

Data Stewardship and RDM: The why and how from researchers' perspective

SEBD 2020 (Online Conference)

22 June 2020



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**
 - **Fully researcher oriented**
- Co-founder and CEO of clearbox.ai, MLOps for responsible AI adoption
- Certified Information Privacy Professional/Europe

Past:

- 7 years research experience, PhD and Post-doc, TU Delft
- Department advisor on RDM, Data stewardship project TU Delft

OS and RDM from the researcher's' perspective



What do researchers think of open science

What are the 3 words that come to your mind when you hear the term open science

Mentimeter

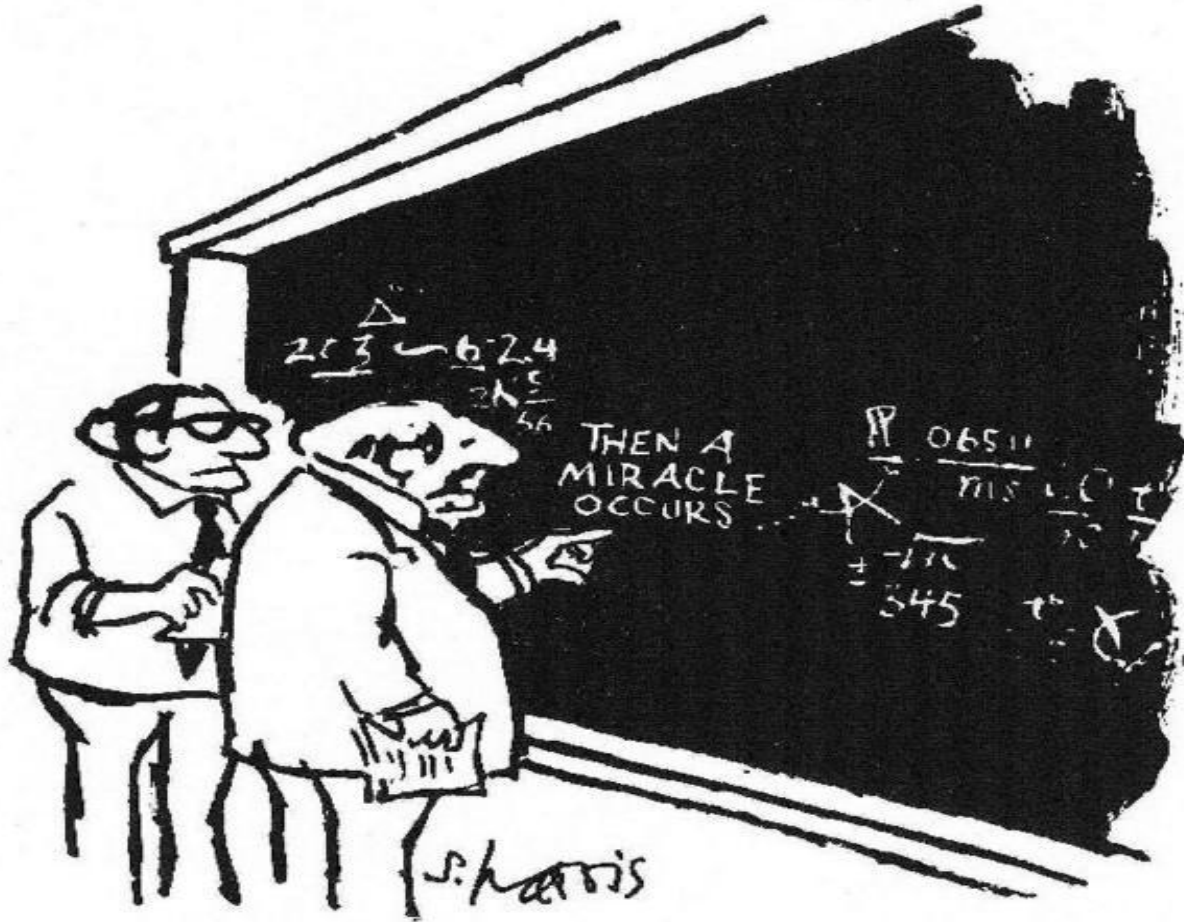




Definition of open science

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).

