# The role of Data Stewards in realising Open Science objectives

Dr. ir. Shalini Kurapati
22 Nov 2022, *Uni Pisa*

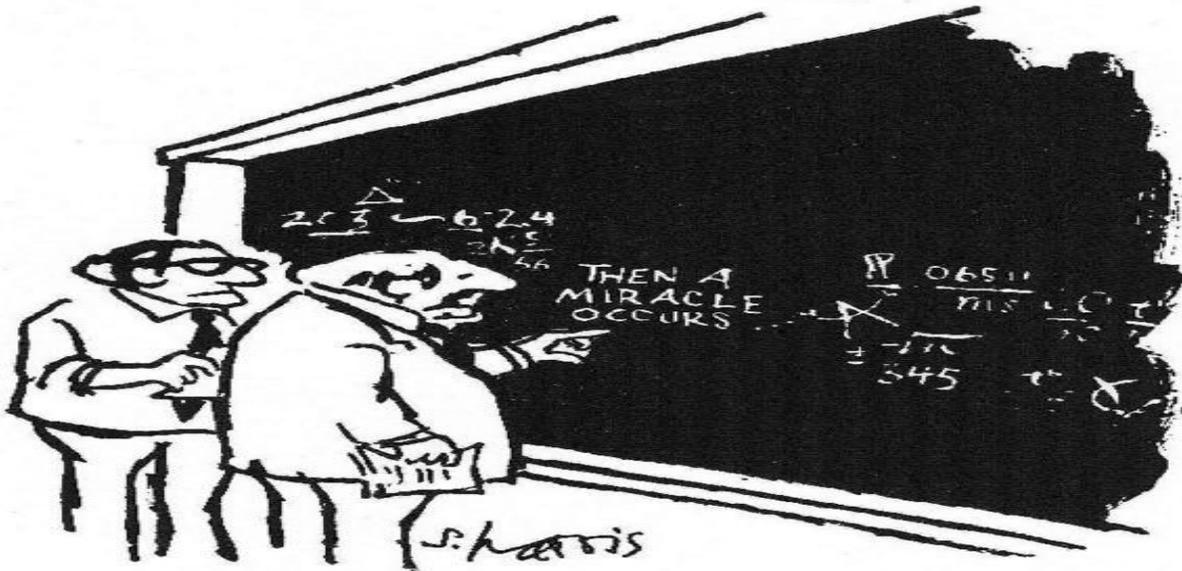# About



Dr. ir. Shalini Kurapati

iD https://orcid.org/0000-0003-1492-4100

– Co-founder and CEO [clearbox.ai](clearbox.ai)
– Past: RDM at TU Delft, advised researchers on data matters
– Open Science and RDM fellow at PoliTo
– 9 years of research experience at the intersection of technology, policy and management
– Certified Information Privacy Professional/Europe

