Being Open and Responsible in Your Own Research Context

Dr Louise Bezuidenhout

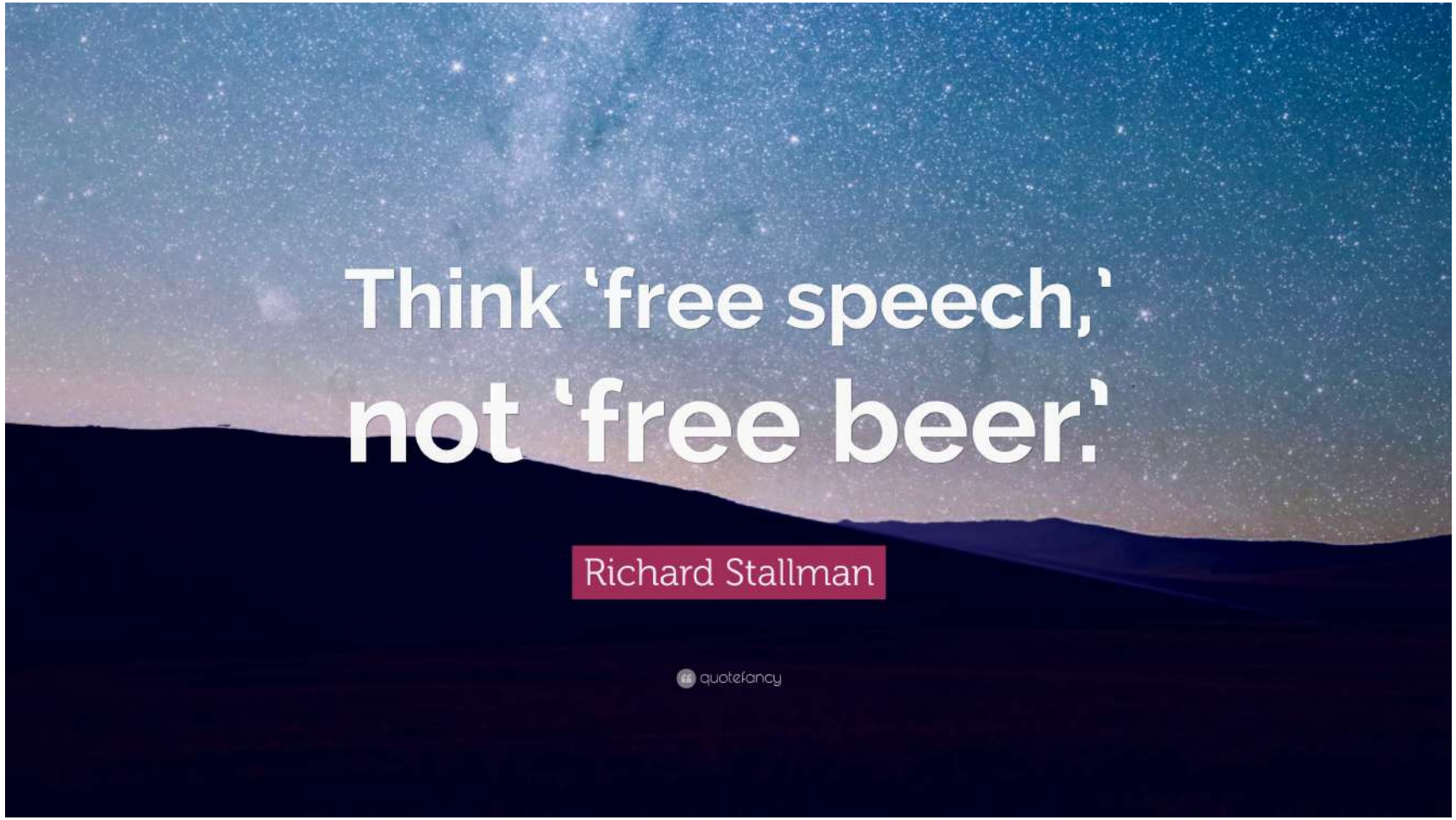
Institute for Science, Innovation and Society
University of Oxford

Plan

- Personal concerns and implementing OS practices at home
  - What have we learned this week?
  - What challenges do we have about implementing these in our research environments?
  - What kinds of assistance can we get?
- RCR and the “bigger picture”
  - Designing just systems
  - Avoiding biases and marginalizations

Recap From Monday

Recap From Monday



Open Science

- **The products of scientific research should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control**

  - Transparency in experimental methodology, observation, and collection of data
  - Public availability and reusability of scientific data
  - Public accessibility and transparency of scientific communication
  - Using web-based tools to facilitate scientific collaboration

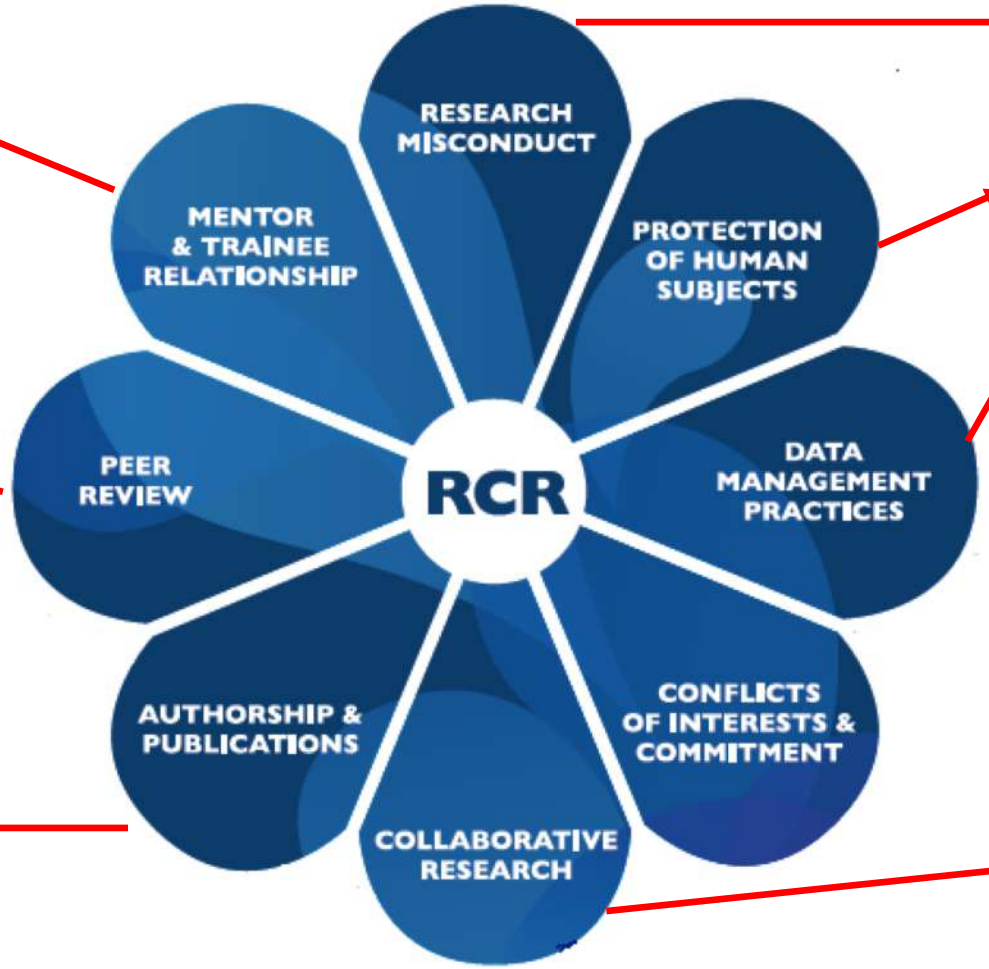
    http://www.openscience.org/blog/?p=269]

Openness and Responsible Conduct of Research

Open Lab Books: Transparency in research practices
Sharing and openness: enhance transmission of values

Open Peer Review:
Transparency in peer review leads to better dialogue and collegial behaviour

Open Access: Improves availability of research outputs
Open publishing: leads to improved citations, credit and collaboration

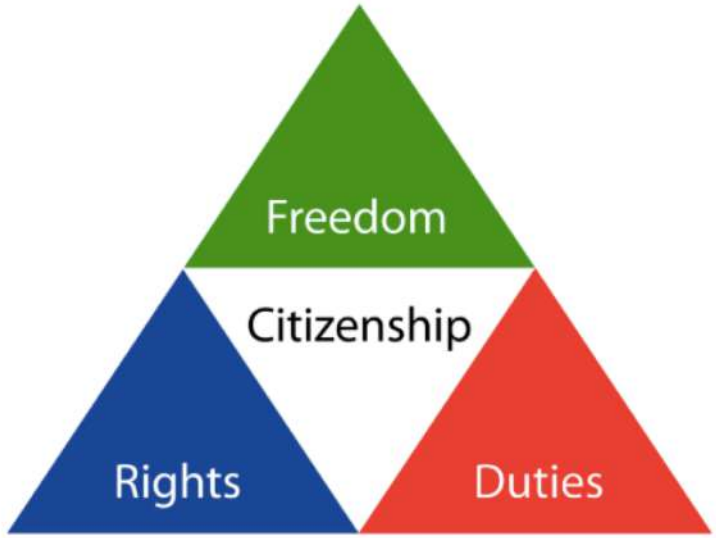


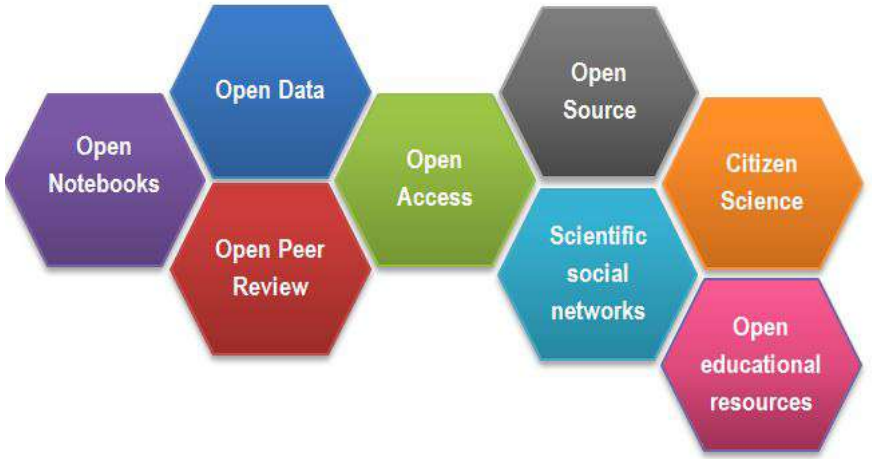
Open Data and Open Methodologies:
Improve transparency and reproducibility of research