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

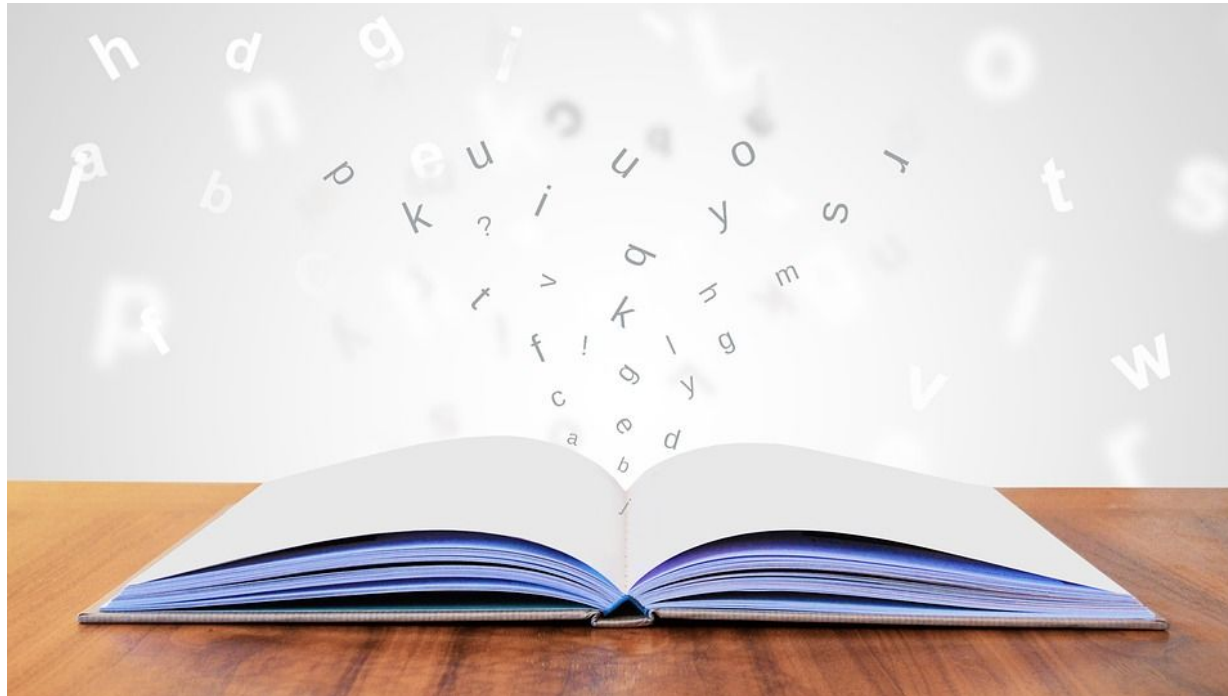
**OPEN SCIENCE:
JUST
SCIENCE
DONE RIGHT**

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



If open science is just science

Why is everyone talking about it now!



