

The importance of FAIR data management and data stewardship in Open Science

28 Oct 2021, *Università del Piemonte Orientale*
(webinar)



Who am I



Dr. ir. Shalini Kurapati

- Open Science fellow at PoliTO (Adjunct) since Oct 2019
 - **Training and Awareness activities on RDM**
 - **Advice (as much as I can) on all data related matters of Open Science**
 - **Task Force Open Science (Cesaer, Unite!), ICDI collaboration**
- Co-founder and CEO of clearbox.ai, MLOps and DataOps for responsible AI adoption
- Certified Information Privacy Professional/Europe (IAPP) for GDPR

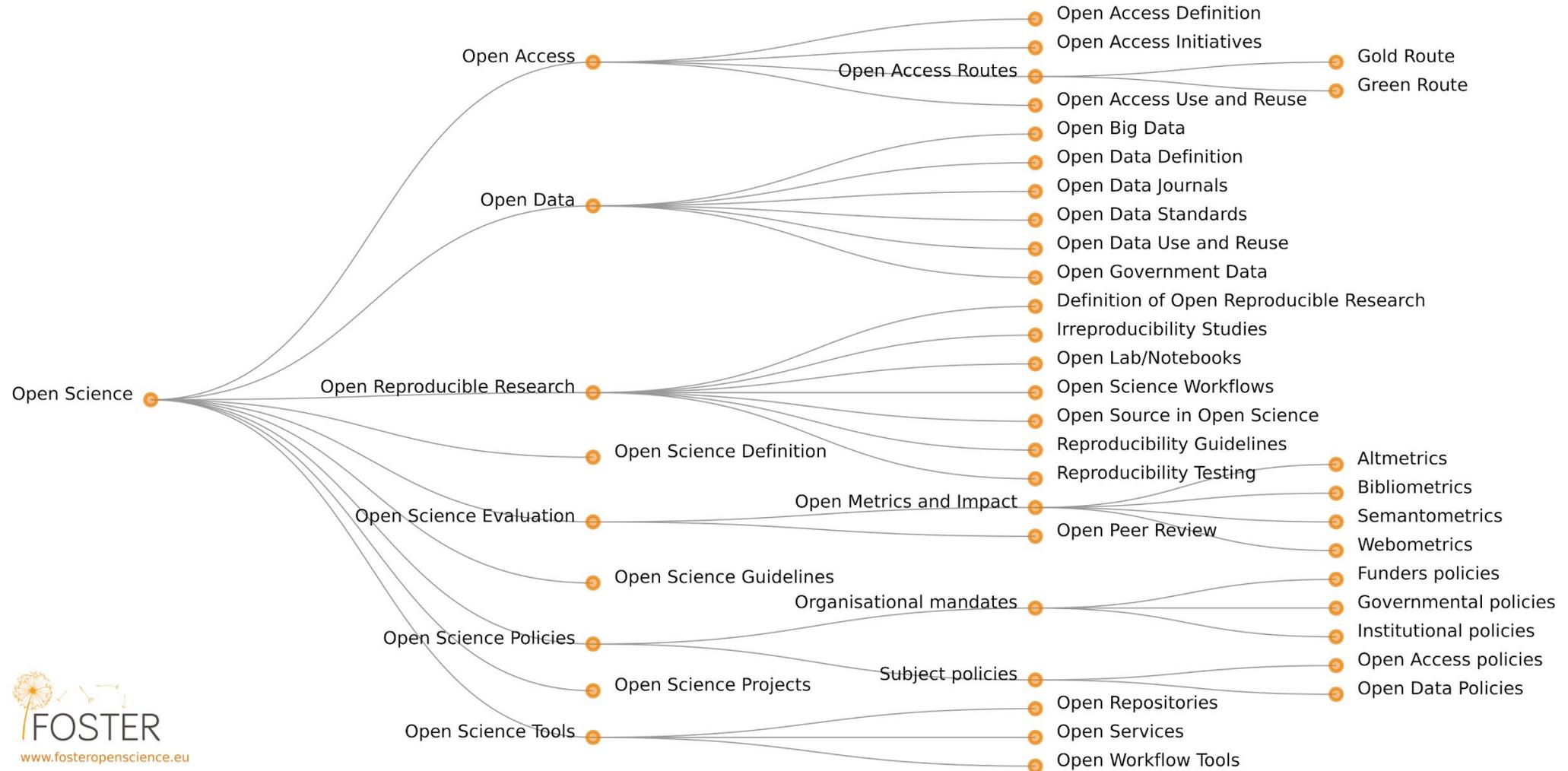
Past:

- 9 years research experience at the intersection of Technology and Policy Management
- Department advisor on RDM, Data stewardship project TU Delft, Netherlands