Open Science Tools:
Improve collaboration

Open and Responsible Science Citizenship

- Consider ethics in daily activities
- Engage in discussions about the “big picture”
- Contribute, safeguard and curate community resources
- Contribute to community-building activities
- Uphold and promote community values

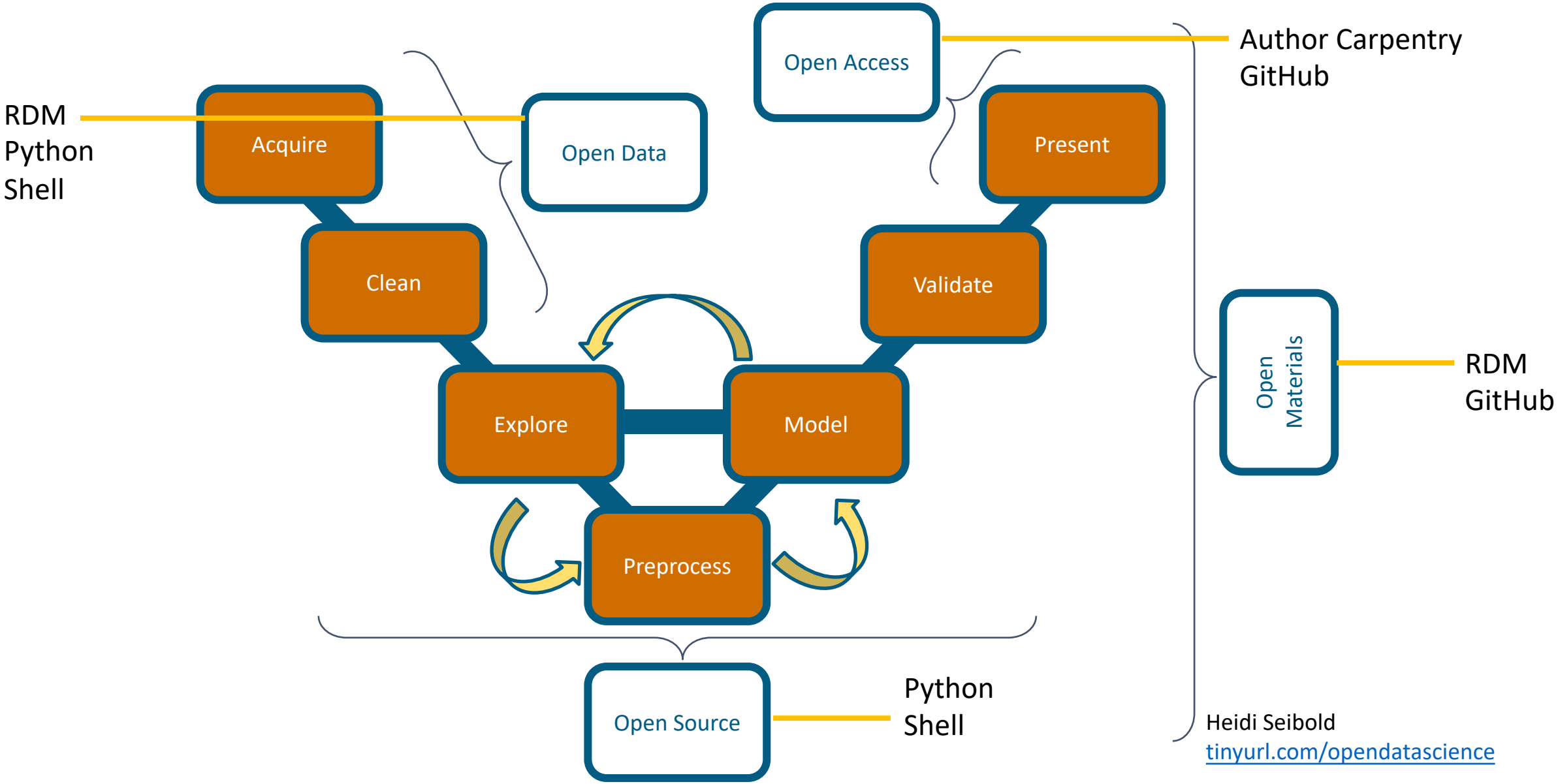




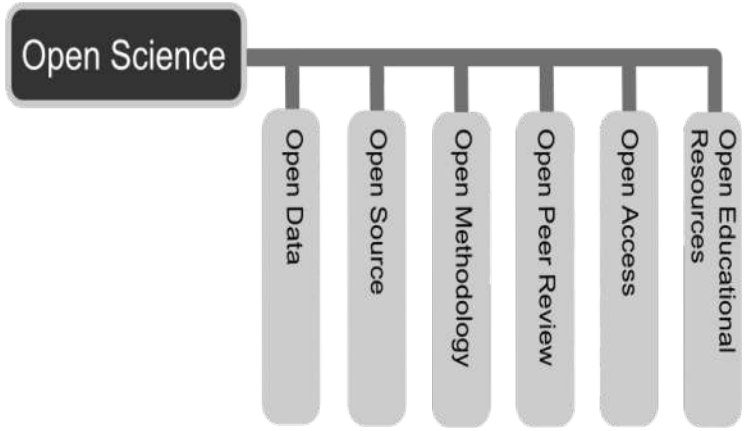
https://www.fosteropenscience.eu/content/what-open-science-introduction

Practical Tools, Ethical Impact

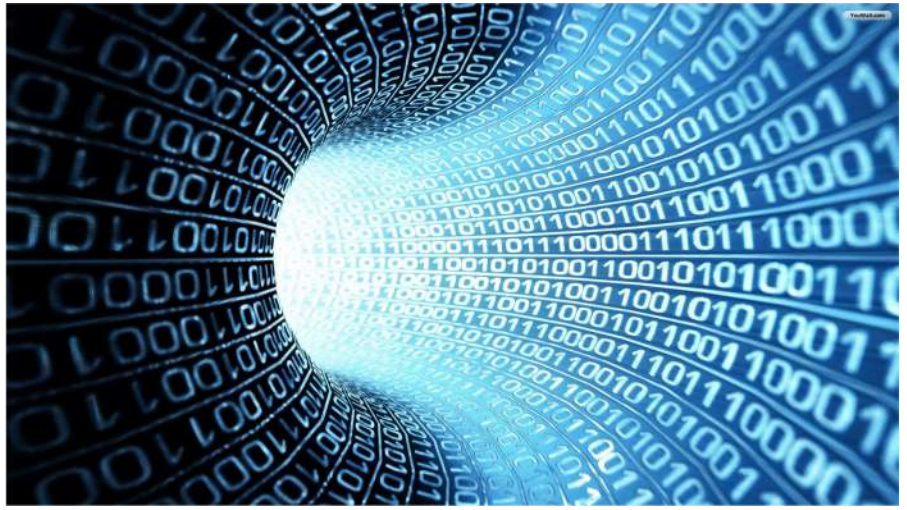
Open Science Throughout The Research Lifecycle



How to be a responsible and open data scientist



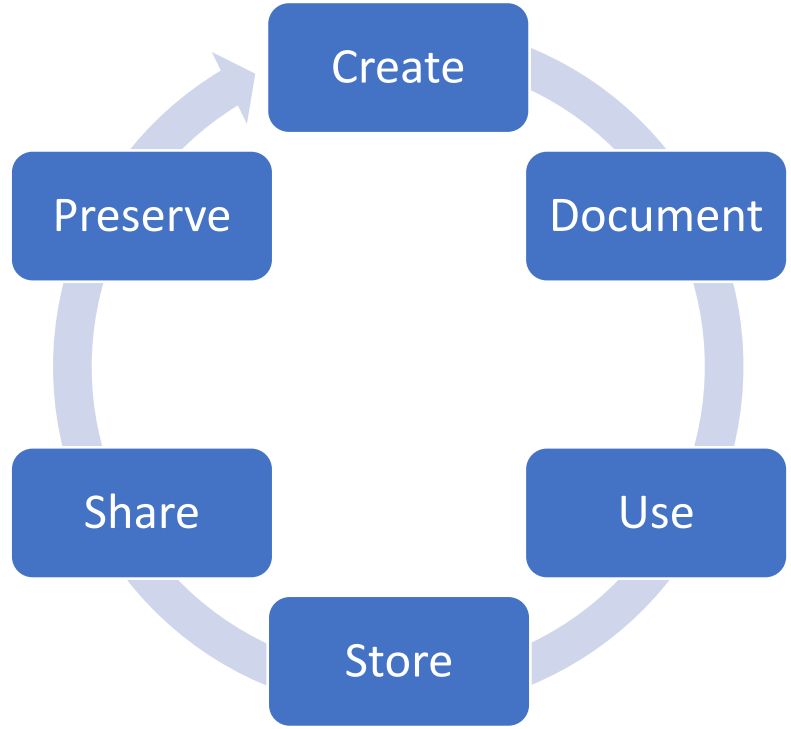
Tools for responsible and open data science



DMP ONLINE

AuthorCarpentry

Doing responsible and open data science research



Everything Has Ethical Considerations

- Research practices and the tools that underpin them have ethical implications
- Being aware of their characteristics/specifications can help you avoid building in ethical complications into your work
  - Legal requirements – ie. GitHub, repository requirements
  - Marginalization of learners/users – using CLI instead of GUI
  - Re-use of data – sharing data, not respecting restrictions
- Even the smallest tools can have significant ethical consequences

Challenges of Being Open at Home

Implementing Open and Responsible Practices in Your Own Research

- Challenges are common – everyone has them
- Physical, social and regulatory contexts influence perceptions of Open Science and ability to engage in Open Science activities