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

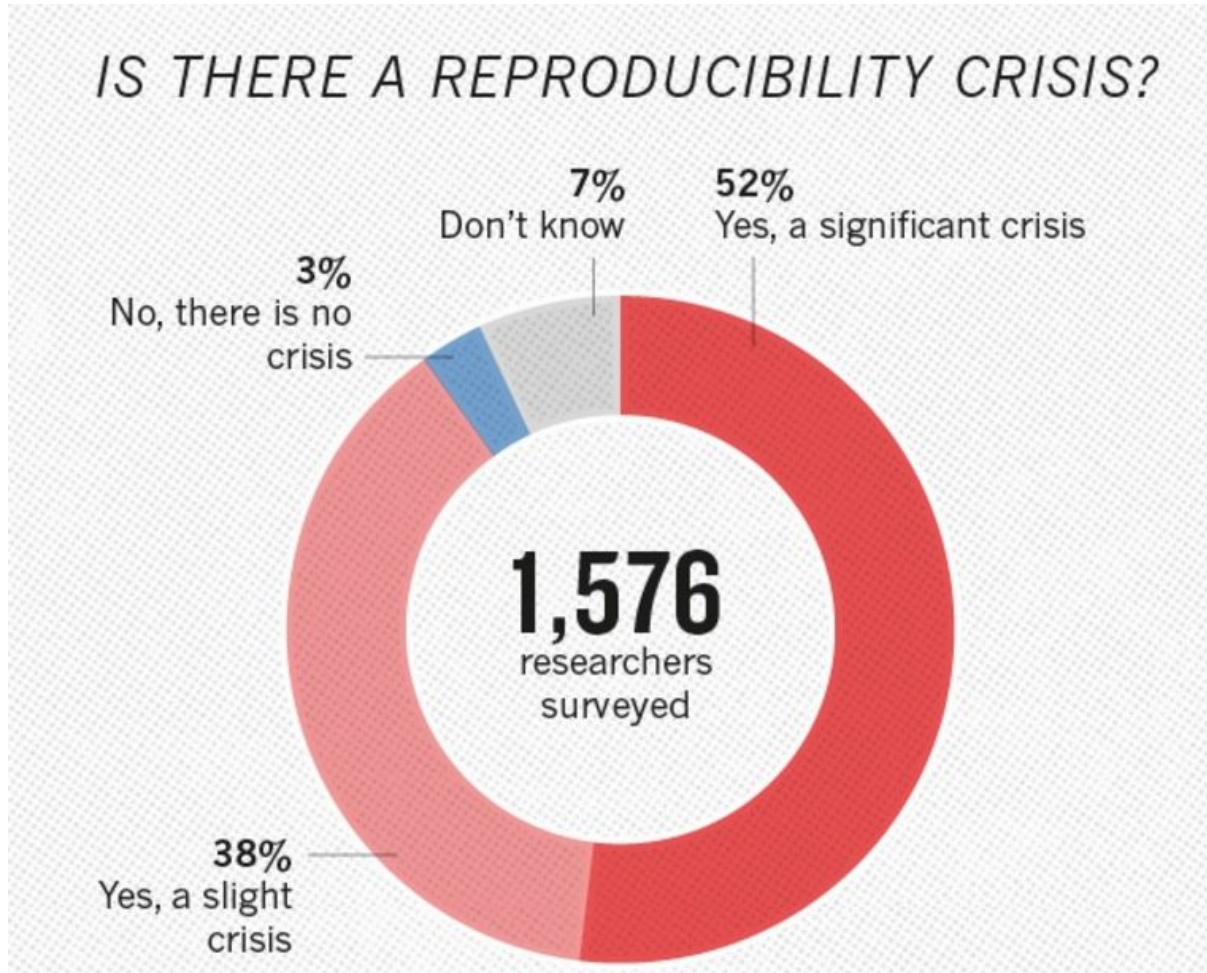
I must get the
result that fits my
narrative so I can
get my paper into
Nature..