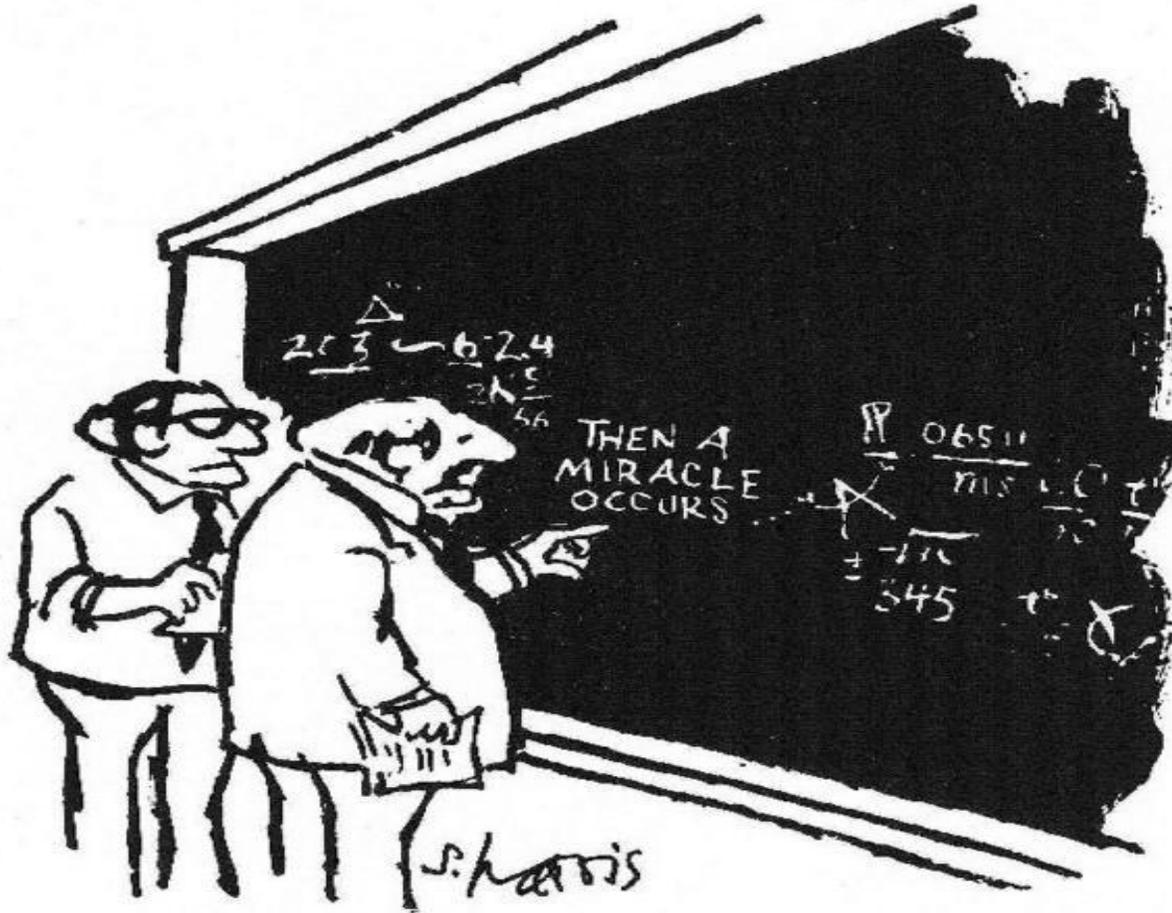
Definition of open science

There is no single doctrine or paper that definitively captures open science. Rather, open science can be defined as a **set of practices** that increase the **transparency** and **accessibility of scientific research** (van der Zee & Reich, 2018).

Open Science Taxonomy



Science \neq Miracles



"I think you should be more explicit here in step two."

Open Science means:

- Evidence based results,
- Transparency, reproducibility, research rigour
- Validation and verification
- Dissemination and access
- And all other things that basically define science.

Open science is nothing new, it's just science

**OPEN SCIENCE:
JUST
SCIENCE
DONE RIGHT**

<https://zenodo.org/record/1285575>

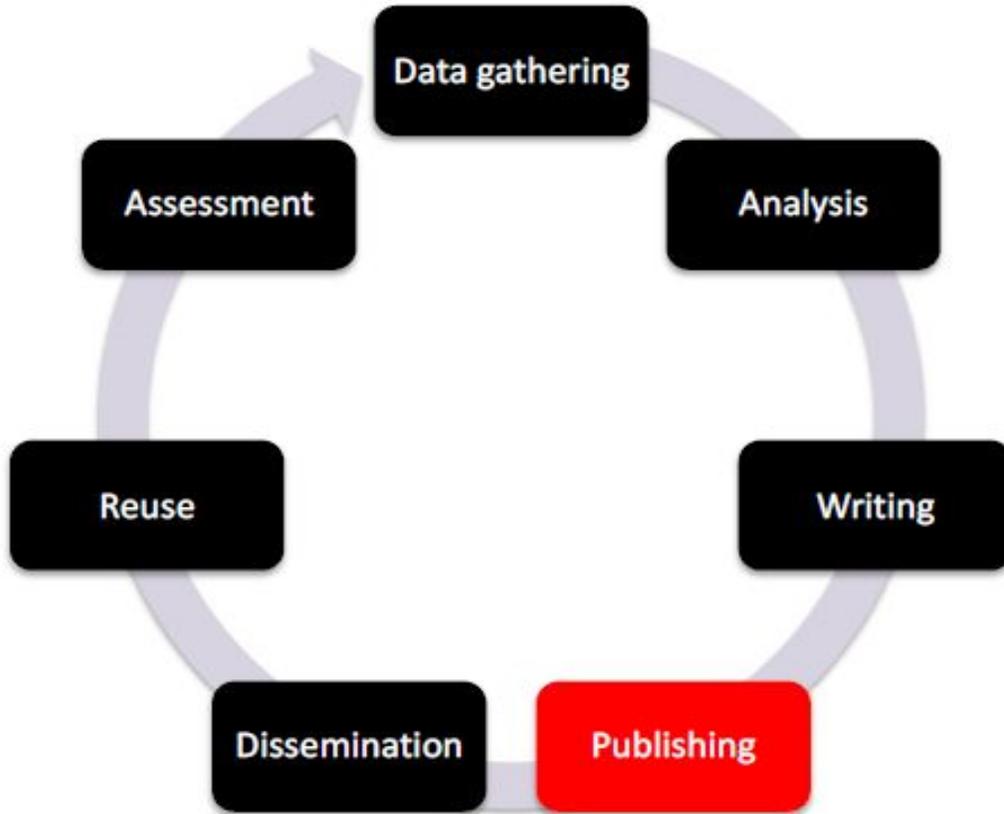


If open science is just science

Why is everyone talking about it now!