SA1/3: *I think it leads to better science*

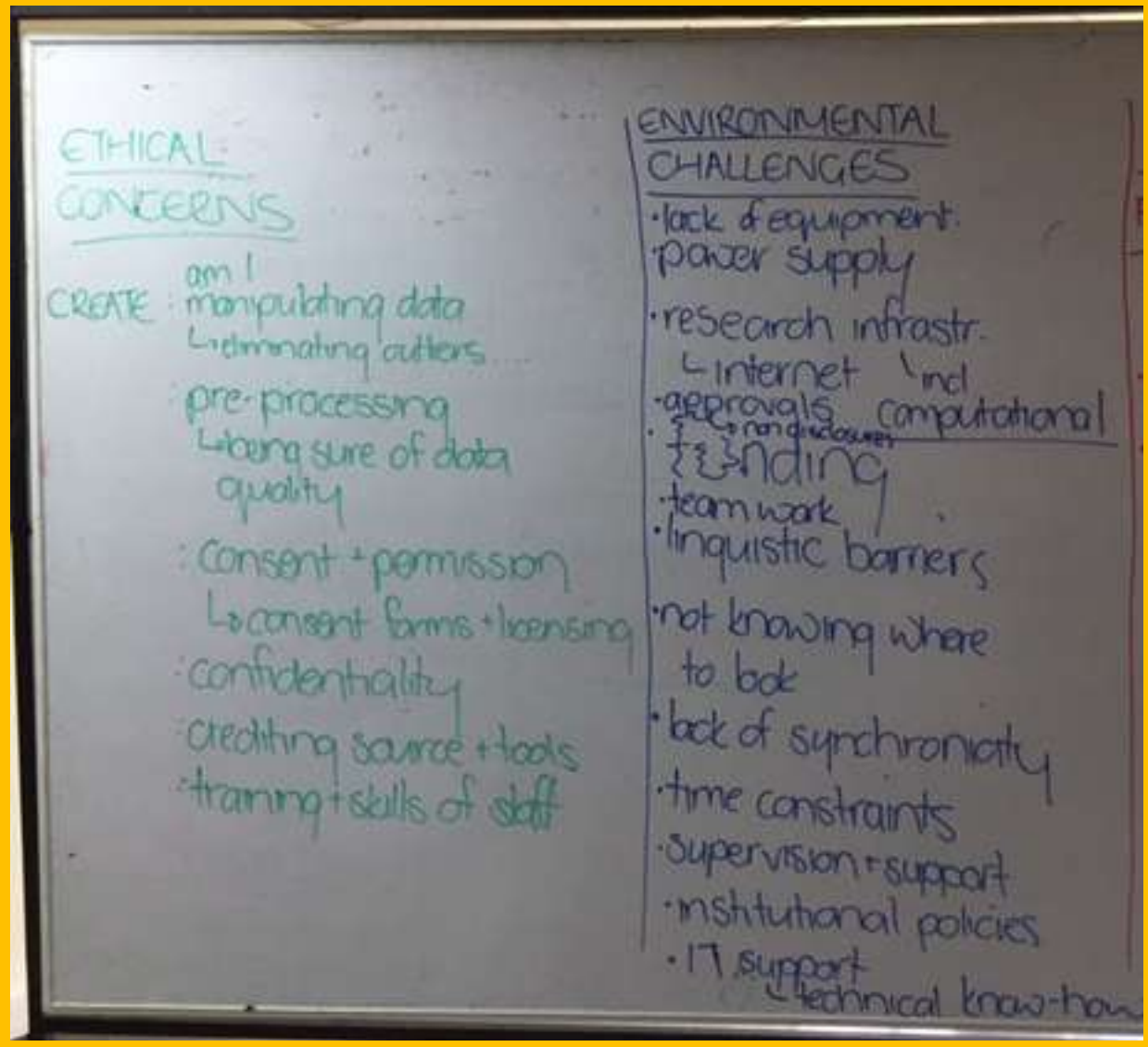
↕

KY1/1: *I won't release data unless I first of all publish*

Experiencing Challenges is Normal

What specific challenges do you anticipate encountering when you return home in terms of your data work?

- Think about specific, or general challenges. Write down 3 on post-its



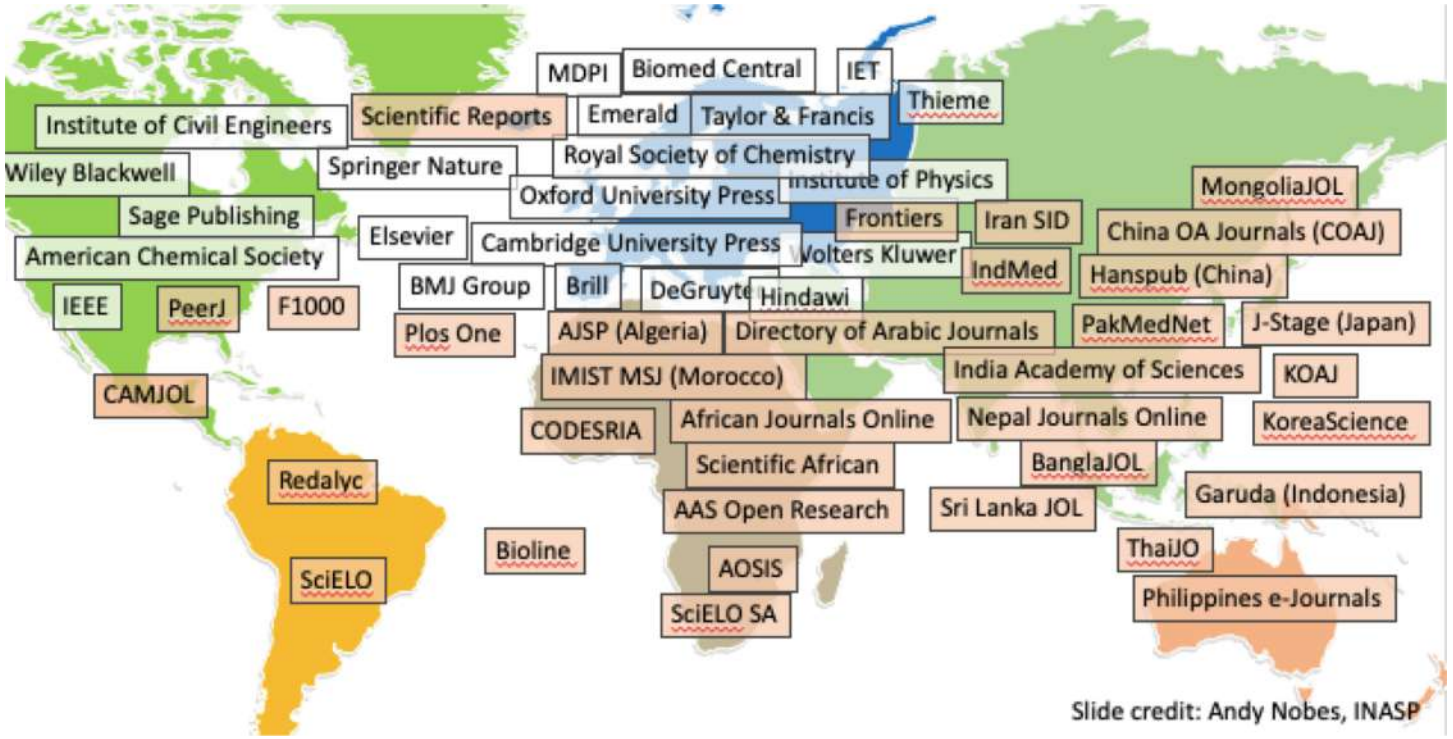
It’s OK To Have Challenges

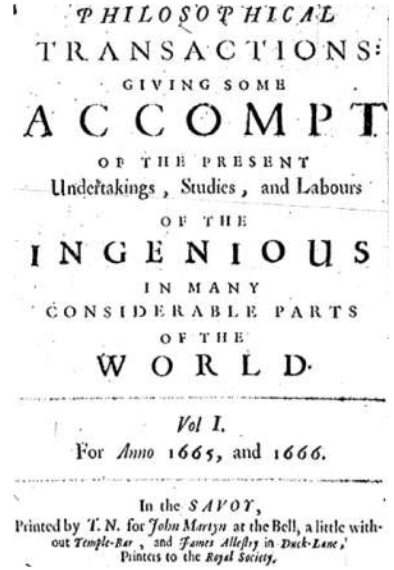
- Challenges can be categorized into a number of different areas:
    1. Cultural resistance and lack of institutional/peer support
    2. Resource limitations
    3. Personal concerns
- Challenges are not insurmountable – many resources can help address them

1. Cultural Resistance and Lack of Support

- A quick background:
  - Inherited colonial academic systems
  - Historic lack of funding and resources limiting research scope
  - “Parachute research”
- Problems include
  - Lack of institutional support
  - Lack of regulations/guidance
  - Lack of trust