# Open Science



"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
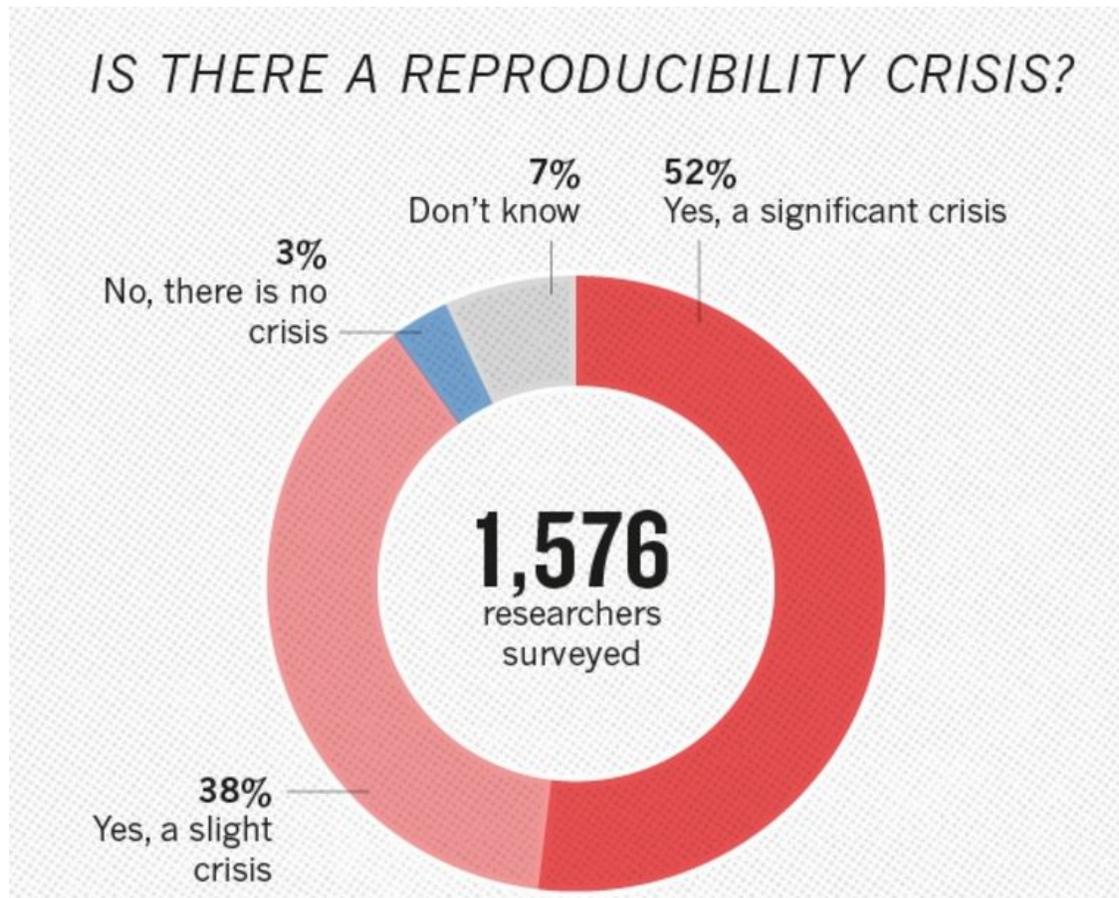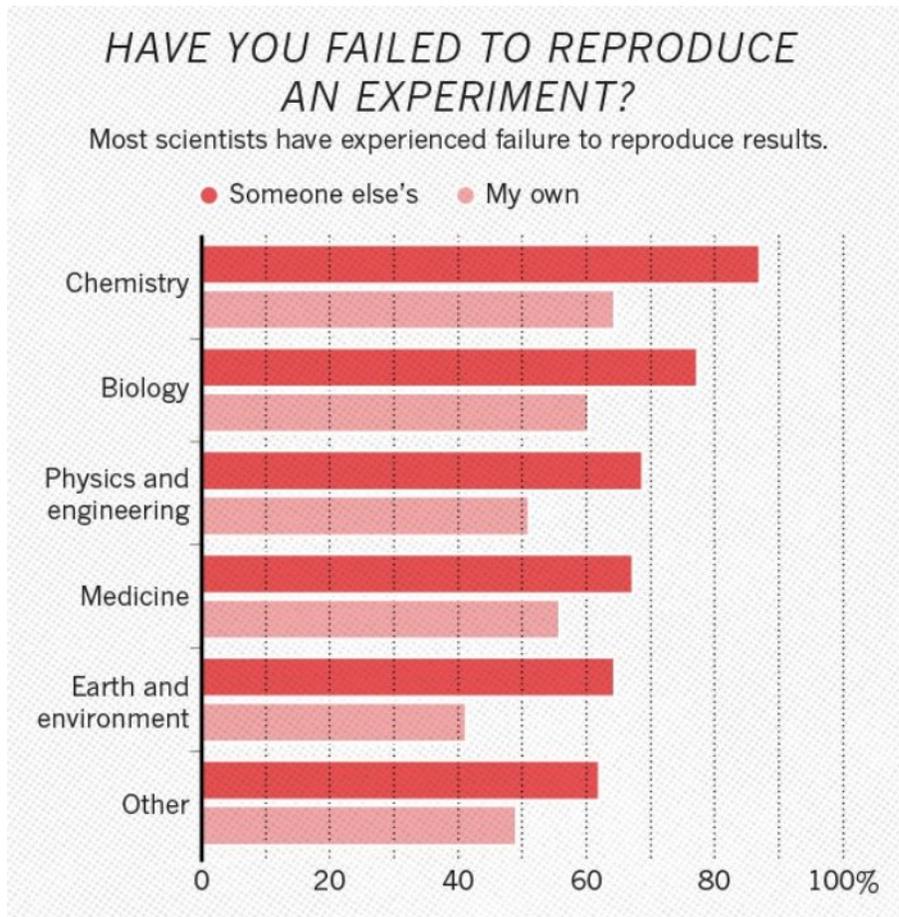- And all other things that basically define science.

