

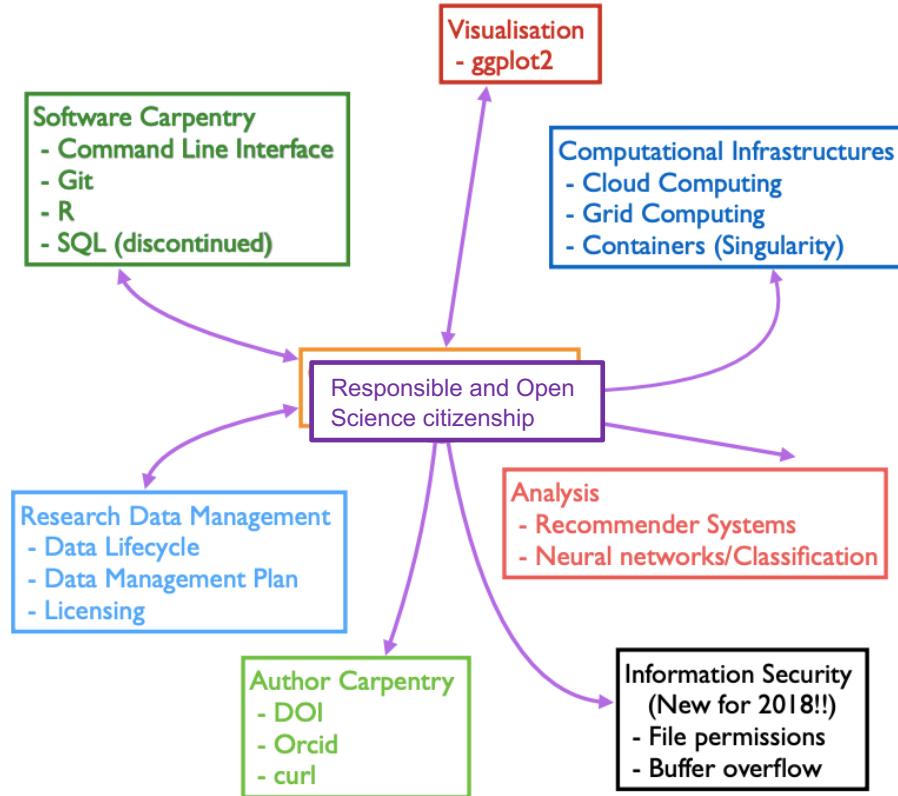


Data
Schools

Open and Responsible (Data) Science Citizenship 1

Material from Louise Bezuidenhout,
presented by Marcela Alfaro Córdoba

Our Curriculum



Plan for the first part of the Morning

1. Responsible conduct of research
2. Open science as a new way of being responsible
3. Being a responsible, open science citizen

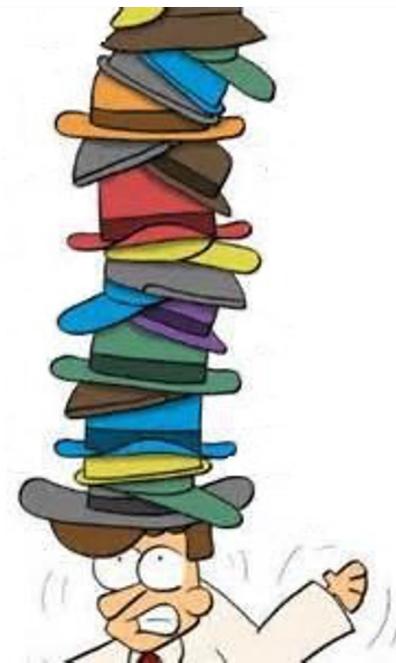
Not just about learning data science ... learning responsible data practices