1a. Getting Your Institution On Board

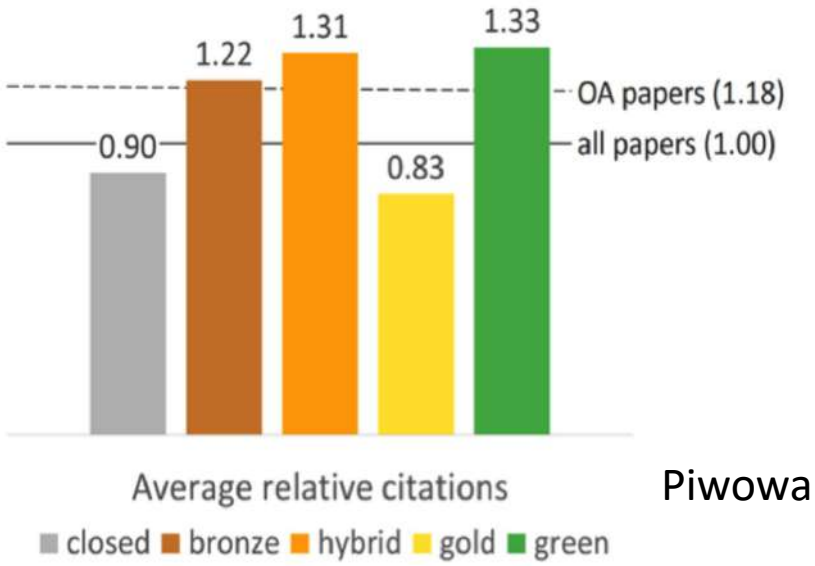




PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

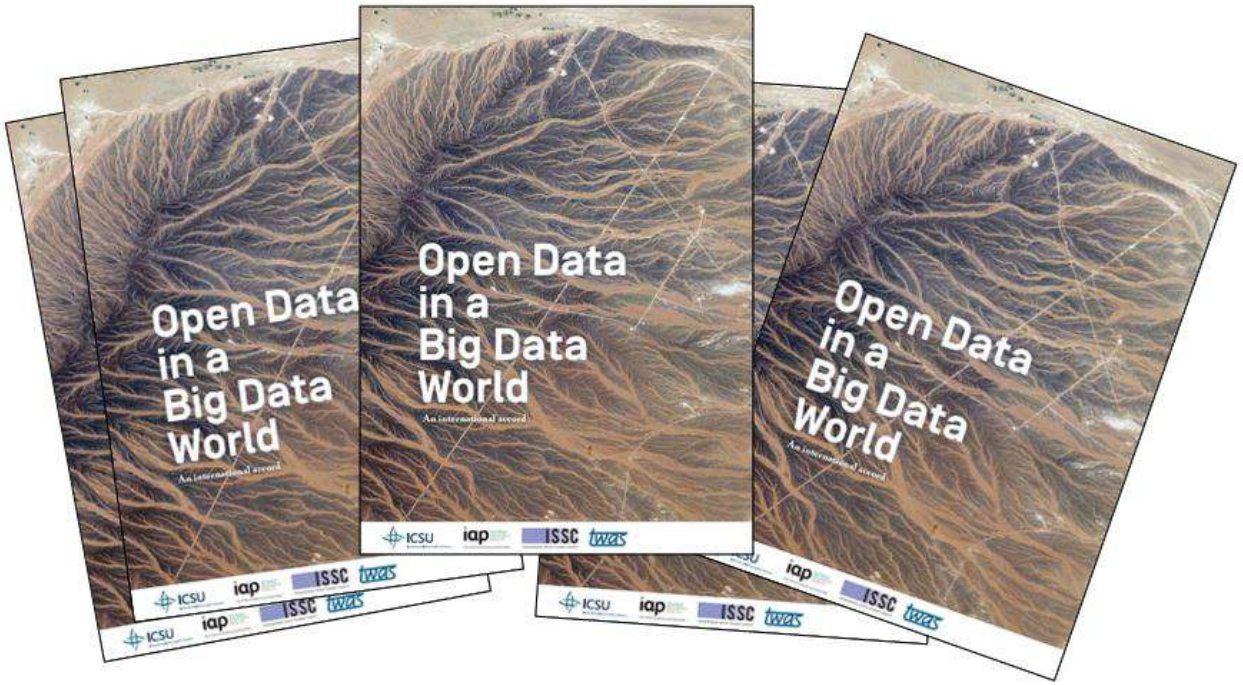


Piwowar 2018

Plan S

Making full and immediate Open Access a reality

1b. Getting Better Protection and Guidance



RDA
RESEARCH DATA ALLIANCE

Data-sharing Agreements

COPE
PROMOTING INTEGRITY IN RESEARCH AND ITS PUBLICATION

1c. Getting Over Issues of Trust

wellcome

Data-sharing Agreements

COPE

PROMOTING INTEGRITY IN RESEARCH AND ITS PUBLICATION

Still Needed: Positive Examples

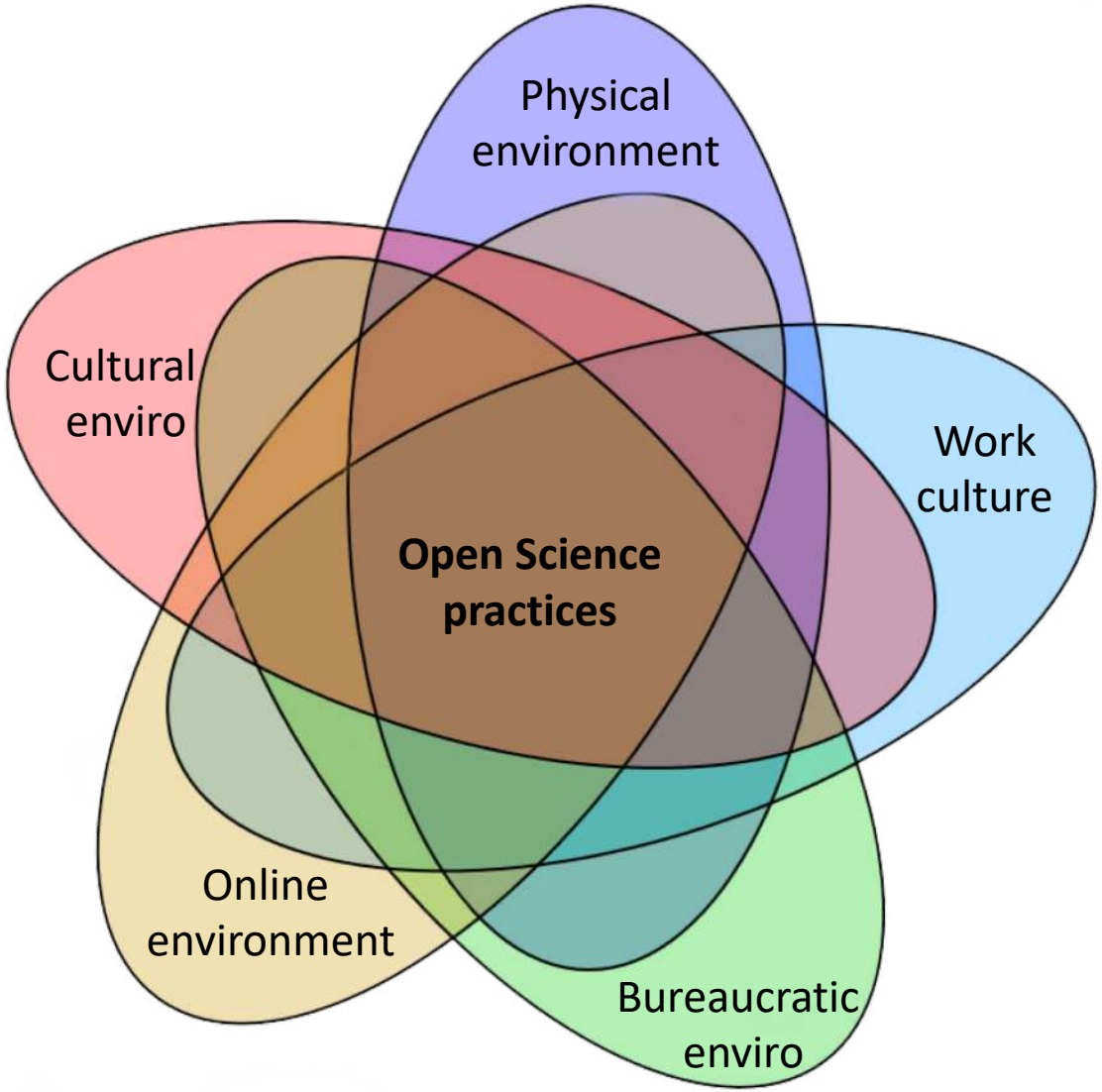
- Need for more positive examples to dispel “urban myths” and lurking ghosts
- Need enthusiastic champions and mentors
- Effective personal networks

- What else can help foster open research cultures and maximize their support?

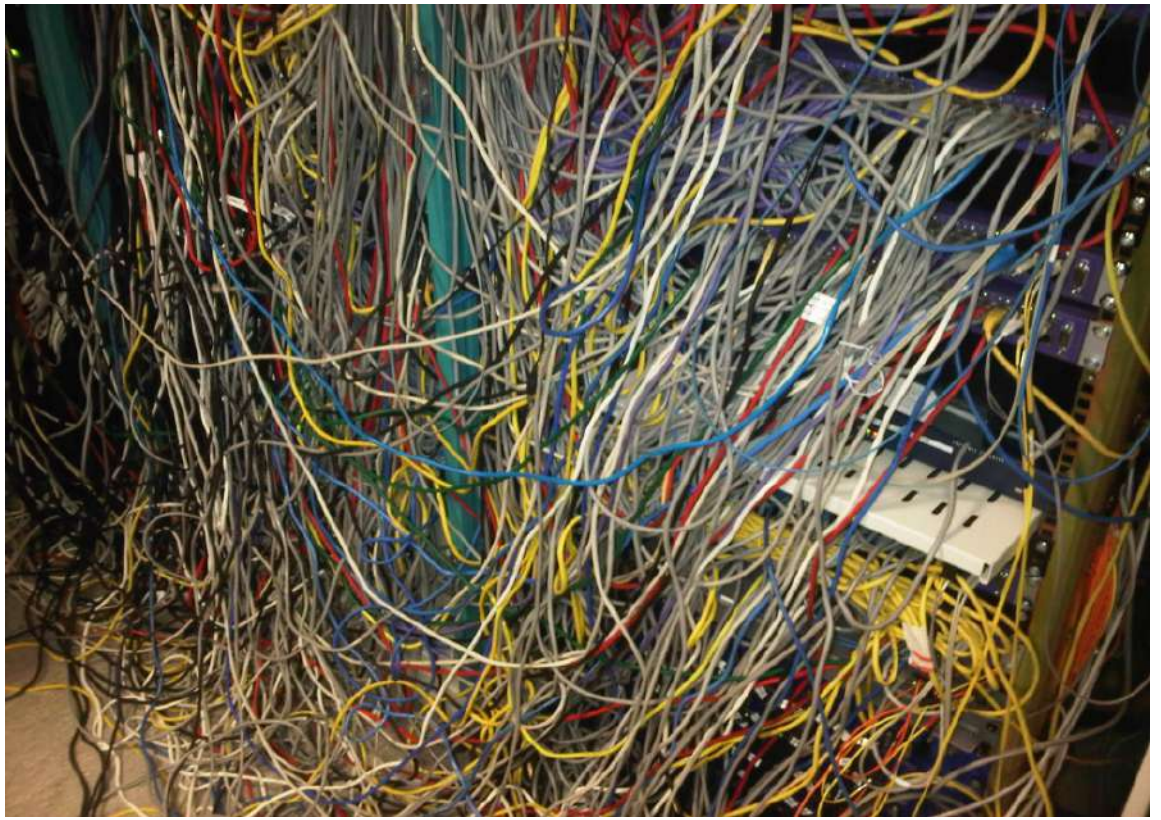
www.bbc.com

2. Infrastructures that Support Openness

- Many institutions struggle with legacies of low-resourcing
- Strategic resource distribution often means that OS activities are under-funded
  - Lack of finances to fund Open Science practices
  - Lack of ICT infrastructures
  - Lack of technical support
  - Lack of guidance