# It's nothing new

# The reproducibility crisis



IS THERE A REPRODUCIBILITY CRISIS?

7% Don't know

52% Yes, a significant crisis

3% No, there is no crisis

1,576 researchers surveyed

38% Yes, a slight crisis

# In the engineering and physics disciplines



HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?
Most scientists have experienced failure to reproduce results.
● Someone else's  ● My own

About 70% cannot reproduce others' experiments

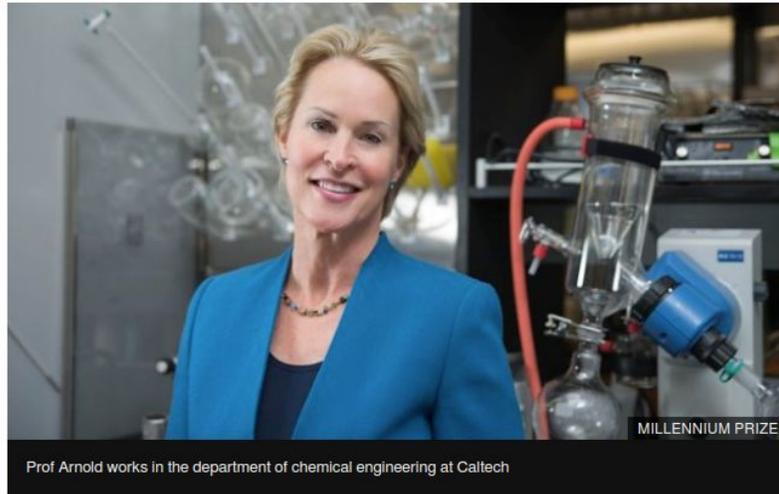and more than 50% cannot reproduce their own experiments!

# Happens even to the best of scientists

### Nobel Prize-winning scientist Frances Arnold retracts paper

🕓 3 January 2020

f  💬  🐦  ✉  ⏦ Share

Nobel Prize


Prof Arnold works in the department of chemical engineering at Caltech

MILLENNIUM 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].

# 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

# Funding bodies are pushing for open science, focus on OA publications and FAIR data



More are following, nationally and regionally

RESEARCH DATA – OPEN BY DEFAULT

**FAIR DATA!**

Accessible

Findable

Interoperable

Re-usable

HORIZON 2020 GRANTEES ARE REQUIRED

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

The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807.

