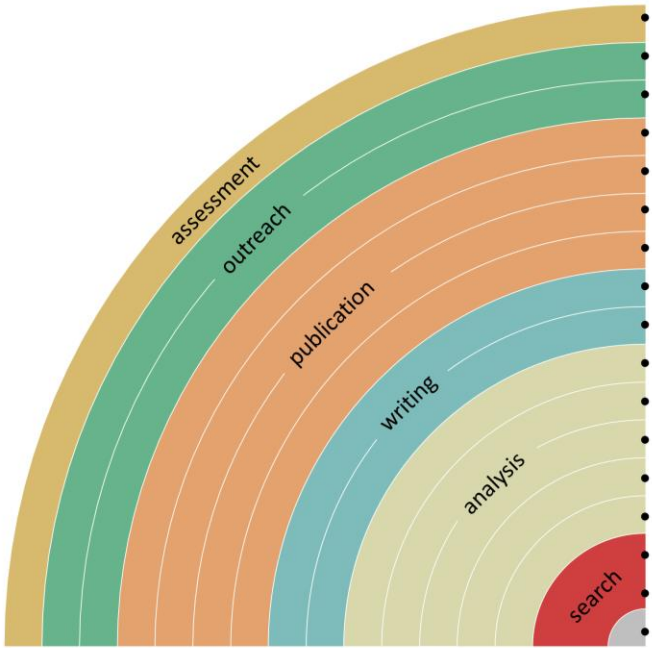
improve science collaboration



Why Open Science?

The rainbow of Open Science

When?