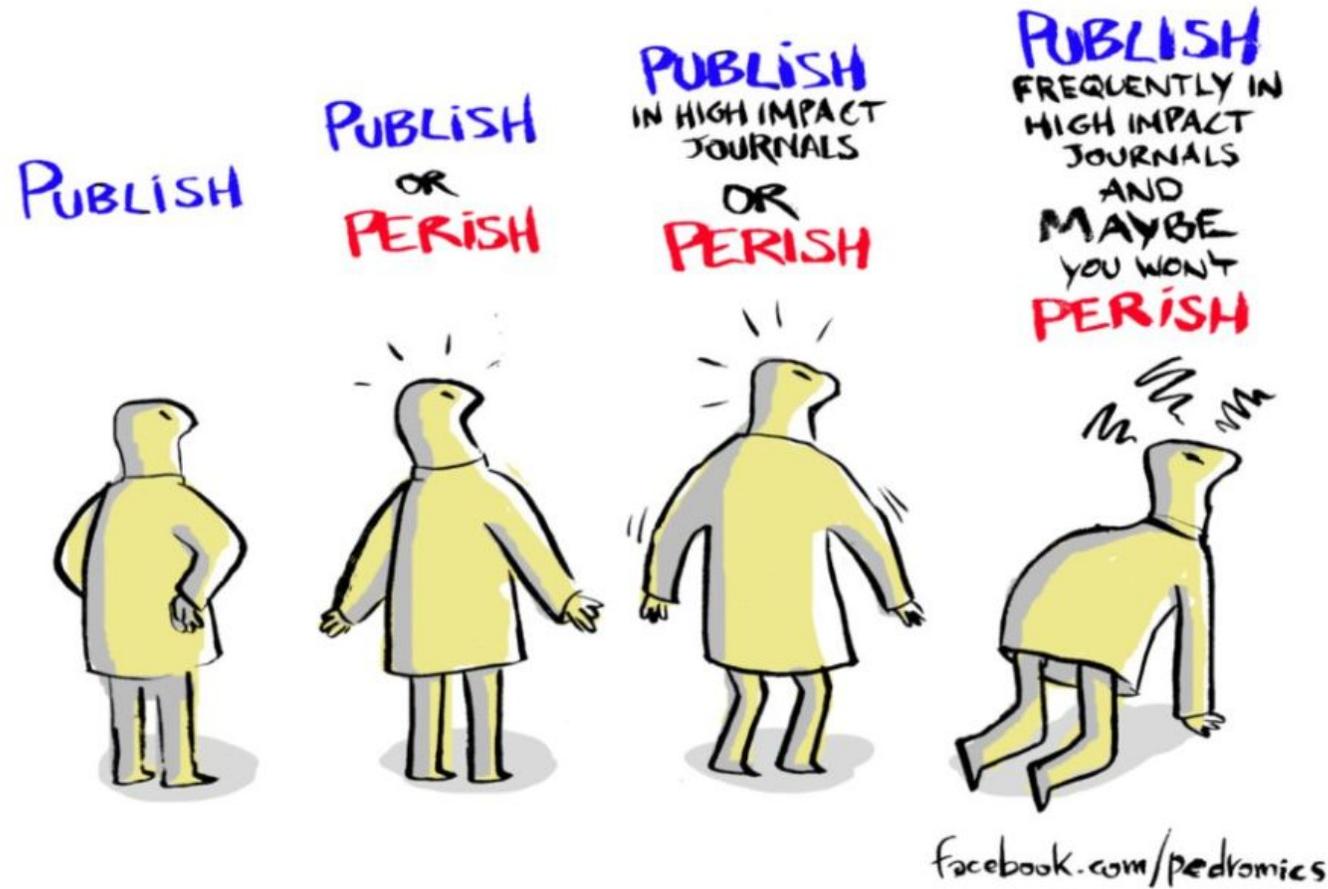
Only Publishing counts.. so far..



The only thing that counts in academia is publication of novel results in high impact journals

<https://www.repository.cam.ac.uk/handle/1810/276106>

THE EVOLUTION OF ACADEMIA



Copyright:

<https://velica.deviantart.com/art/Publish-or-Perish-645355248>

You see the problem?

19th century
scientist

I must find the
explanation for this
phenomenon in order
to truly understand
Nature...



21st centurt
~~scientist~~
academic

I must get the
result that fits my
narrative so I can
get my paper into
Nature..



facebook.com/pedromics

<https://www.euroscientist.com/open-scientists-in-the-shoes-of-frustrated-academics-part-i-open-minded-scepticism/>

Some consequences: In extreme cases

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

Ewen Callaway

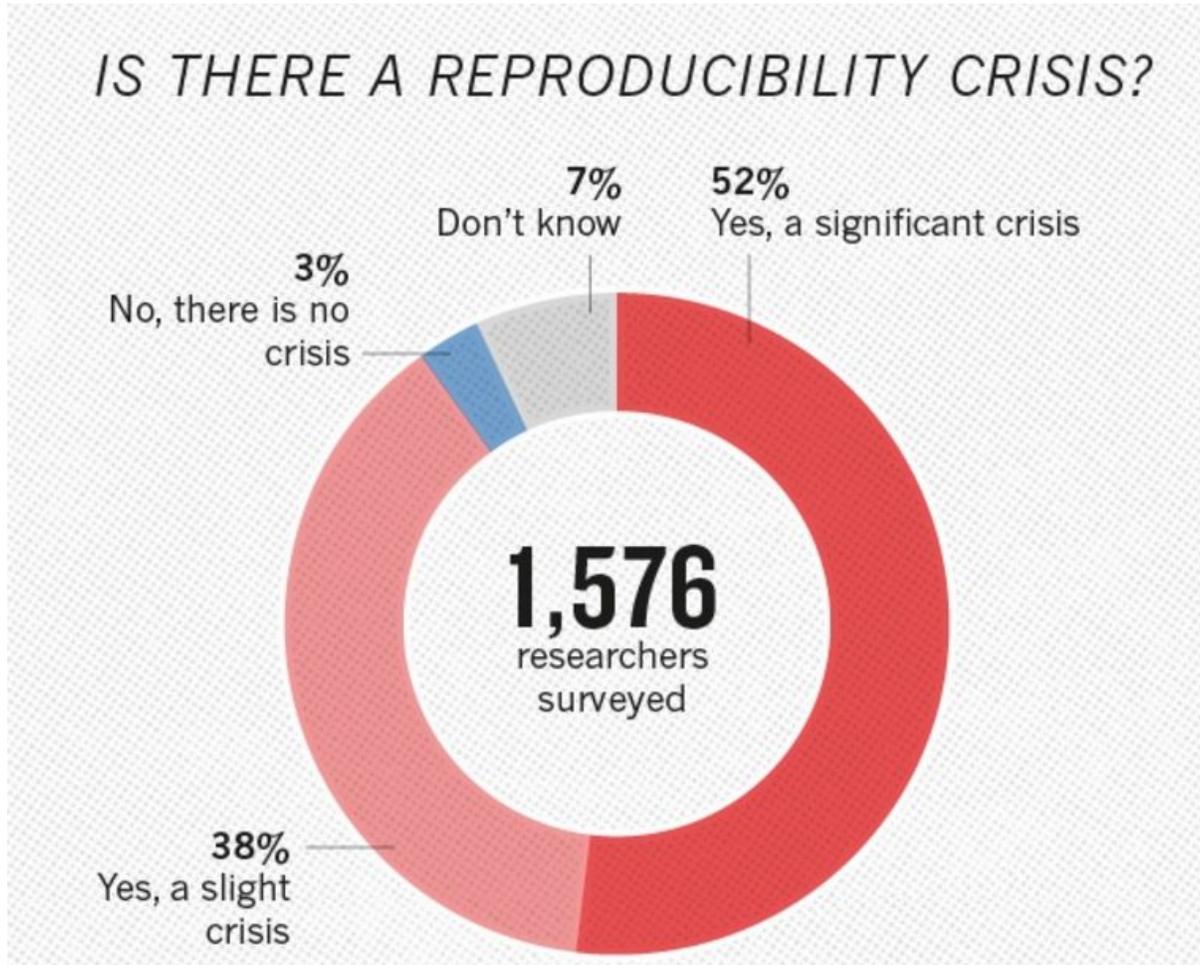
When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report (go.nature.com/tqmp5c) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.