facebook.com/pedromics

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

The reproducibility crisis



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

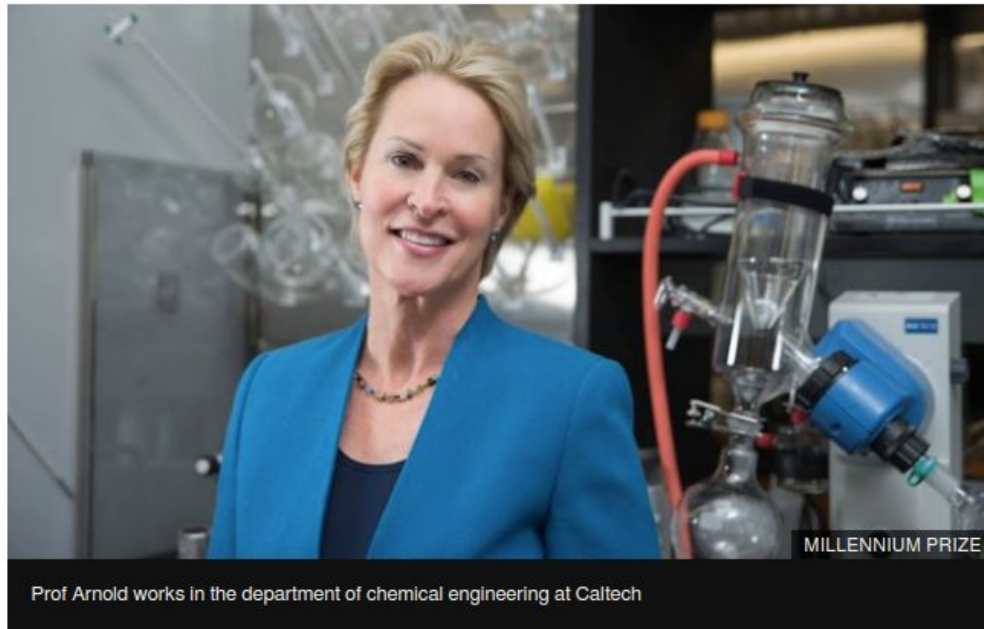
Happens even to the best of scientists

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



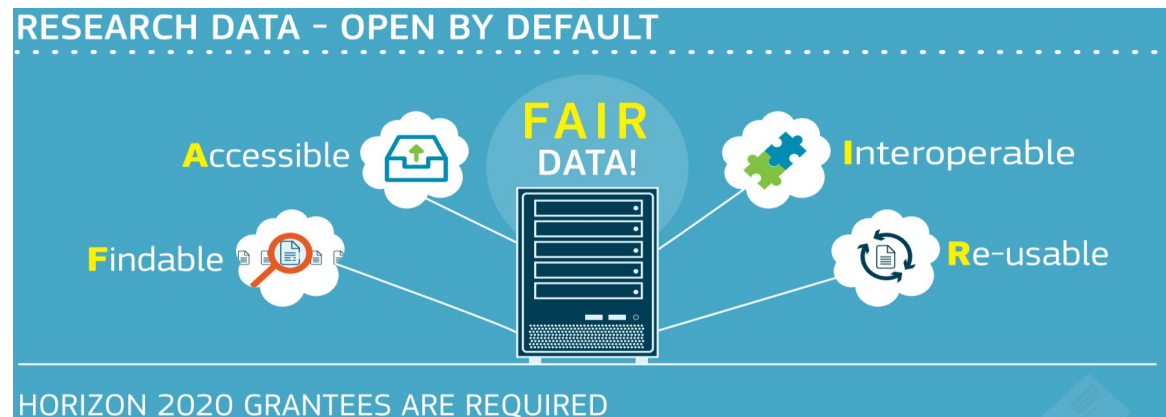
- Selective reporting
- Pressure to publish
- Insufficient supervision and training
- **Supporting data / methods / code not available**

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

<https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>



POLITECNICO
DI TORINO

Top journals need it already!

nature > nature research > editorial policies > reporting standards and availability of data, materials, code and protocols