DOI: 10.5281/zenodo.1147025

TÉCNICO LISBOA

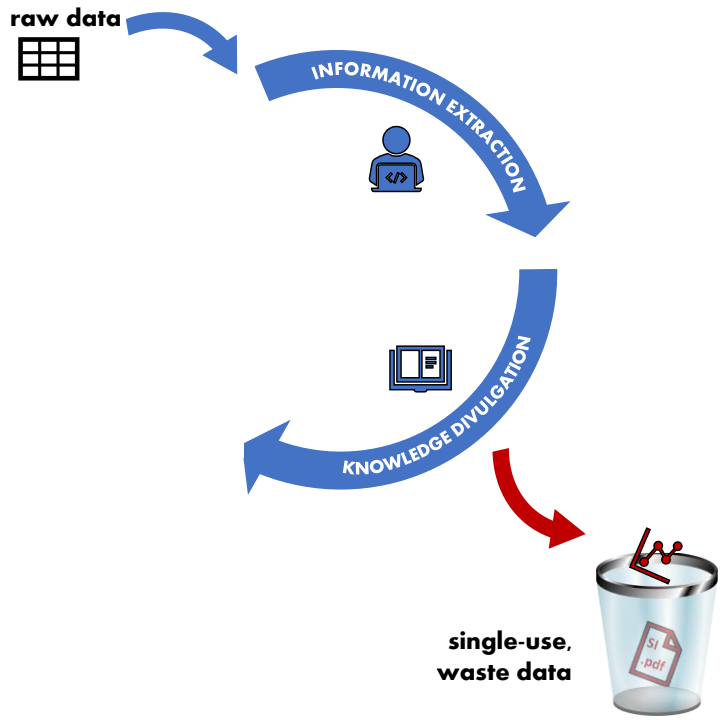
Data management

The bookkeeping

Data life cycle

P.S.F. Mendes et al., ChemCatChem, 2021

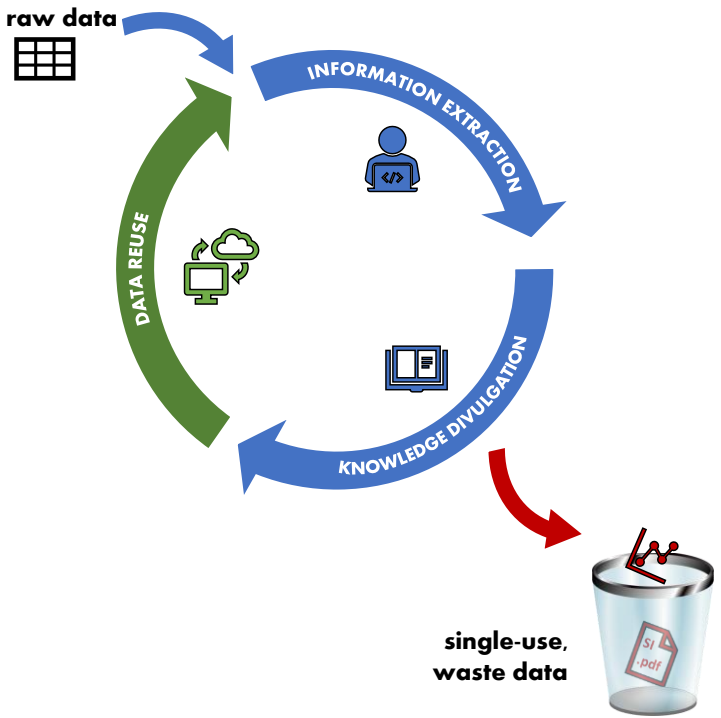
towards data circularity



Data life cycle

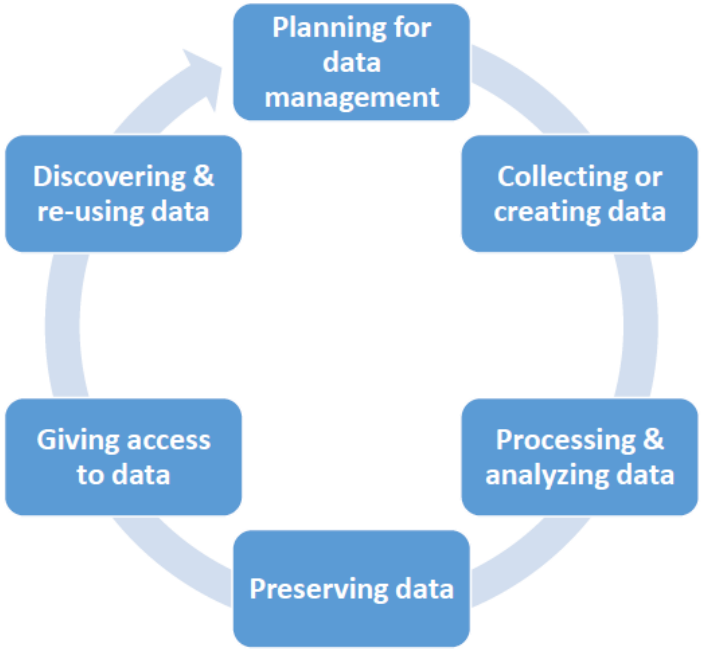
P.S.F. Mendes et al., ChemCatChem, 2021

towards data circularity



What is data management about?

The active management of data throughout their lifecycle



https://www.slideshare.net/OpenAccessBelgium/20181024-oa-weekrdmmyriammertens

DMPs: key topics

- Types of data
- Contextual details (metadata)
- Storage, backup and security
- Archiving and long-term access
- Access and sharing
- Policies for re-use
- Roles and plan oversight

https://datamanagement.hms.harvard.edu/plan/data-management-plans

Tools

Online tools to create DMPs

- https://dmptool.org/
- https://dmponline.dcc.ac.uk/

Info and checklists

- https://libguides.wustl.edu/drmr/dataprep
- https://datamanagement.hms.harvard.edu/plan/data-management-plans

TÉCNICO LISBOA

Data sharing

The practice

Sharing Data I: how?

FINDABLE

Data is available on the internet

ACCESSIBLE

Access conditions are understandable

INTEROPERABLE

Formats are standardized

REUSABLE

Data is well-described

https://www.nature.com/articles/sdata201618

It should be possible for others to discover your data. Rich **metadata** should be available online in a **searchable resource**, and the data should be assigned a **persistent identifier** (e.g. DOI, Handle…).

Data and metadata should be conform to **recognized formats and standards** to allow them to be combined & exchanged (file formats, metadata schemas, controlled vocabularies, keywords, ontologies, qualified references & links to other related data).

Findable Accessible Interoperable Reusable

It should be possible for humans and machines to gain access to your data (**retrievable by their PID** using a **standard protocol** such as http), under specific conditions or restrictions where appropriate (**authentication and authorization** steps if necessary). There should be **metadata, even if the data aren't accessible**.

Lots of **documentation** is needed to support data interpretation and reuse. It is clear how, why & by whom data were created & processed (**provenance**). The data should conform to **community norms** and be **clearly licensed** so others know what kinds of reuse are permitted.