Responsible Conduct of Research

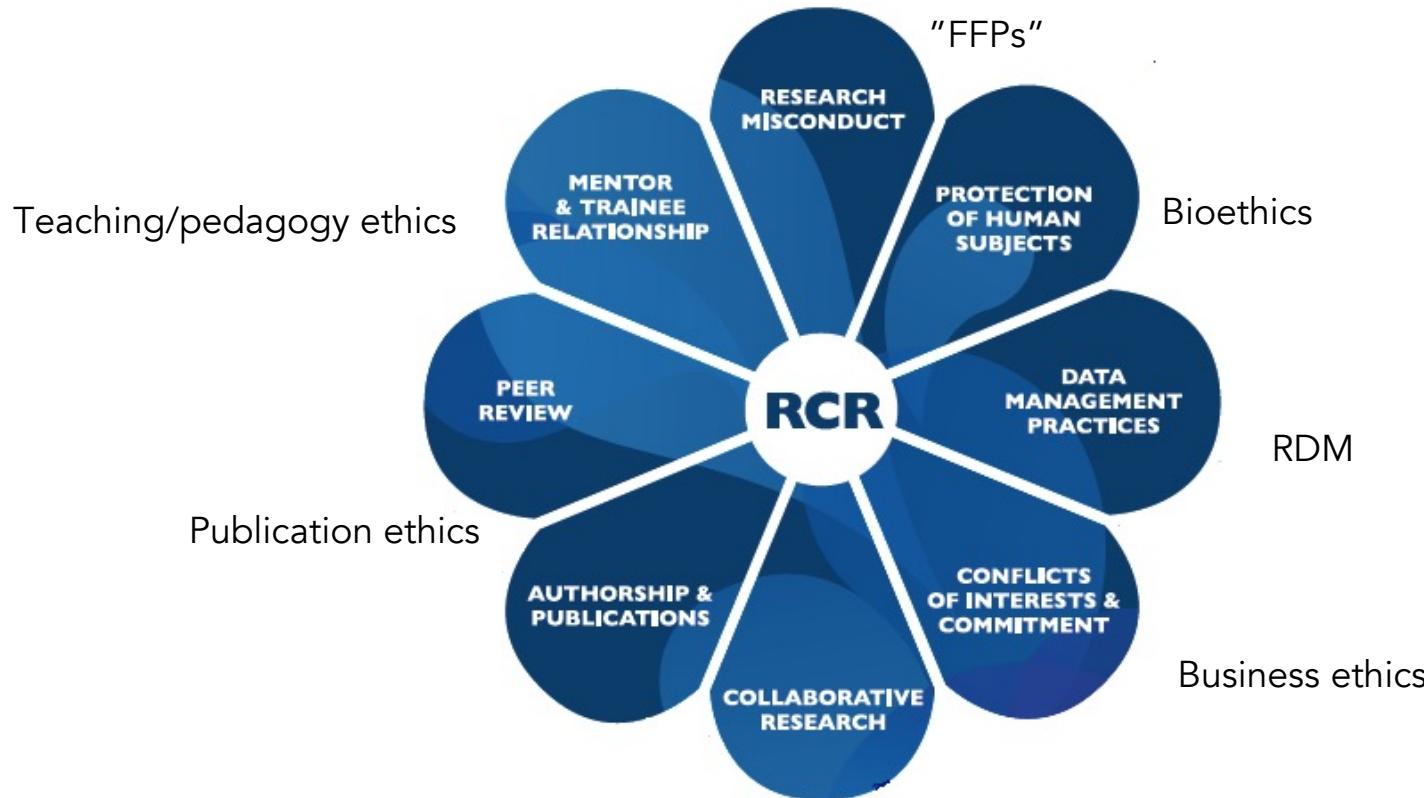
- Not just about being
 - “good at your work”
 - producing data
 - getting on with your colleagues
- As knowledge producers, educators and recipients of public funds we have additional responsibilities as researchers

Balancing Multiple Roles as a Researcher

- Data producer
 - Data user and/or collaborator
 - Author
 - Employee
 - Teacher/mentor
 - Recipient of public funds
 - Recipient of public trust
 - Citizen/legally-obligated individual
- *How do all these roles fit together?*



Responsible Conduct of Research



Responsibility as a Product of Integrity

- Practice of scientific investigation with integrity

integrity

/ɪn'teɡriti/ 

noun

1. the quality of being honest and having strong moral principles.

"a gentleman of complete integrity"

synonyms: honesty, uprightness, probity, rectitude, honour, honourableness, upstandingness, good character, principle(s), ethics, morals, righteousness, morality, nobility, high-mindedness, right-mindedness, noble-mindedness, virtue, decency, fairness, scrupulousness, sincerity, truthfulness, trustworthiness

"I never doubted his integrity"

2. the state of being whole and undivided.