nature research



Search



Login

Editorial policies

Authorship

Competing interests

Confidentiality

Plagiarism and duplicate publication

Image integrity and standards

Preprints & Conference Proceedings

Peer-review policy

Reporting standards and availability
of data, materials, code and

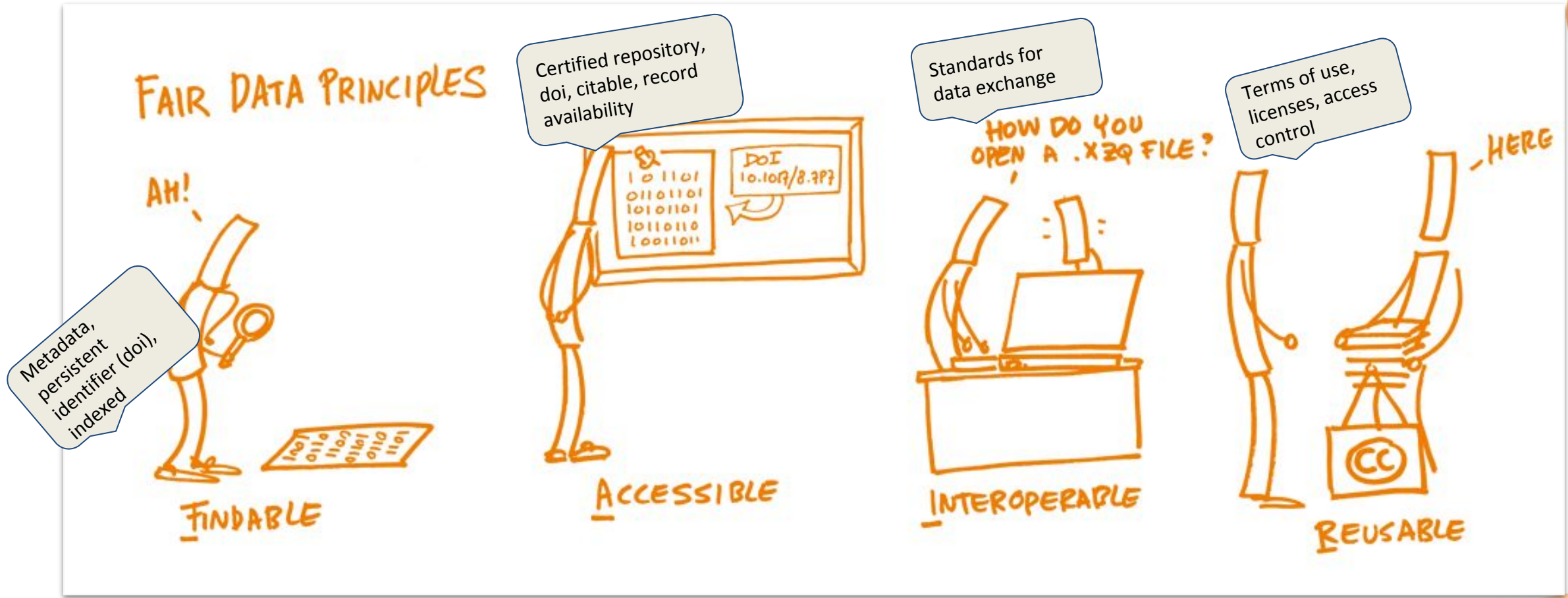
Reporting standards and availability of data, materials, code and protocols

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature Research journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications**. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.



POLITECNICO
DI TORINO

What is FAIR data?