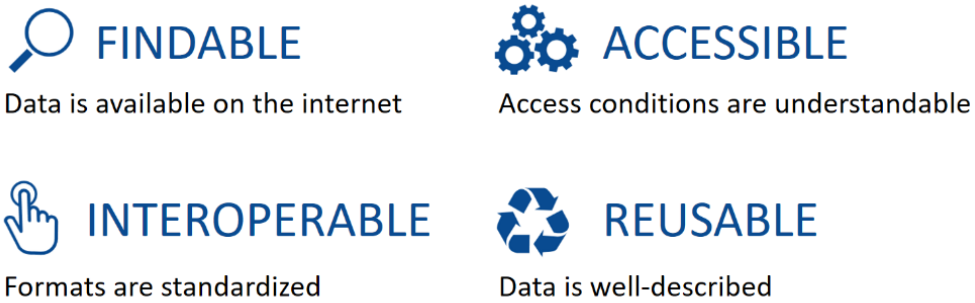
TÉCNICO LISBOA

How is data shared today and what can we learn from it?

Drill 1

Sharing Data I: how?

- FAIR principles



- Community accepted standards: check at https://fairsharing.org/

Sharing Data I: get your data ready to share

Findable

✓ Describe the content of the files (metadata): the user can know the content without opening the files

The data file contains the glycerol conversion and selectivities over time for the studied catalysts, in combination with N_2-sorption data of those catalysts. In addition, an overview of literature data on glycerol hydrogenolysis is given (Fig 1 in the manuscript).

A Bouriakova et al. (submitted)

- 10wt% Cu/-Al2O3, indicated as Cu
- 1wt%Ba-10wt% Cu/-Al2O3, indicated as Ba-Cu
- 1wt%Ce-10wt% Cu/-Al2O3, indicated as Ce-Cu
- 1wt%Cs-10wt% Cu/-Al2O3, indicated as Cs-Cu
- 1wt%La-10wt% Cu/-Al2O3, indicated as La-Cu

Interoperable: machine readable

✓ Tabular data = Tabular format (preferably as e.g. .csv)

✓ No (paid) software requirements

https://www.openaire.eu/data-formats-preservation-guide

https://5stardata.info/en/

Sharing Data II: get your data ready to share

- Reusable (beyond the original purpose)
  - ✓ Raw data: to enable re-processing of your data (e.g. to calculate different metrics)
  - ✓ Hypothesis & calculation details (metadata): to enable reproducing the data treatment (e.g. calibration factors, metrics formulae)
  - ✓ Experimental procedures & operating conditions (metadata): to enable reproducing the experimental results

Sharing Data II: where?

Find the best repository for your data

Re3Data

Generic

Field-specific

Materials Informatics community

HTE JCAP

- Open: data must be accessible to anyone without any barriers
- Findable: is there a well-known repository for this type of data or for the community in our field?
- Long-term data storage & accessibility: for how long will my data be stored and accessible?
  - Choose a license

Sharing Data III: share the shared data

- Cite your data via a DOI

- Cross-reference publications and data

  - ✓ Data availability statement in paper

A Bouriakova et al. (submitted)

Data Availability Statement

The raw data corresponding to Fig. 1 and Fig. 3 are openly available in Zenodo at http://doi.org/10.5281/zenodo.3739138.

  - ✓ Reference to the pa

Alexandra Bouriakova; Pedro S. F. Mendes; Benjamin Katryniok; Jeriffa De Clercq; Joris W. Thybaut

This repository includes experimental data associated with the following publication: A. Bouriakova et al., "Dopant induced stabilization of alumina supported copper – impact on the catalytic performance in the hydrogenolysis of glycerol to 1,2-propanediol", submitted to Catalysis Communications in April 2020

TÉCNICO LISBOA

Get your data ready to share

Drill 2

Take-home messages

- We need to fix science: Open (& Slow) Science
- Open Science is about breaking barriers to knowledge and improving collaboration in science (it's on the way and it's not perfect)
- Share as much as you are comfortable but share it!
- Make a difference by sharing!
  - Already published, complete datasets
  - Open data instead of supporting information
  - Open source code
  - Preprints of new papers

Acknowledgments

material pioneers TU Delft 4TU.ResearchData OSCD Open Science Community Delft

TÉCNICO LISBOA

GHENT UNIVERSITY

Laura Pirro

Sébastien Siradze

Joris Thybaut

Alexandra Bouriakova

BOF: PDO/2018/0019

European Research Council

TÉCNICO LISBOA