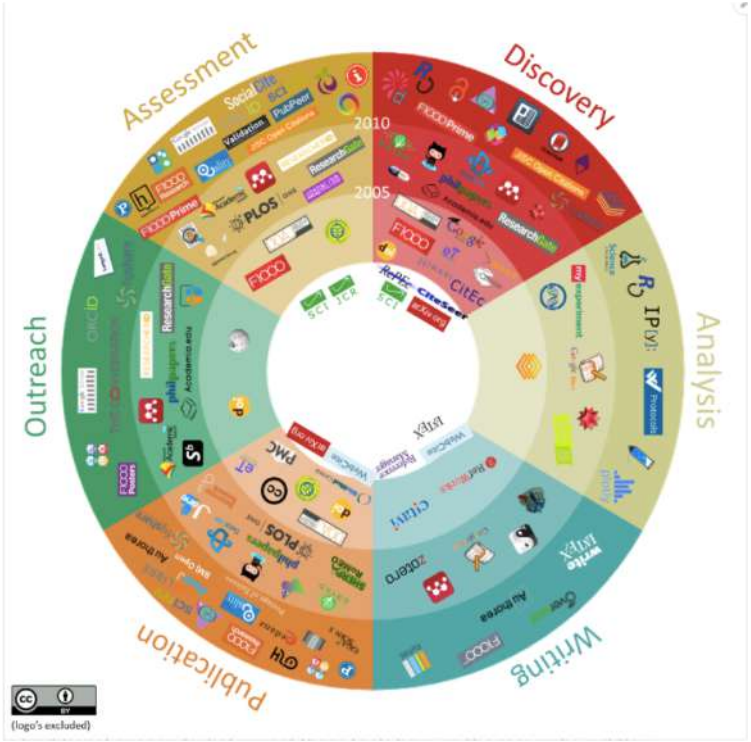


Just Because the Resources Are Online …



…. doesn’t mean they’re accessible

2a. Lack of Resources



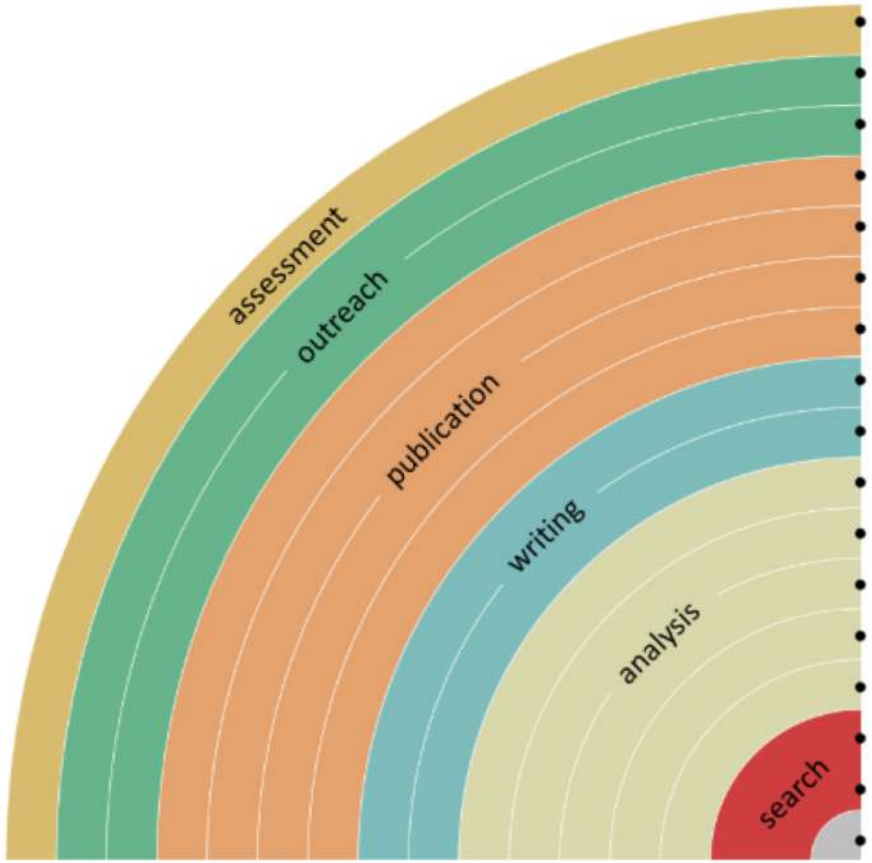
zenodo

APC waivers and discounts

BMC offers waivers and discounts for article processing charges (APCs) for papers whose corresponding authors are based in low-income countries.

BMC offers APC waivers to papers whose corresponding authors are based in countries classified by the World Bank as low-income economics as of July 2017.

You can make your workflow more open by …



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

DOI: 10.5281/zenodo.1147025

Knowing Where to Look

Research Databases and data sources

There is a wealth of research data in various databases around the world – much of it publicly available. Here are a few examples of where to look:

- Global Partnership for Sustainable Development Data www.data4sdgs.org/
- Flowminder: http://www.flowminder.org/
- Worldpop: http://www.worldpop.org.uk/
- University of Connecticut Research Database Locator: http://rdl.lib.uconn.edu/byTitle.php
- Listing of Open Access Databases (LOADB): http://www.loadb.org/
- Research4Life programme:
  - **AGORA** - Access to Global Online Research in Agriculture http://www.fao.org/agora/en/
  - **HINARI** - Access to Research for Health programme http://www.who.int/hinari/en/
  - **OARE** - Online Access to Research in the Environment http://web.unep.org/oare/
  - **ARDI** - Access to Research for Development and Innovation http://www.wipo.int/ardi/en/

African databases:

- OpenAFRICA: https://africaopendata.org/
- African Development Bank Statistical Data Portal http://dataportal.opendataforafrica.org/
- Directory of Data Repositories in Africa (DODRIA) https://researchdatadirectoryonafrica.com/
- FAO Agricultural databases http://www.fao.org/statistics/databases/en/

Offline databases:

- TEEAL (The Essential Electronic Agricultural Library) https://teeal.org/
- eGranary Digital Library https://www.widernet.org/eGranary/
- **Wiki Project Med Foundation** http://medbox.iiab.me/home/
- See also the Wikipedia list of academic databases and search engines

Thanks to Andy Nobes, INASP

2b. Lack of Expertise and Training

OPEN SCIENCE MOOC

FREE | OPEN | LEARNING

THE CARPENTRIES

CODATA-RDA Research Data Science Schools

RStudio Community

Forums

If you're looking for a forum in your native language, please check out the local user groups page at the Python Wiki.

- Python Forum (English)
- Python-Forum.de (German)
- /r/learnpython (English)

Support Networks

Academic support networks - organisations and NGOs

There are many international organisations and NGOs providing support to academics, ranging from free resources and access, training, Networking and subject-specific advice. Some useful organisations are listed below

AuthorAID www.authoraid.info

Eifl (Electronic Information for Libraries) www.eifl.net

Equator Network www.equator-network.org

CoDATA (Committee on Data of the International Council for Science) www.codata.org

Global Health Network https://tghn.org/

Global Young Academy https://globalyoungacademy.net/

Healthcare Information for All www.hifa.org

INASP www.inasp.info

Mendeley network https://www.mendeley.com/research-network/community

MedicineAfrica http://medicineafrica.com/

OWSD (Organisation for Women in Science in the Developing World) www.owsd.net

Scholars at Risk Network https://www.scholarsatrisk.org/

ResearchGate https://www.researchgate.net/

Research4Life http://www.research4life.org/

TWAS (The World Academy of Sciences for the advancement of science in developing countries) https://twas.org/

Indepth Network http://www.indepth-network.org/

International Health Policies http://www.internationalhealthpolicies.org/

Wessex Global Health Network

http://www.wessexghnetwork.org.uk/

Thanks to Andy Nobes, INASP

Still Needed

- Local investment in Open Science infrastructures may take time
- However, global infrastructures and practices are changing rapidly
- Need more LMIC voices in these discussions to make sure that they work for researchers in lower-resourced contexts

- What else can help researchers overcome resource limitations?

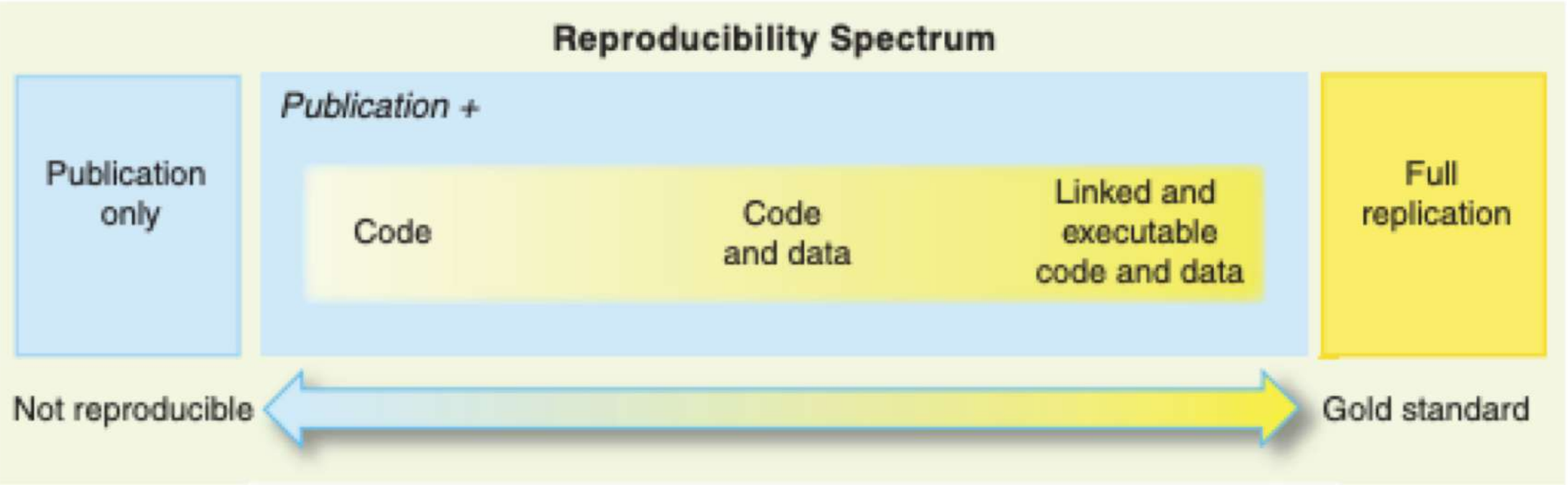
3. Personal Concerns

- As researchers we have concerns about implementing some Open Science practices
- These concerns are legitimate, and often relate to cultural and regulatory challenges
  - Concerns about being scooped
  - Concerns about scrutiny of data and methods
  - Misuse of data
  - Unintended harms

3a. Knowing Your Rights/Responsibilities

3b. Openness as a Continuum

”Your primary collaborator is yourself 6 months from now, and your past self doesn’t answer emails" (Russ Poldrack)