Fraud is not the main issue..

In everyday scientific practice, fraud is minimal, but the main issue is the reproducibility

The reproducibility crisis



<https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

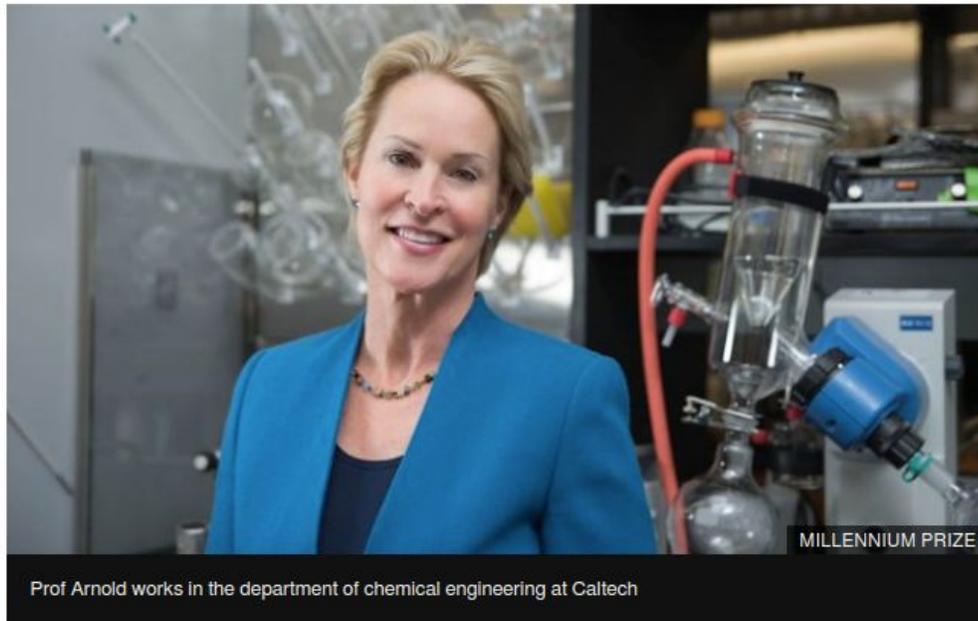
Happens even to the best of scientists

Nobel Prize-winning scientist Frances Arnold retracts paper

🕒 3 January 2020

f 🗨️ 🐦 ✉️ Share

Nobel Prize



“It has been retracted because the results were not reproducible, and the authors found data missing from a lab notebook.

Reproduction is an essential part of validating scientific experiments. If an experiment is a success, one would expect to get the same results every time it was conducted.”

<https://www.bbc.com/news/world-us-canada-50989423>

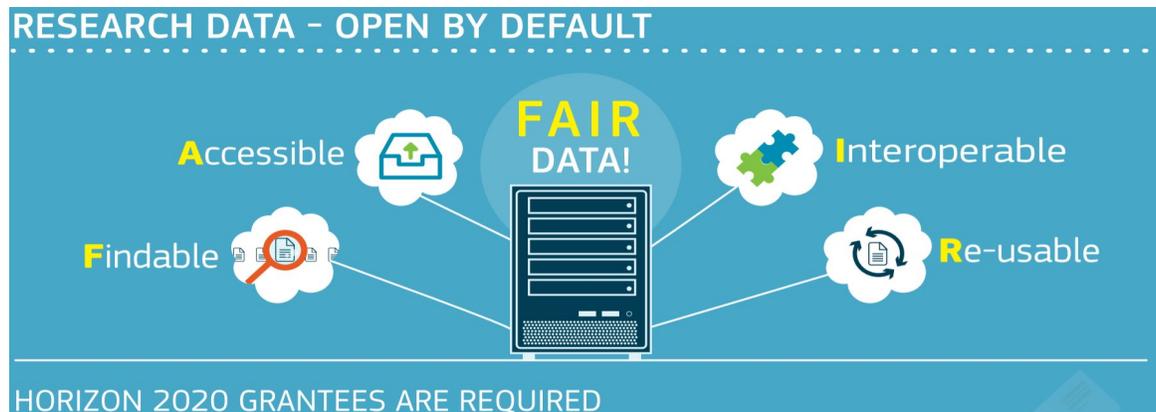
Reasons for the crisis



- Selective reporting
- Pressure to publish
- Insufficient supervision and training
- **Supporting data / methods / code not available**

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, [online] 533(7604), pp.452-454. Available at: <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970> [Accessed 26 Apr. 2018].