# The variety...

Go to **www.menti.com** and use the code **7813 0903**

Mentimeter

## What kind of research data do you produce/ use?
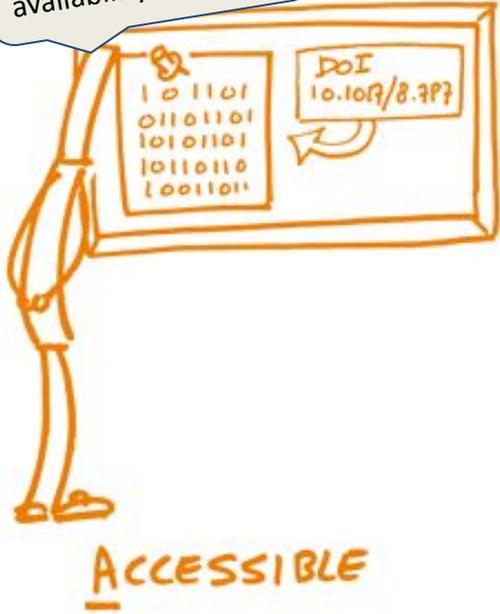
# It's highly discipline specific.. FAIR is not same for all

# It's a loot of work!! And new- not just researchers

https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs
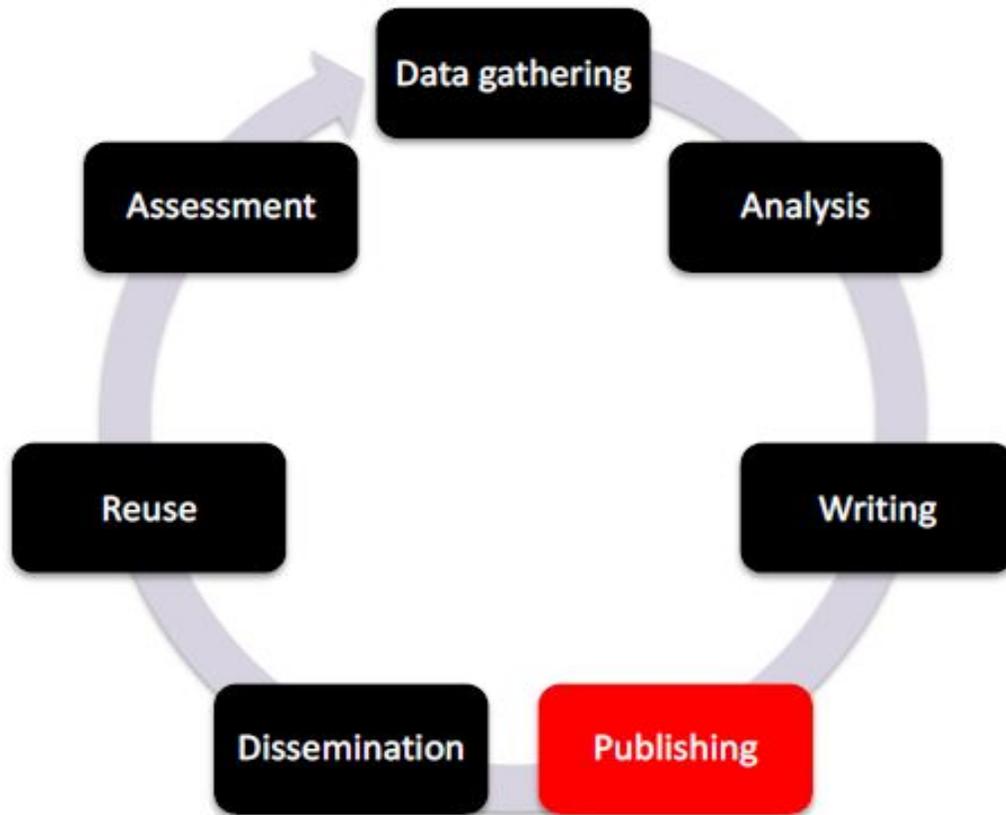
The only thing that counts in academia is
publication of novel results in high impact
journals

# Most importantly



Culture shift.. engaging with researchers.. showing the practical ropes

# In any context, RDM is key!



We cannot achieve open science or FAIR data without RDM.

It is not simple but crucial!

Solutions at the intersection of technology, culture and awareness and policy

**Immediate reactions to "data" requirements:**

- It would take me 5 years to find all my data!

- The PhD/postdoc who had the data left the lab

- Should we write down all protocols?