Peng 2011

3c. Managing Risk

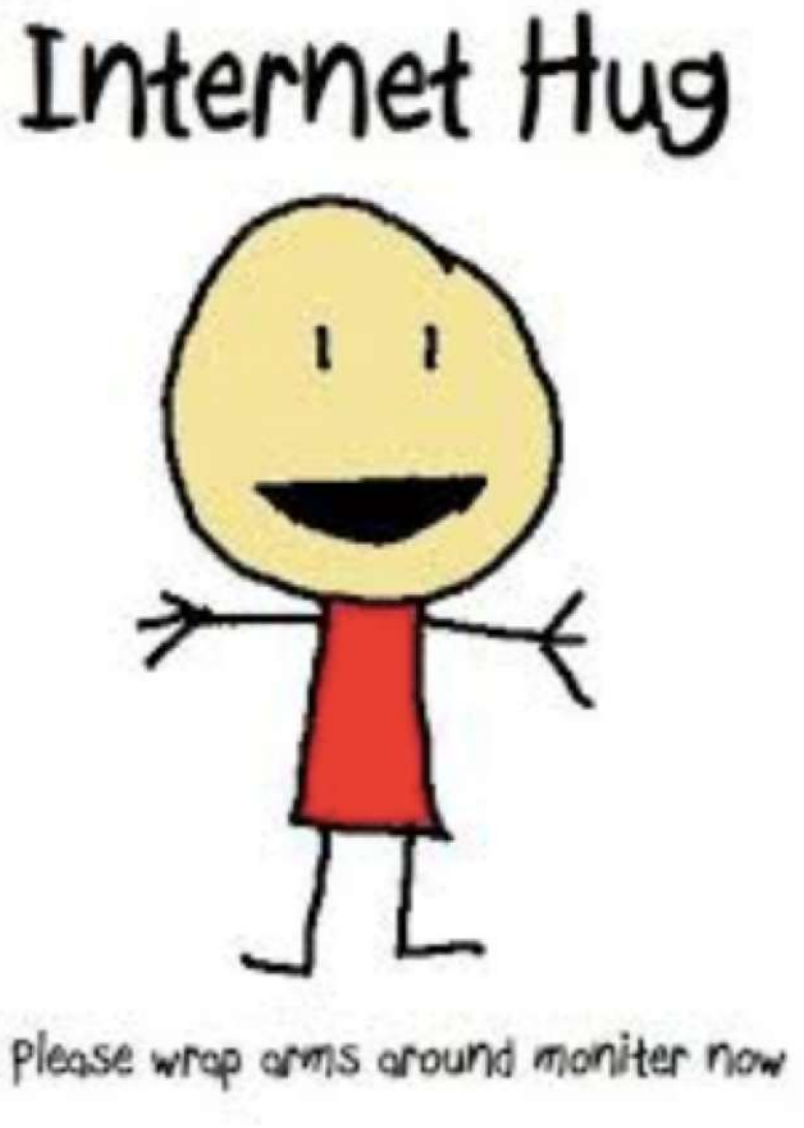
- Unintended harms are an unavoidable element of research
- Using trusted infrastructures can offset some concern as they set requirements on users and contributors
- Discuss concerns with peers – often they will have good advice

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Still Needed: More Evidence

- Lack of evidence of LMIC concerns
- Tendency to treat LMICs concerns as “same as HIC but more”
- Need more evidence about what is working, what is preferred and what is still needed
- Creating, joining and interlinking networks of support is key to fostering Open Science

- What else can help researchers overcome concerns about being open?



Openness Is A Lifelong Journey

Publish Preprints

FAIRify data

Make code available

Publish Lab-Notebooks

Use version control

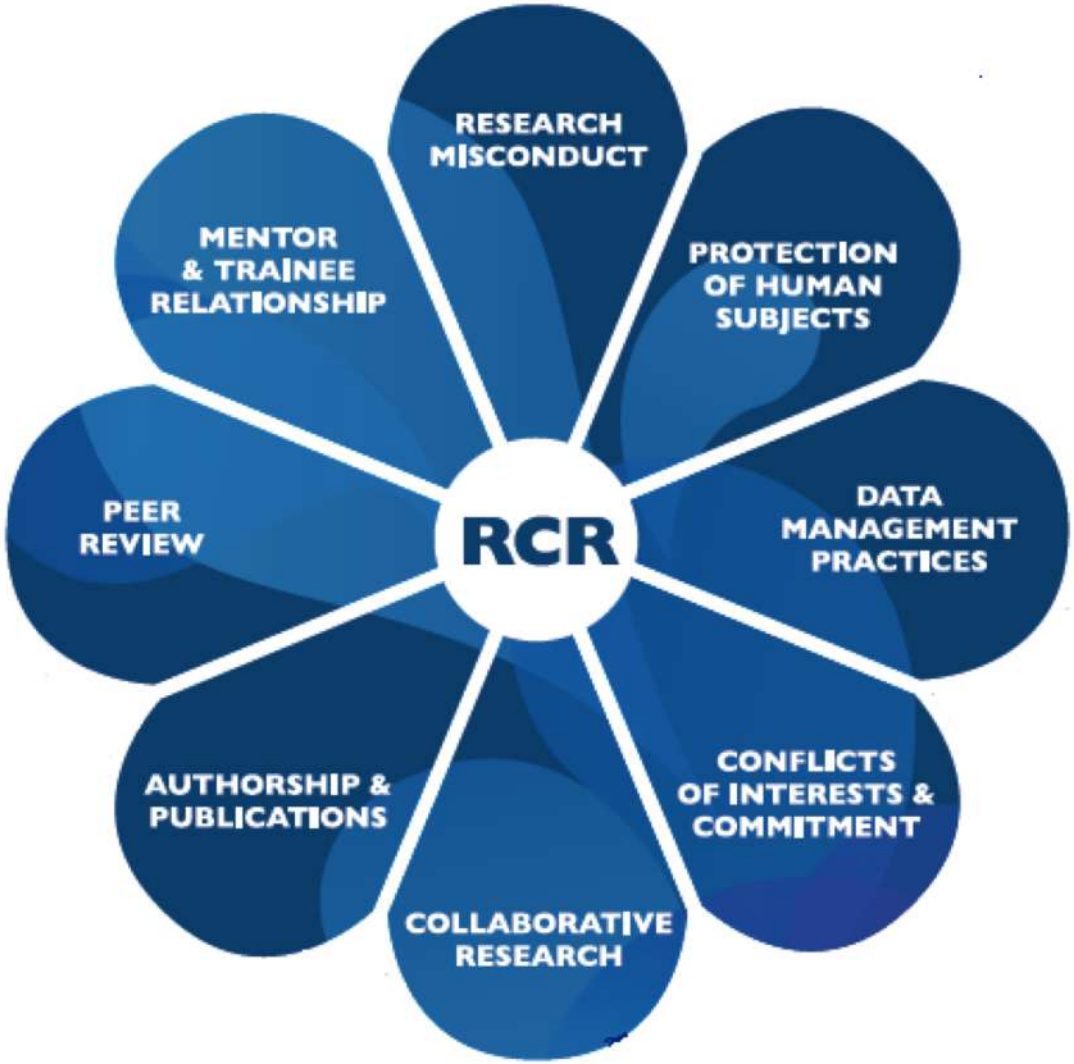
Preregister your project

Do science communication

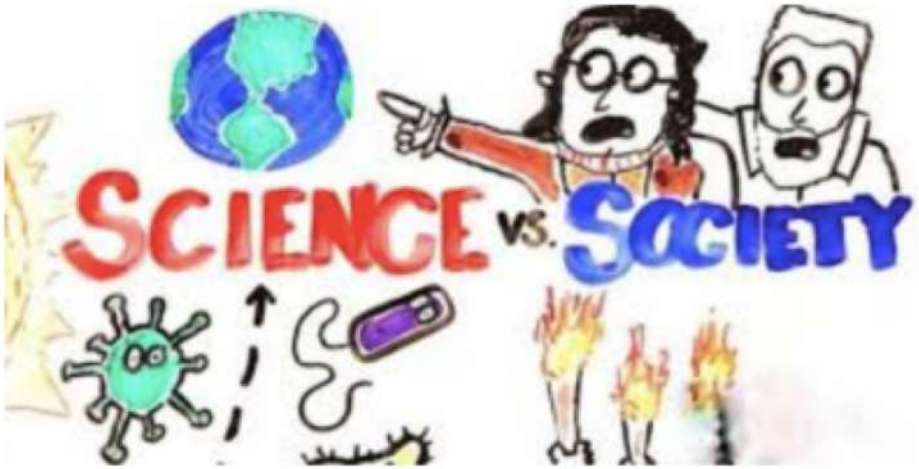
Learning to Look at the Bigger Picture

Open Science: an Extension of RCR Values

- A just distribution of resources (public funds and research products)
- A way of maximizing the benefits of research
- A safeguard against possible harms arising from research
- As a means of improving accountability and transparency
- An enactment of collegiality



Looking at the Bigger Picture: Ethical Challenges of Data Science



Challenges Beyond the Lab

- technology affects communication, collaboration and knowledge exchange within scientific, work and home settings
- need to help people to use those innovations *more productively and safely*
- need to improve ways in which new technologies can be designed and developed to be *more responsive* to societal acceptability and desirability

Not just about being open/closed. It's about making sure that you use openness as a tool to secure just futures.

You Are Part of the Bigger Picture

- Not just problems of someone else's making"

- The data that you select in your analyses can produce biases
- The algorithms that you design can perpetuate biases and stereotypes
- The websites, platforms, sharing pathways that you design, endorse or populate can perpetuate discrimination
- The data you generate can be re-used, re-combined, re-purposed in unexpected ways

Bias

Bias (parcialidad)

Inclination or prejudice (perjudicar) for or against one person or group, especially in a way considered to be unfair.

Discrimination (discriminación)

Unequal treatment of persons on the basis of ‘protected characteristics’ such as race, sexual identity etc.