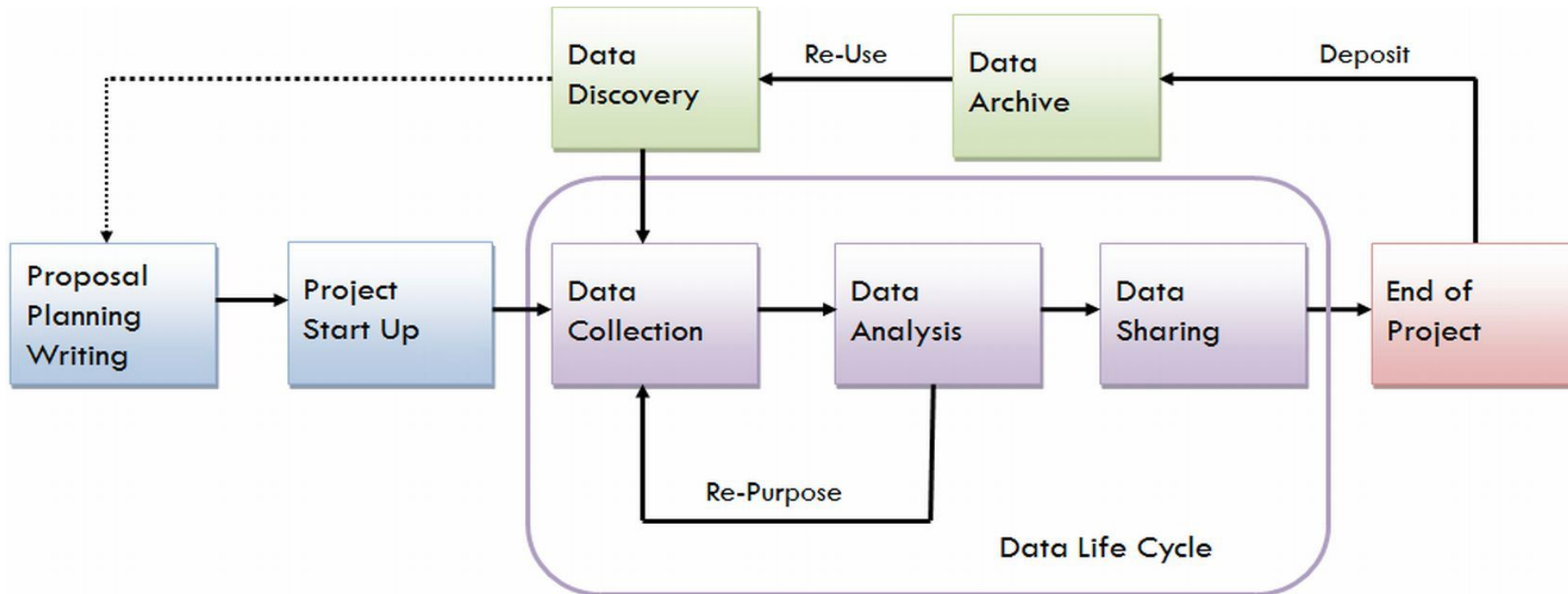


Hochstenbach, P. (2018). *Open Research Data Material - FAIR data principles*. [image] Available at: <https://hochstenbach.wordpress.com/> [Accessed 26 Apr. 2018].

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



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*

Similar questions asked to 600 researchers

Invited Forum: Challenges in Making Data Available
Empirical Article



Data Sharing in Psychology: A Survey on Barriers and Preconditions



**Bobby Lee Houtkoop¹, Chris Chambers², Malcolm Macleod³,
Dorothy V. M. Bishop⁴, Thomas E. Nichols^{5,6,7}, and
Eric-Jan Wagenmakers¹**

¹Psychological Methods Programme Group, University of Amsterdam; ²Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University; ³Centre for Clinical Brain Sciences, University of Edinburgh; ⁴Department of Experimental Psychology, University of Oxford; ⁵Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford; ⁶Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford; and ⁷Department of Statistics, University of Warwick