Funding bodies are pushing for open science, focus on FAIR data



More are following, nationally and regionally

Horizon Europe & Open Science



Open Science across the programme

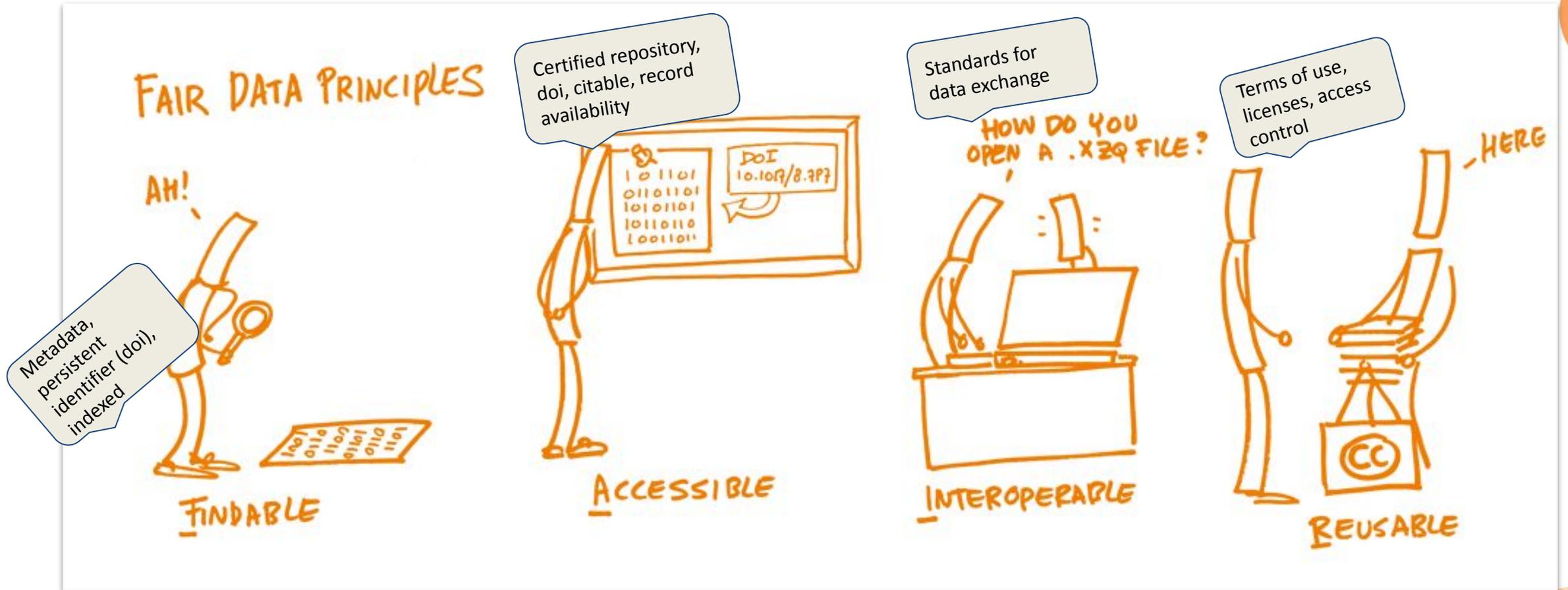
Open Science

Better dissemination and exploitation of R&I results and support to active engagement of society

Mandatory Open Access to publications: beneficiaries shall ensure that they or the authors retain sufficient intellectual property rights to comply with open access requirements

Open Access to research data ensured: in line with the principle "as open as possible, as closed as necessary"; Mandatory Data Management Plan for FAIR (Findable, Accessible, Interoperable, Re-usable) and Open Research Data

What is FAIR data?