Bias in algorithms

Unjustified and/or unintended deviation in the distribution of algorithm outputs, with respect to one or more of its parameter dimensions

(Desviación injustificada y / o no intencional en la distribución de salidas de algoritmos)

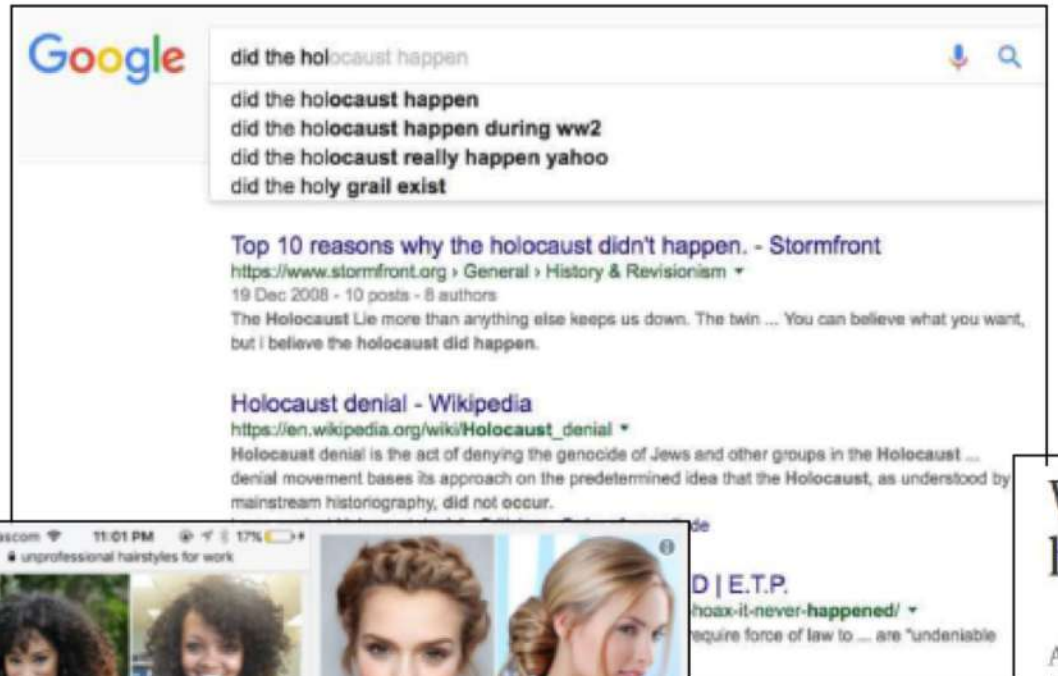
Value Ladened Nature of Algorithmic Design

"Algorithms are inescapably value-laden. Operational parameters are specified by developers and configured by users with desired outcomes in mind that privilege some values and interests over others…[O]peration within accepted parameters does not guarantee ethically acceptable behaviour… for example, profiling algorithms that discriminate against marginalised populations"

(Mittelstadt, Allo, Taddeo, Wachter, Floridi, 2016)

(La operación dentro de los parámetros aceptados no garantiza un comportamiento éticamente acceptable)

Current Challenges



Example 1: Algorithmic Decisions on Bail

USA | JUSTICE

Courts use risk algorithms to set bail: A step toward a more just system?

PROGRESS WATCH Court systems in more than two dozen US cities and states are using algorithms that assess flight risk without considering race, gender, or socioeconomic status, in an attempt to remove implicit bias from the equation.

By **Gretel Kauffman**, Staff | AUGUST 3, 2016

PROPUBLICA ... Dylan Fugett was rated low risk. (Josh Ritchie for P... Donate

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

In 2014, then U.S. Attorney General Eric Holder warned that the risk scores might be injecting bias into the courts. He called for the U.S. Sentencing Commission to study their use. "Although these measures were crafted with the best of intentions, I am concerned that they inadvertently undermine our efforts to ensure individualized and equal justice," he said, adding, "they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."

The sentencing commission did not, however, launch a study of risk scores. So ProPublica did, as

Northpointe and COMPAS

Eliminating Human Bias?

- In the early 2000s the US criminal justice system began using risk assessments to assist decision-making.
- Assessments are based on algorithmic calculations to predict, for instance, how likely an individual is to re-offend or fail to attend court for sentencing.
- Used to determine whether an individual should be granted bail or how long their sentence should be
- ‘Low risk’ offenders given shorter sentences and perhaps even kept out of jail entirely.
- Overcome human bias, or …?

Proprietary Software to Determine Risk?

- Risk assessments are now used across a wide number of states at all stages of the legal process
- Software and scores provided by for-profit companies such as Northpointe
  - Scores derived from 137 questions, either answered by defendants or pulled from criminal records. These questions related to factors such as personal offender history, family offender history, drug taking amongst friends and personal views on offending. Race was not one of the questions.
- Risk assessment scores are usually made available to the defendant's legal team
- Criteria through which the scores are generated are typically regarded as proprietary to the companies that develop them and are not released.

The Difficult Nature of Identifying Biases

- Only 20% of those predicted to commit a violent crime had gone on to do so
- Of those deemed likely to re-offend, 61% went on to be arrested, when misdemeanours such as driving with an expired license were included;
- Black people were almost twice as likely to be falsely labelled as at risk of future offending than white people;
- White people were mislabelled as low risk more often than black people;
- Even when statistical tests were run to isolate the effect of race from criminal history, recidivism, age and gender, black people were:
  - 77% more likely to be labelled as at risk of committing a future violent crime than white people
  - 45% more likely to be labelled as at risk of committing any kind of crime

Northpointe and COMPAS

Northpointe, the company that sells COMPAS, said in response that the test was racially neutral. To support that assertion, company officials pointed to another of our findings, which was that the rate of accuracy for COMPAS scores — about 60 percent — was the same for black and white defendants. The company said it had devised the algorithm to achieve this goal. A test that is correct in equal proportions for all groups cannot be biased, the company said.

Monkey Cage

A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.

By Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel
October 17, 2016

Most Read Politics

1 'Poor man's version of Don King': Trump continues his war of words with LaVar Ball

2 Analysis
President Trump and accusations of sexual misconduct: The complete list

3 White House military personnel removed amid investigation into contacts with foreign women during Trump's Asia trip

Even the Smallest Decisions Can Introduce a Bias

- Unequally wrong for false positives in different populations = unfair (Pro Publica argument)
- Equally right in predicting reoffending = fair (Northpointe argument)
- Base populations have different levels of reoffending so algorithm cannot be equally wrong and equally right for both populations
- Technical measures to ‘correct’ for societal unevenness?
- **Transparency and accountability is necessary to enable individuals to challenge algorithm-based decision making that affects their lives**

Even the smallest technical decisions can influence biases

Using a single assessment of “right” or “just” can cause biases to perpetuate

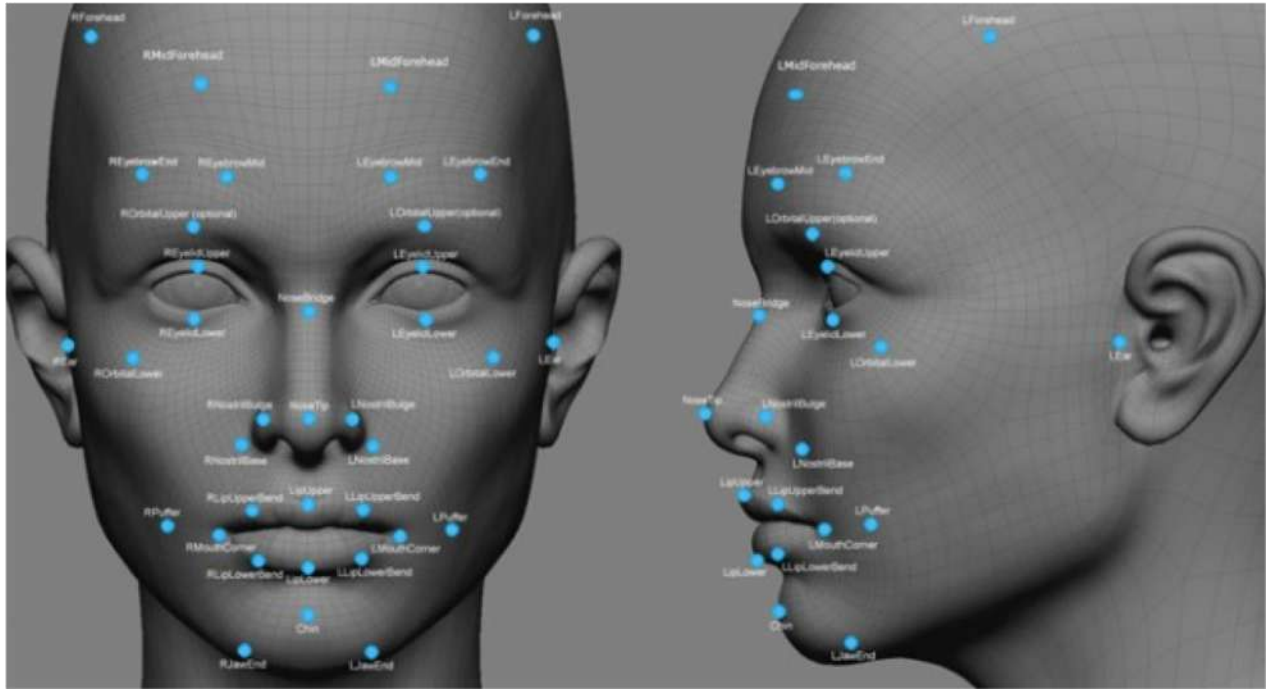
Campaigning for Justice: ProPublica

- Increasing amount of discussion about use of Northpointe COMPAS in judiciary
  - Proprietary algorithms determining individual futures
  - Inability to scrutinize processes through which decisions are made unjust
  - Uncritically accepting algorithmic decisions can mean that the justice system is failing in duty of care

PROPUBLICA

Example 2: Facial Recognition Software

- Joy Buolamwini (MIT)
- Software created by brand-name tech firms such as Amazon uncovered much higher error rates in classifying the gender of darker-skinned women than for lighter-skinned men.
- Other problems – unable to reliably detect Asian eyes
- Location of software companies and demographics = non-representative datasets used in algorithm development