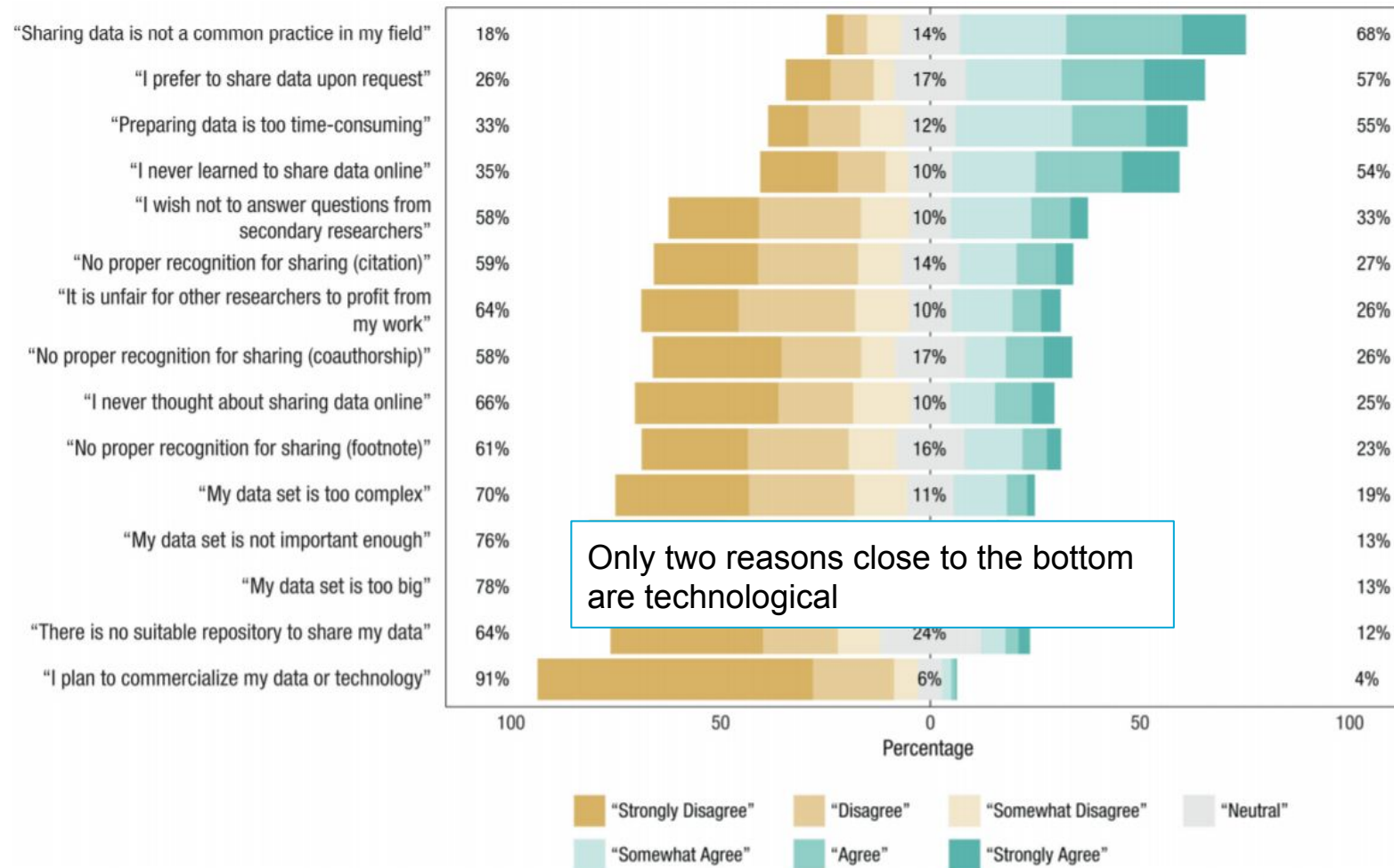
Advances in Methods and
Practices in Psychological Science
2018, Vol. 1(1) 70–85
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2515245917751886
www.psychologicalscience.org/AMPPS

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



POLITECNICO
DI TORINO

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 does RDM support mean?