- Data management is a waste of time

- Nobody will understand my data

- People can just ask me for it when they need it

# In my experience



- They are there for any data questions
- Help researchers who "don't have data" and who "don't have problems"
- First point of contact, general practitioners
- Every occasion is good!

- Change happens one person at a time

# Objective: Improve daily practice not enforcement

# TU Delft Experience

# In POLITO

# HE Guidance

https://zenodo.org/record/5527043

# Data stewardship: Is new and evolving



Figure 2.5 ZonMw/ELIXIR data stewardship roles in the data stewardship landscape[58]

Study in Denmark
https://zenodo.org/record/3609516#.Y
pZm31RByUk

**TABLE 3.** Titles given to the role of a DS.

| Preferred term | n |
|---|---|
| Don't know | 37 |
| Data manager | 29 |
| DPO | 13 |
| Data Steward | 10 |
| Data librarian | 6 |
| Information Management Specialist | 4 |
| Arkivar / Digitaliseringskonsulent | 3 |
| Data scientist | 2 |
| Dataansvarlig | 2 |
| IT specialists | 2 |
| Records manager | 2 |
| Consultant | 2 |
| Colleague | 2 |
| Administrative medarbejder - IT | 2 |
| Project managers / dept. coordinator | 2 |
| Business intelligence analyst | 1 |
| Data Engineers | 1 |
| Data handler | 1 |
| Data koordinator (coordinator) | 1 |
| Data specialists | 1 |
| Data / system owners | 1 |
| Datadisciple | 1 |
| IT Risk Management | 1 |
| IT-koordinator | 1 |
| Systemanasvarlig og digitaliseringskonsulent | 1 |
| Projektmedarbejder | 1 |
| Research Data Officer | 1 |
| Research supporter | 1 |