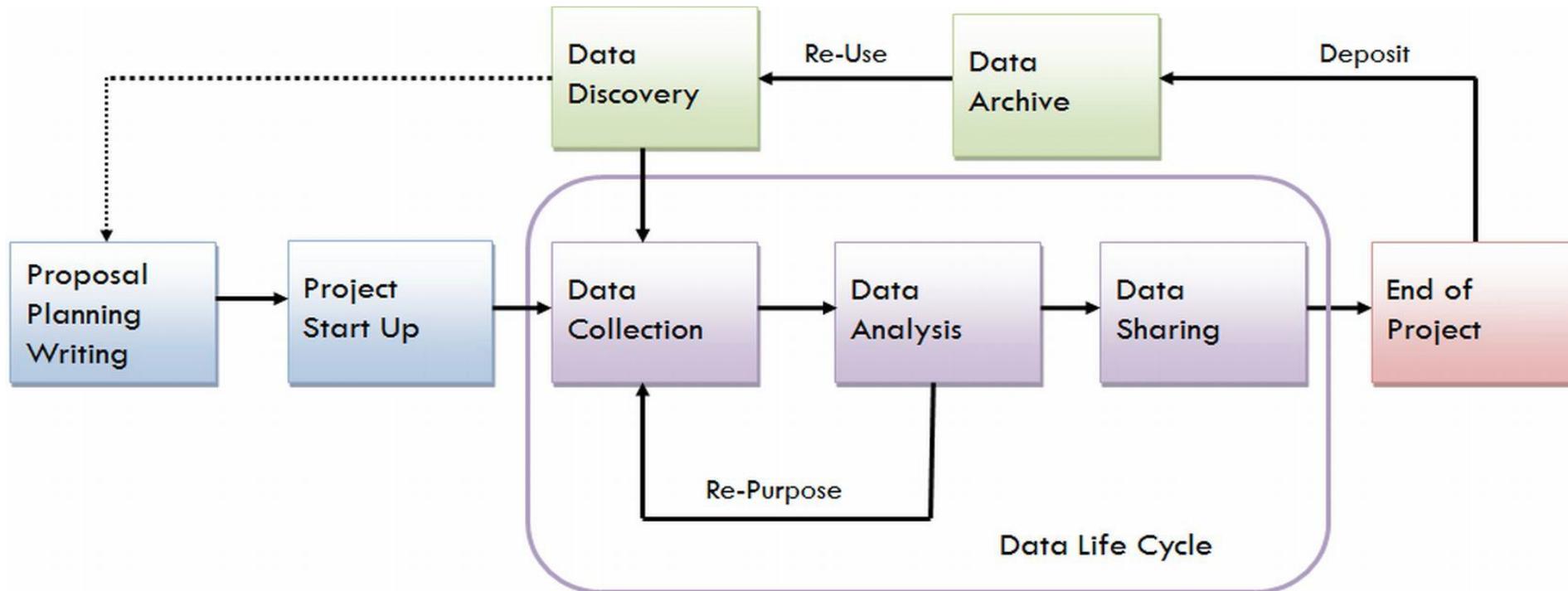


Hochstenbach, P. (2018). *Open Research Data Material - FAIR data principles*. [image] Available at: <https://hochstenbach.wordpress.com/> [Accessed 26 Apr. 2018].

You can have a closed/restricted access and still be FAIR

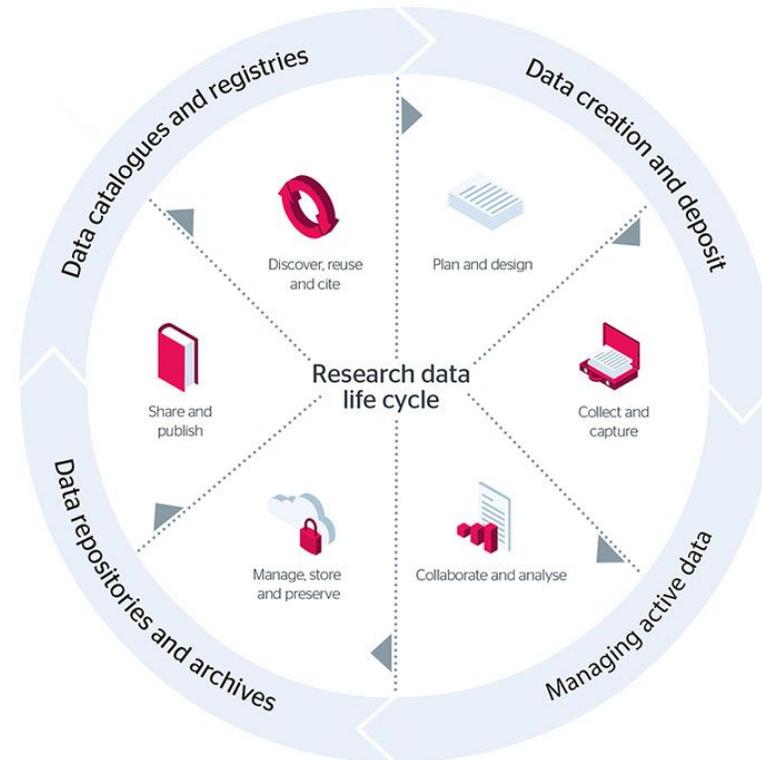
RDM/Data stewardship key

Data stewardship and good RDM practices from the beginning are very crucial to achieve FAIR data



Data Management Plans for FAIR data

A written plan on how you plan and execute your research life cycle



<https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data>

What does a DMP cover: A checklist

1. Administrative Data
2. Data Collection & Organisation
3. Storage and Backup
4. Documentation and Metadata
5. Ethics and Legal compliance
6. Selection and Preservation
7. Data Sharing
8. Responsibilities and Resources

<http://www.dcc.ac.uk/resources/data-management-plans/checklist>



DMP Online

Welcome

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

Join the growing international community that have adopted DMPonline:



17,622 Users



203 Organisations



23,083 Plans



89 Countries

Some funders mandate the use of DMPonline, while others point to it as a useful option. You can [download funder templates](#) without logging in, but the tool provides tailored guidance and example answers from the DCC and many research organisations. Why not sign up for an account and try it out?

Sign in

Create account

* Email

* Password

[Forgot password?](#)

Remember email

Sign in

- or -

Sign in with your institutional credentials

Creating the Plan

Create a new plan

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

* What research project are you planning?

mock project for testing, practice, or educational purposes

* Select your research organisation

- or - No research organisation associated with this plan or my research organisation is not listed