"upholding territorial integrity and national sovereignty"

synonyms: unity, unification, wholeness, coherence, cohesion, undividedness, togetherness, solidarity, coalition

"internal racial unrest threatened the integrity of the federation"

Key Ethical Norms

- Responsible research involves the awareness and application of professional norms and ethical principles in all areas relating to research
- Beneficence (do good)
- Non-maleficence (cause no harm)
- Accountability
- Transparency
- Care/Equity
- Collegiality



Codes of Conduct

- Ethics are often outlined in disciplinary codes of conduct
- Can be helpful ways of summarizing areas of activity to think about in relation to RCR

ACM Code of Ethics and Professional Conduct

1. GENERAL ETHICAL PRINCIPLES.

1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.

1.2 Avoid harm.

1.3 Be honest and trustworthy.

1.4 Be fair and take action not to discriminate.

1.5 Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.

1.6 Respect privacy.

1.7 Honor confidentiality.

2. PROFESSIONAL RESPONSIBILITIES.

2.1 Strive to achieve high quality in both the processes and products of professional work.

2.2 Maintain high standards of professional competence, conduct, and ethical practice.

2.3 Know and respect existing rules pertaining to professional work.

2.4 Accept and provide appropriate professional review.

2.5 Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.

2.6 Perform work only in areas of competence.

2.7 Foster public awareness and understanding of computing, related technologies, and their consequences.

2.8 Access computing and communication resources only when authorized or when compelled by the public good.

2.9 Design and implement systems that are robustly and usably secure.

3. PROFESSIONAL LEADERSHIP PRINCIPLES.

3.1 Ensure that the public good is the central concern during all professional computing work.

3.2 Articulate, encourage acceptance of, and evaluate fulfillment of social responsibilities by members of the organization or group.

3.3 Manage personnel and resources to enhance the quality of working life.

3.4 Articulate, apply, and support policies and processes that reflect the principles of the Code.

3.5 Create opportunities for members of the organization or group to grow as professionals.

3.6 Use care when modifying or retiring systems.

3.7 Recognize and take special care of systems that become integrated into the infrastructure of society.

4. COMPLIANCE WITH THE CODE.

4.1 Uphold, promote, and respect the principles of the Code.

4.2 Treat violations of the Code as

<https://www.acm.org/code-of-ethics>

Do you know a code of conduct or a code of ethics and professional conduct for your area? If so, which ones?



- 1** Go to **PollEv.com**
- 2** Enter **UCSCSTAT5F22**
- 1** Text **UCSCSTAT5F22** to **37607**
- 2** Text in your message

Total Results: 0

RCR as a Collaborative Endeavour



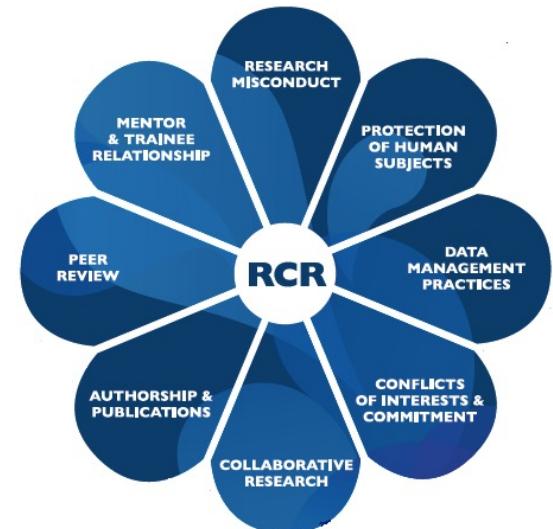
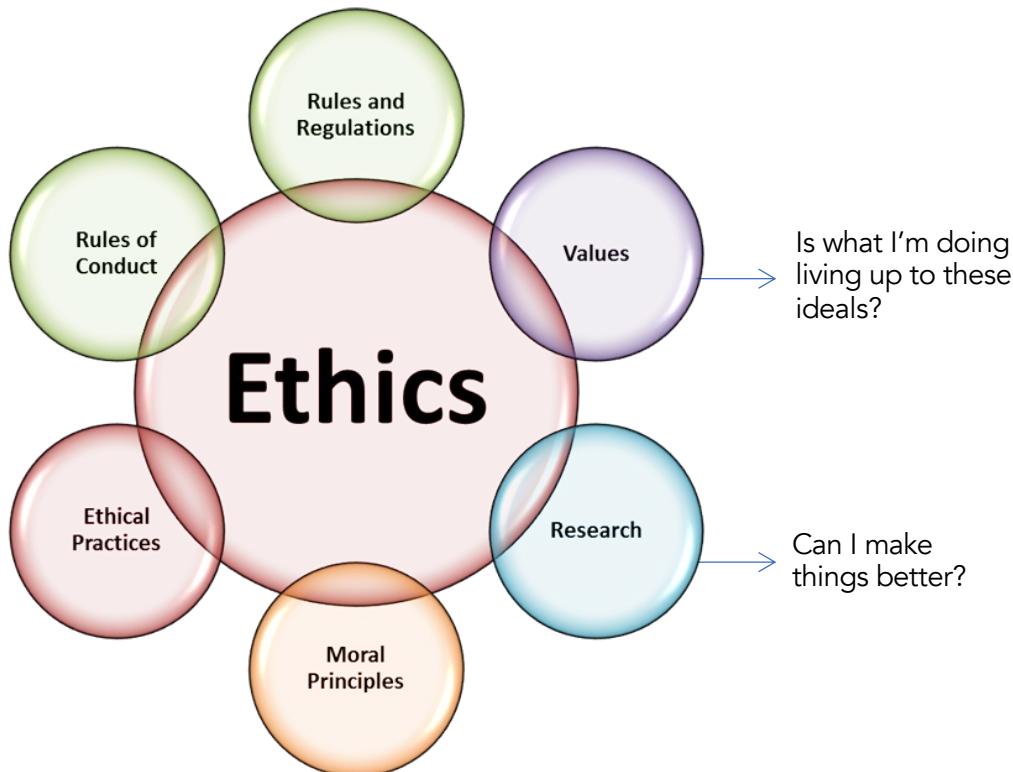
Institutional Responsibilities