Advice



Storage



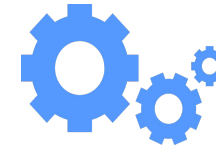
Costs



Compliance



Data
Management
Plans



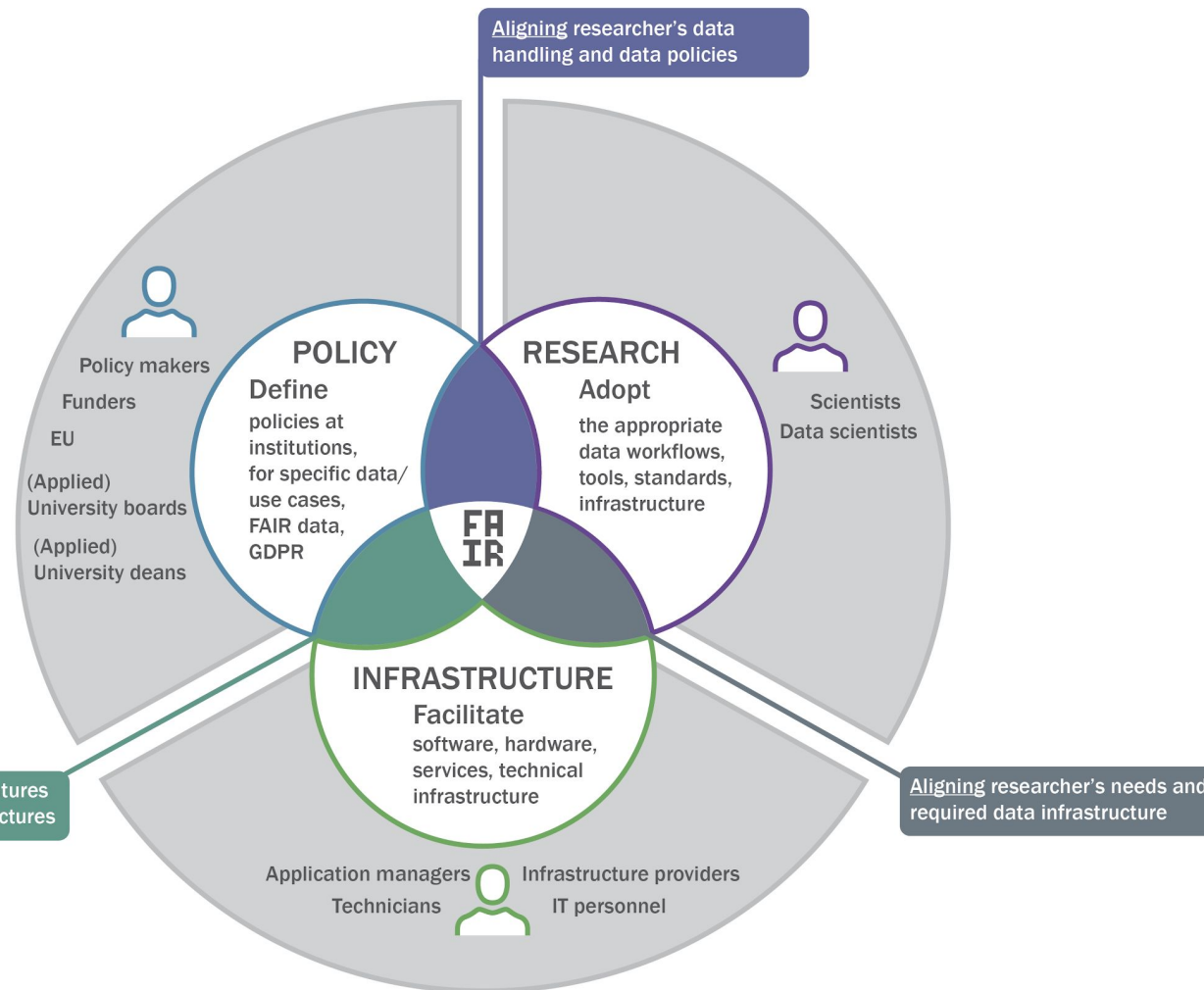
Tools



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/>

We need strong policy measures



Evaluation of Research Careers fully acknowledging Open Science Practices

Rewards, incentives and/or recognition for researchers practicing Open Science

It's time for open science skills to count in academic careers (Part 1: Talks)

Authors: Shalini Kurapati, Marta Teperek, Maria Cruz, Angus Whyte

Disclaimer: In the spirit of openness and transparency, we would like to share that Shalini Kurapati wrote parts of this blog post based on the zenodo record of the presentations even though she wasn't present during the event. Her account was verified by the remaining authors who were present.

To read Part 2 of this blog post follow this [link](#).

Open Science is not always easy – skills are urgently needed

Open science is becoming a ubiquitous and recurring theme in the current academic environment. Researchers are increasingly expected to publicly share their research outputs (data, code, models etc.) as well as their publications. This often requires considerable effort from researchers to manage and curate their research outputs to make them shareable.

<https://openworking.wordpress.com/2018/12/03/its-time-for-open-science-skills-to-count-in-academic-careers-part-1-talks/>



POLITECNICO
DI TORINO

Concluding thoughts

- Need for **proper infrastructure** and **policy support** from institutions
- Training, awareness and people infrastructure to drive cultural change
- **Recognition** is the main drivers for both scientific and non-scientific staff to pursue open science
- And most importantly: everyone can contribute to changing cultures and daily practices.





Thank
You!

shalini.kurapati@polito.it

@shalini_kr