* Select the primary funding organisation

- or - No funder associated with this plan or my funder is not listed

Create plan

Cancel

It's a new way of working for researchers... after years of publish or perish

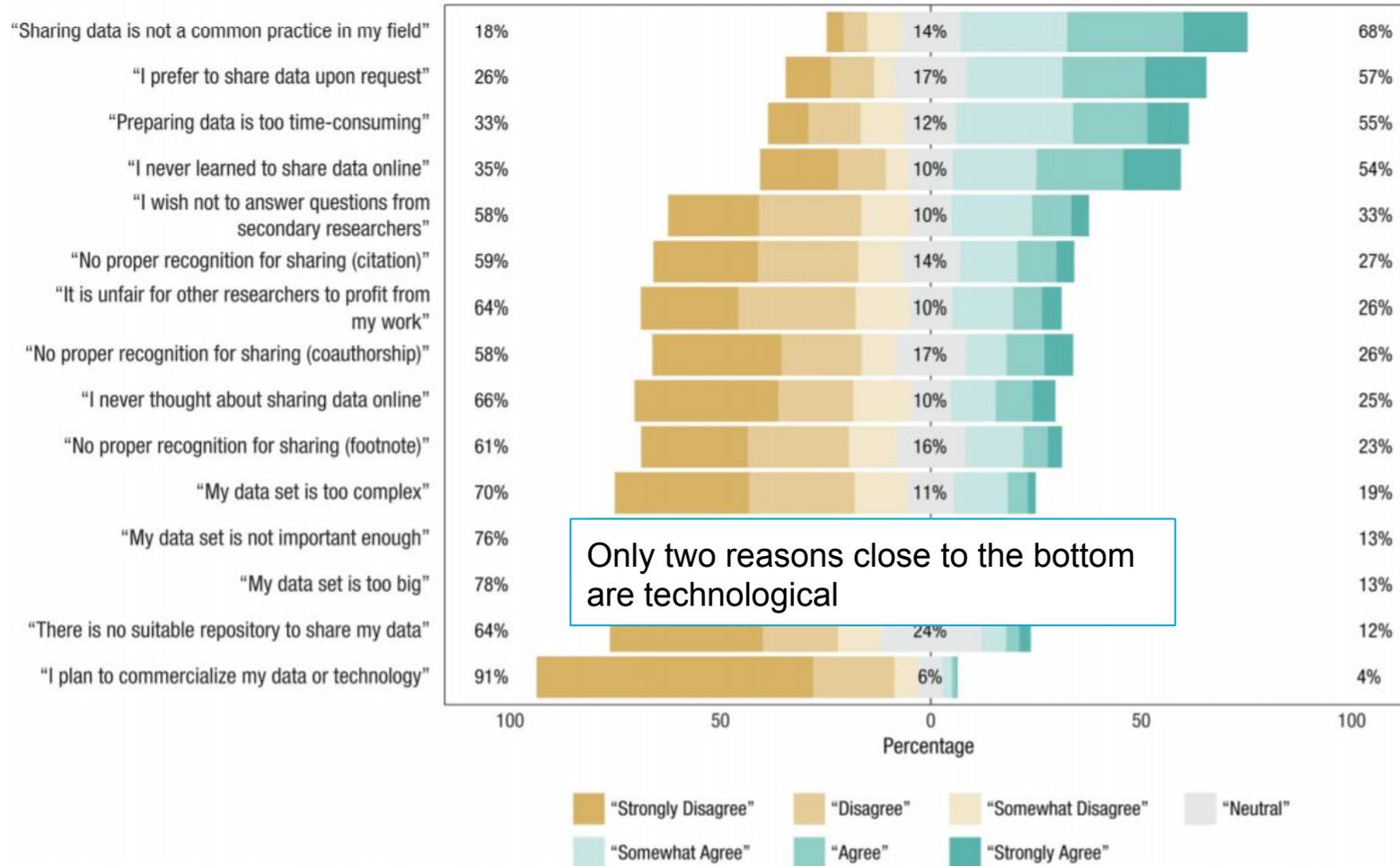


Let's see the PoV of researchers

My data contains personal/sensitive information	My data is too complicated	People may misinterpret my data	My data is not very interesting
Commercial funder doesn't want to share it	We might want to use it in a(nother) paper	People will contact me to ask about stuff	Data Protection/ National Security
It's too big	People will see that my data is bad	I want to patent my discovery	It's not a priority and I'm busy
I don't know how	I'm not sure I own the data	Someone might steal/plagiarise it	My funder doesn't require it

*Credit:
Dr Jenny Molloy, Open Knowledge
Foundation*

To what extent do you agree with the following statements about barriers related to data sharing?



<https://doi.org/10.1177/2515245917751886>



Similar findings in other reports (from other disciplines), e.g.

Life sciences, social sciences and humanities:

Van den Eynden et al. (2016)

<https://doi.org/10.6084/m9.figshare.4055448.v1>

All disciplines:

Johnson et al. (2016)

<http://doi.org/10.5281/zenodo.177856>

The key challenges are cultural/policy related and not technological



FAIR data is not the same for all



- The culture, awareness and the attitudes towards RDM and (FAIR) open science varied starkly across departments
- RDM is key to achieve FAIR data
- Researchers need support

Graphic source:
<https://www.andis.org.au/working-with-data/fairdata/training>

What do the Data Stewards do?



Researcher oriented:

- They are there for any data questions
- General practitioners
- Usually have research experience
- Help researchers who “don’t have data” and who “don’t have problems”

Objective: Improve daily practice not enforcement



What does RDM support mean?