- Education
 - Infrastructures that support responsible research
 - Capacity for dealing with concerns/whistleblowing/misconduct
 - Engagement and compliance with inter/national regulation
-
- *The roll-out of RCR can be patchy. We will discuss this more in the next ethics lecture.*

RCR: Changing Practices and Environments

- Understanding of responsible research, researchers and research contexts is changing
- Individual responsibility – expect compliance with regulations, active engagement with RCR behaviours
- Institutions/National/International systems – expect investment in RCR-supporting infrastructures, develop systems of educating, monitoring and mediating
- International science community – expect monitoring, support and capacity building

RCR as an Extended Form of Research Ethics



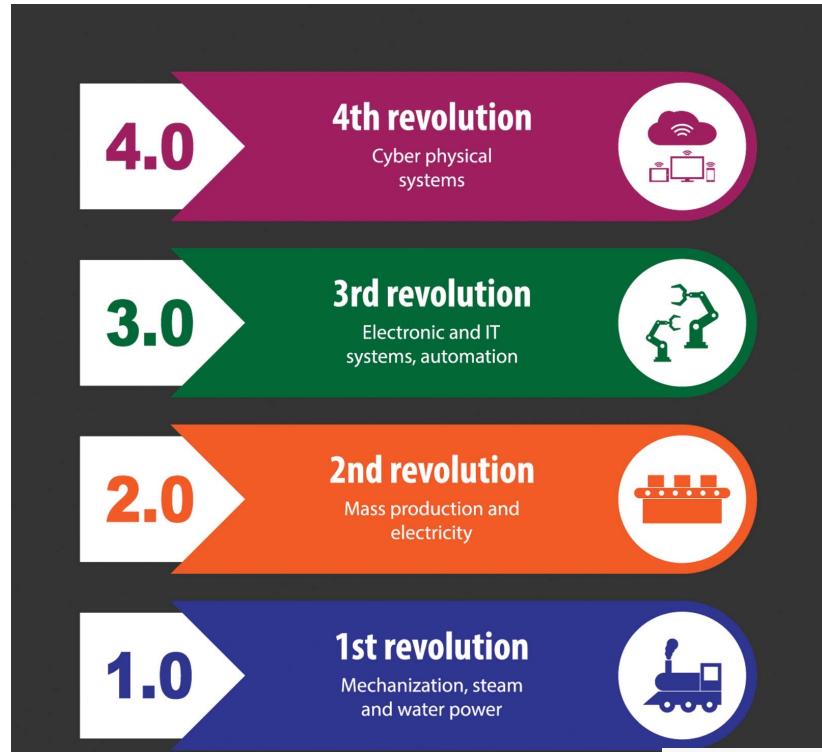
Fostering ethical excellence instead of avoidance of misconduct

Extending the Reach of Ethics in Research

- Produce verifiable and re-usable data
- Protect scientists and societies from harm
- Enable collaboration
- Ensure investments (financial, trust, time etc) are recompensed
- Embeds science within cultural/social priorities

The Digital Revolution

- Data “deluge”
- Big Data
- AI
- Increasing opportunities and prioritization of interdisciplinary research



The Digital Revolution: Coming at the Right Time?