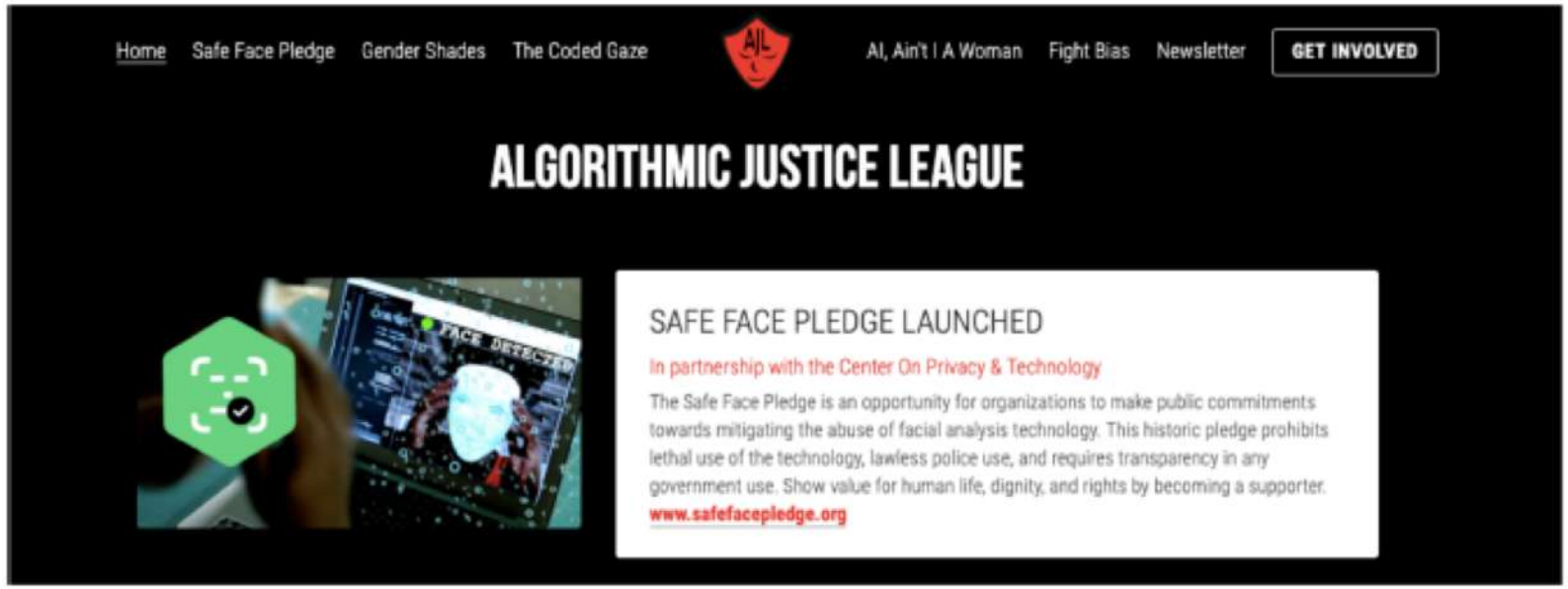


Significant Harms From Deployment of Algorithms

- Called on Amazon to stop selling its facial recognition software to police.
- Caution about the fast-moving adoption of facial recognition by police, government agencies and businesses from stores to apartment complexes
- Computer vision systems that enable self-driving cars to “see” the road shows they have a harder time detecting pedestrians with darker skin tones.

Algorithmic Justice League

- Algorithmic activism
- Name and shame companies
- “Safe Face Pledge” – address bias, facilitate transparency, promote dignity and human rights



On Correcting for ‘Real World’ Bias

algorithms are inherently politicised *[as connected to social policy and political power]… and reflect our current world view, our current social policy … If we are not explicit about that as well, if we are not transparent about that, that we value equality between men and women, then we are again creating bias at another level of the system* (Jirotka 2016)

What Causes Bias?

…. among the major factors the contribute to bias in the results that [systems] produce is because there is bias in the data. So you actually have to look at the data as far as the performance is concerned, to make sure you have a representative sample of the population you are trying to model (Mittelstadt, Allo, Taddeo, Wachter, Floridi, 2016)

Bias in data selection

Use of unrepresentative datasets in algorithm development

we have to think about how to rebalance the data so that that discrimination is not propagated through the algorithms. How does one come up with a fair set of data, which can actually challenge the biases that might naturally be there …

Not as easy as it sounds …

A Vision for Algorithmic Design

We want our algorithms in a sense to follow a higher values, ***moral values*** *that we think are more important than giving an exact reflection of the world. And that I think is a very interesting, but also in a sense very shady area in which, are we going to use the data as it is? Or are we going to change the data, or not change but adapt the way we look at the data to serve our purpose of being non-discriminatory…*

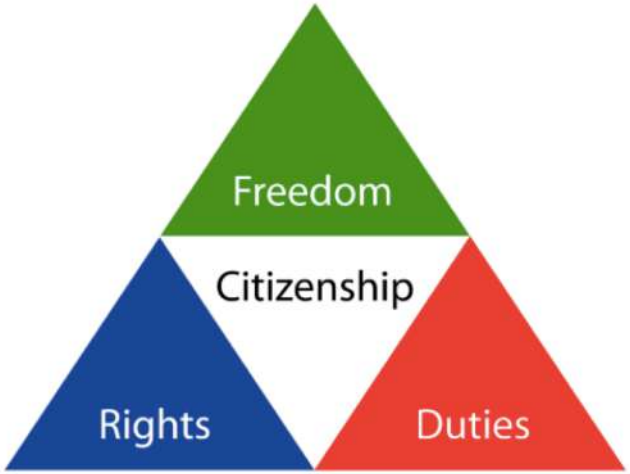
What Can I Do?

- Get independent researchers to check your code/data selection/results to expose biases
- Always critically examine the decisions you're making in your research and ask "why do I think that way"?
- Be critical of the code and results you're using – how did they get to the point they did?
- Think about how other cultures will respond to your decisions

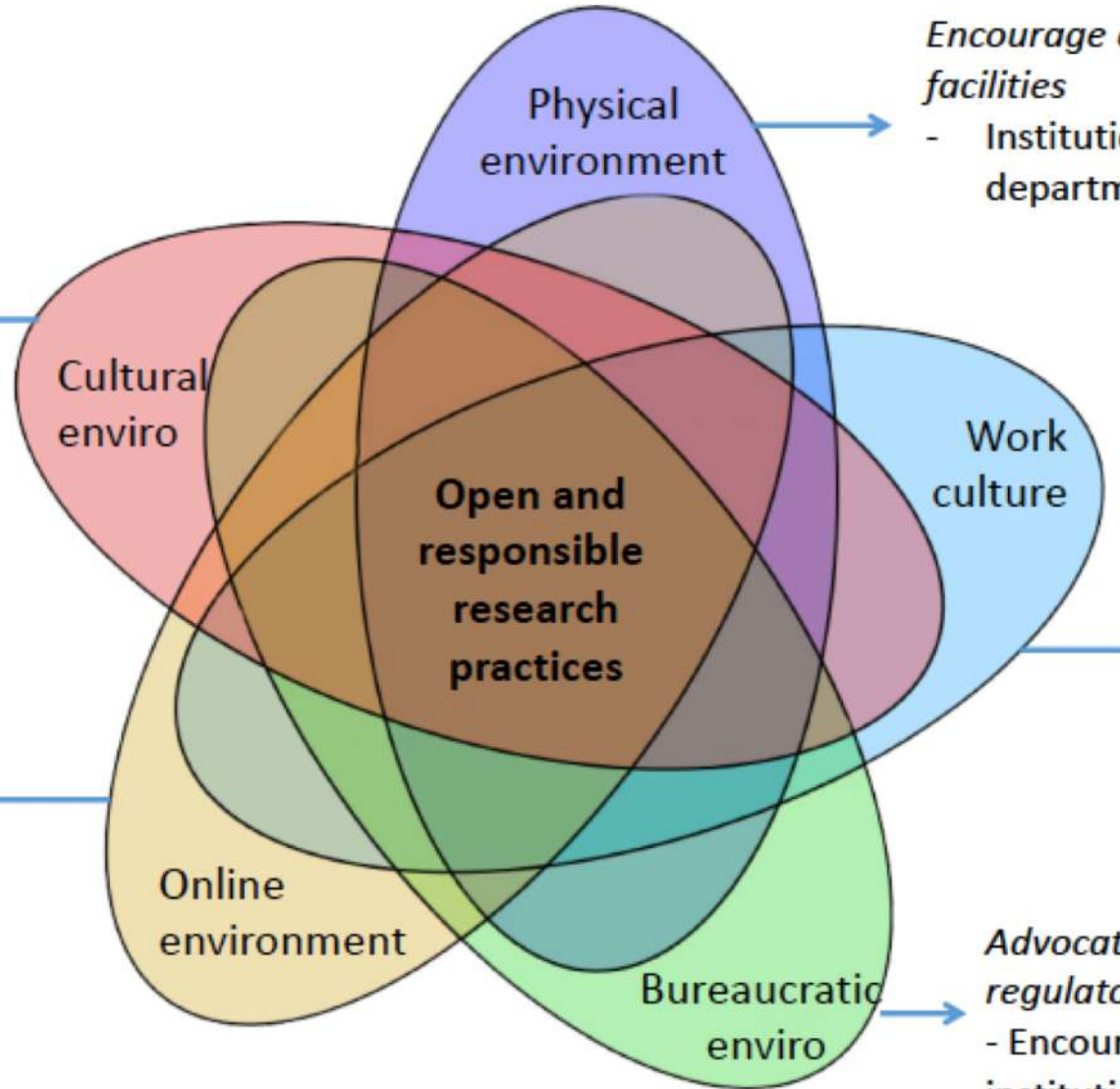
Do the data/coding choices you made contribute to just present and futures? Are you upholding the moral values of societies?

Individual Activities … Global Impact

- Being a responsible and open science citizen involves more than just making sure that your own data practices are ethical
- Being a “citizen” of the data community comes with responsibilities to the scientific community, public and future
- Not just about responsible and critical use of data, also about scrutinizing evolving systems



Extending Data Science Citizenship Responsibilities



Encourage development of open storage facilities

- Institutional repositories, centralized departmental sharing

Engage with public on data-related issues

- Use expertise to engage with common concerns and misconceptions
- Contribute to public skill development

Improve openness and responsibility in your work environment

- Advocate for RDM practices
- Engage colleagues in discussions on RCR, RDM, FAIR
- Encourage data dissemination beyond disciplines

Scrutinize data and platforms currently in use for biases

- *and re-use data*
- Carefully scrutinize your own code and design for biases
- Flag up platforms are exclusionary
- Biases in algorithms etc
- Advocate for improving community resources

Advocate for development of data regulatory frameworks

- Encourage development of data policies at institutional and policy level

Outline for Next Week

Over the course of next week, reflect on the tools that you are going to be taught. Think about:

1. How you can safeguard *beneficial* outcomes of your activities in data gathering, infrastructure building and data dissemination?
2. How can you discuss these issues with your colleagues and peers?
3. How can you scrutinize the systems/datasets you will work with to make sure that biases do not creep into your research systems?
4. How can responsible and open science citizen strengthen these activities?

Thank You

Please feel free to contact me with any further questions!

Louise.Bezuidenhout@insis.ox.ac.uk