Advice



Storage



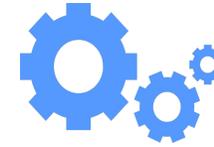
Costs



Compliance



Data
Management
Plans



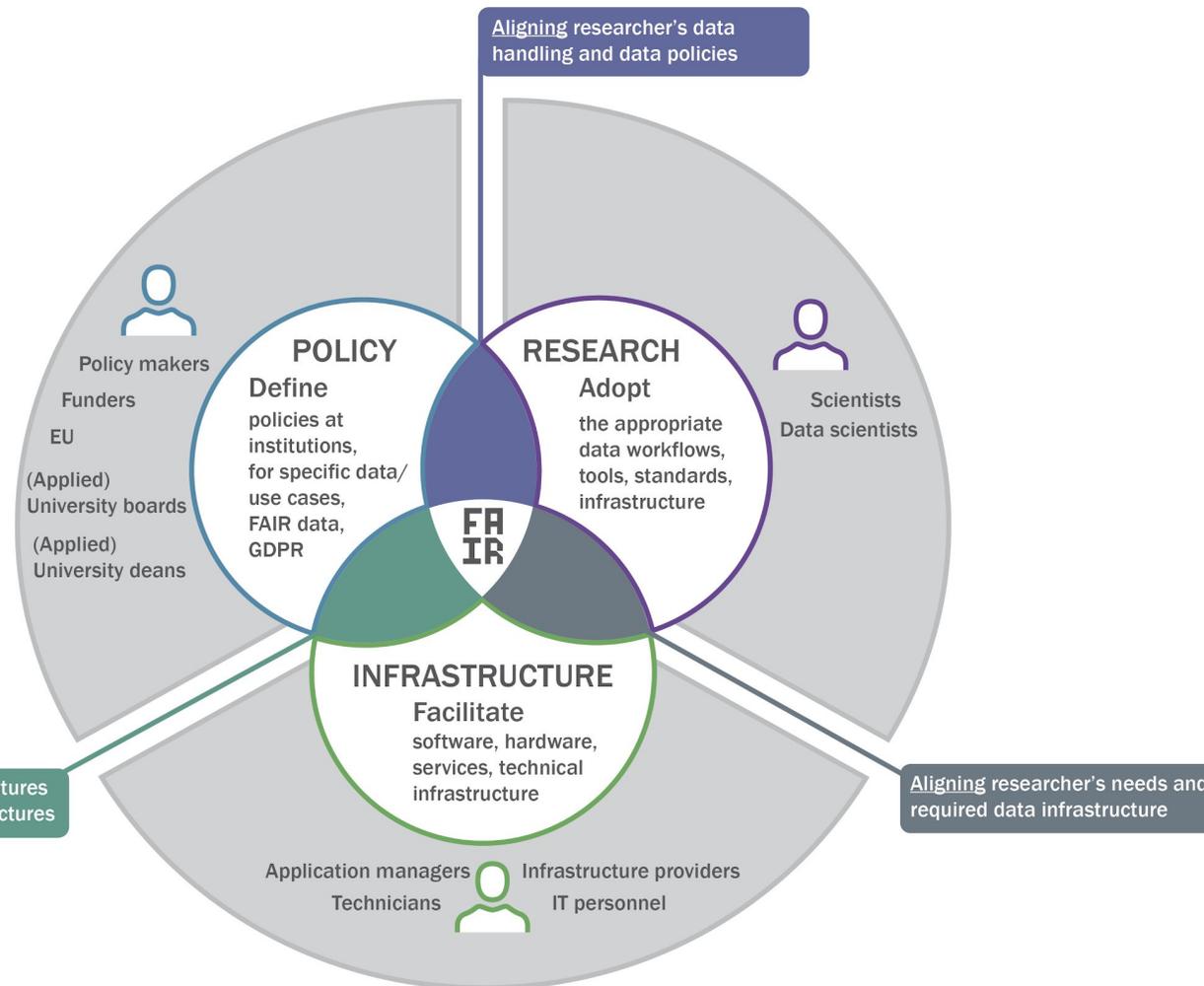
Tools



Training

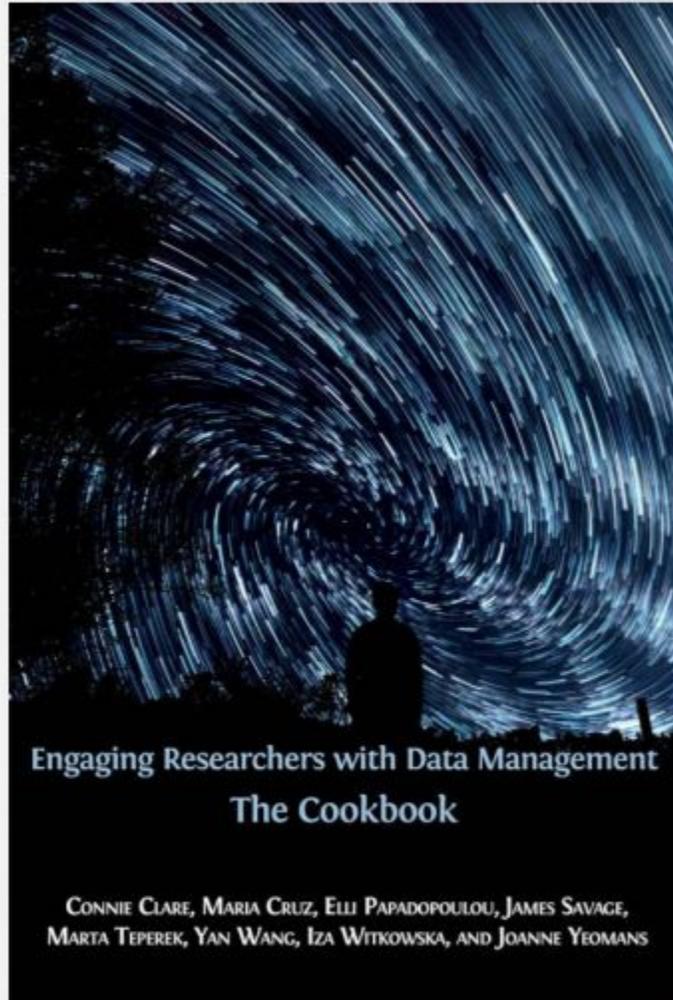
It crosses people, tech and policy

Implementation areas for data stewardship



<https://www.dtls.nl/2019/10/21/professionalizing-fair-data-stewardship-in-the-life-sciences-defining-job-criteria-skills/>

To get started with RDM support



Engaging Researchers with Data Management: The Cookbook

October 2019

DOI: 10.11647/OBP.0185

<https://www.openbookpublishers.com/product/1080>

(or go here for a draft preview:

<http://bit.ly/RDMbook>)

Institutional Personnel

Number of researchers



Number of PhD students



Number of full time employees (FTE) providing RDM Support centrally



Target Audience (Researchers)



Main Drivers



Cost (Materials, Infrastructure, Time)



Ease of implementation



Who Is this Book For?

This book has been written for anyone interested in RDM, or good research practice more generally. It will be particularly useful to those interested in how to effectively engage with researchers about research data management. This might include librarians, data managers, data stewards, archivists, members of ICT (Information and Communication Technology) departments, colleagues from legal and financial support, faculty management, senior executives at institutions, funders, policymakers, publishers, members of the commercial sector, and researchers at any career stage who want to change practices among their peers. In short, if you have read this far, then this book is for you.

Why Read this Book?

We hope that reading this book will:

- inspire you to implement new activities to engage with researchers about research data;
- help you find the activities most suitable for your institutional setting (according to size, research profile, resources available for data management, target audience, etc.);
- inform you about the ease of implementing each case, identifying the specific challenges associated with them and possible tips to overcome these;
- give you a general overview of what other institutions around the world do to engage their researchers with research data;
- provide you with tangible suggestions for actions that you could present to senior management at your institution;
- stimulate collaboration. We hope that reading our case studies and learning about the initiatives adopted by contributing institutions will lead to new connections and cooperation.

<https://www.openbookpublishers.com/product/1080>

Summary

- Importance of Open Science
- Role of FAIR data management and
- Researcher perspective
- Data Stewardship and RDM support
- How to get started with RDM support

There is no looking back really



Open Science is the
only way forward in the
post truth world



Thank
You!

shalini.kurapati@polito.it

@shalini_kr