# Open Science, more changes to follow
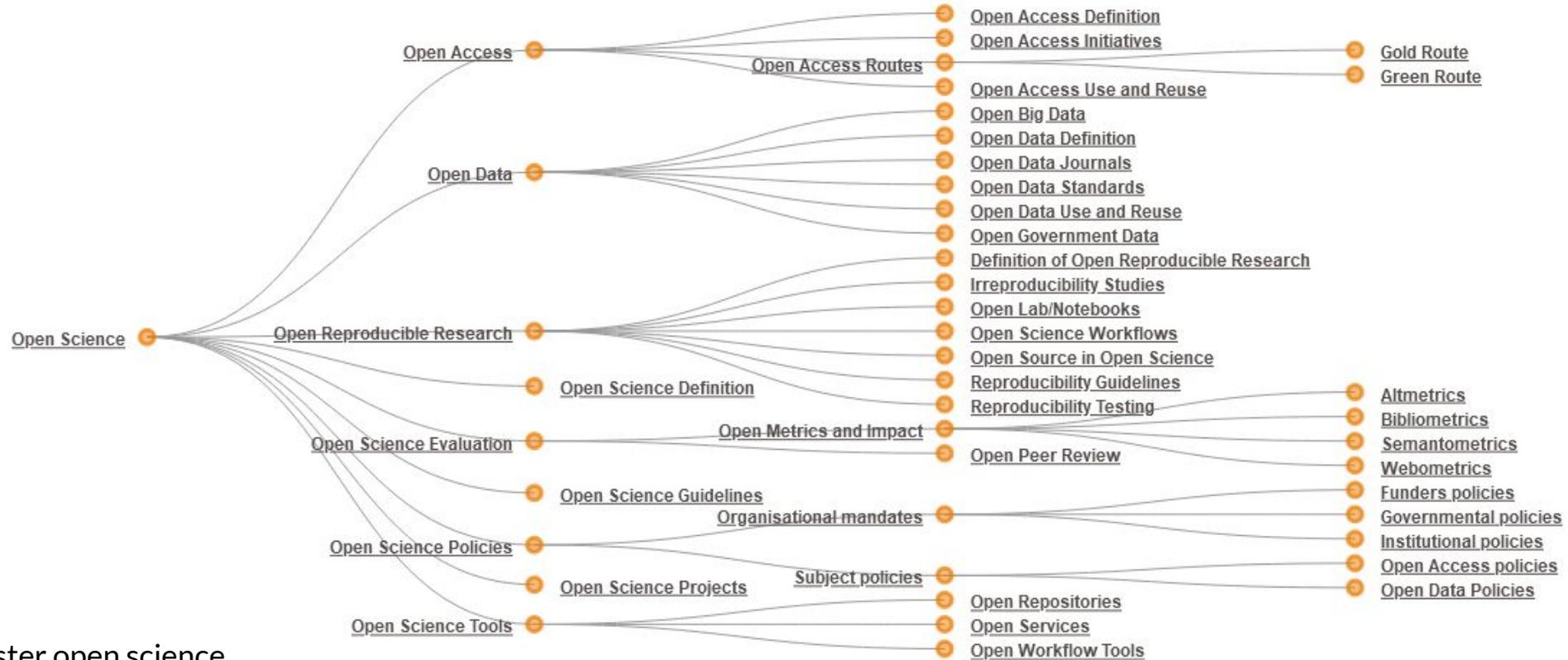


foster open science

# There is no looking back really
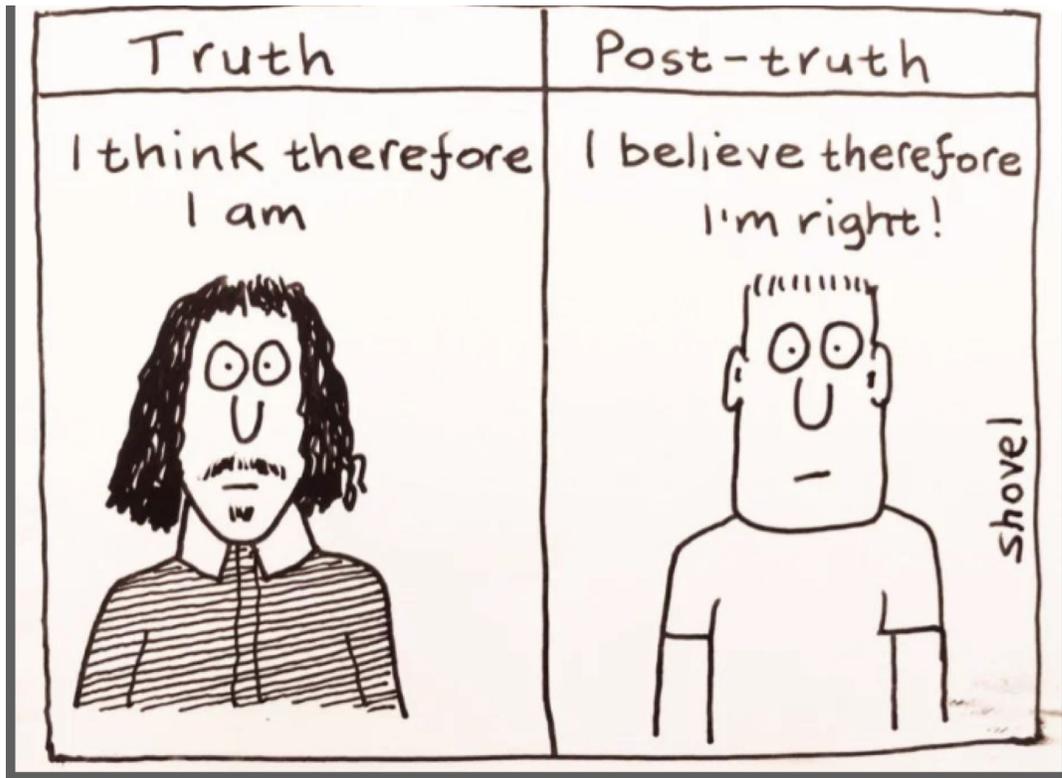


Open Science is the only way forward in the post truth world

# Thank you!

Twitter @shalini_kr