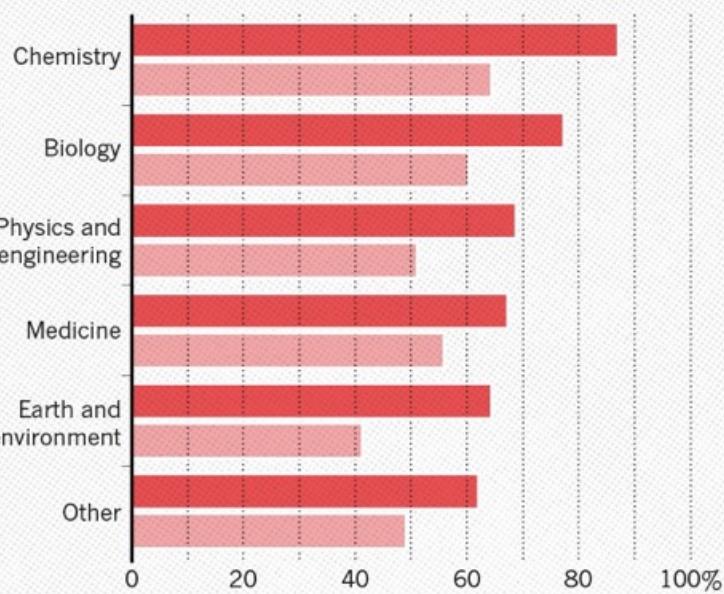
- Advances in computing have not only changed research, but come at a time of significant change in research
- A number of issues are changing the way we think/talk/govern research
 - Reproducibility crisis
 - Increased scrutiny of public investment in research
 - Increased pressure for public access to research resources
 - Sustainable Development Goals (SDGs)
 - Blurred boundaries between academia, commerce and government
 - Citizen science movement
- Driving forward new models of research practice

The Reproducibility Crisis in Research

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

● Someone else's ● My own



Same Data, Different Conclusions

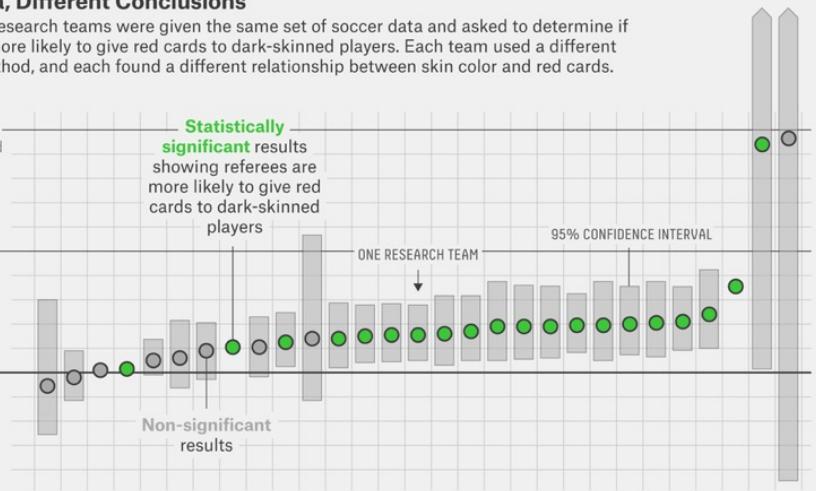
Twenty-nine research teams were given the same set of soccer data and asked to determine if referees are more likely to give red cards to dark-skinned players. Each team used a different statistical method, and each found a different relationship between skin color and red cards.

Referees are
three times as
likely to give red
cards to
dark-skinned
players

Statistically
significant
results
showing referees are
more likely to give red
cards to dark-skinned
players

Twice as likely

Equally likely

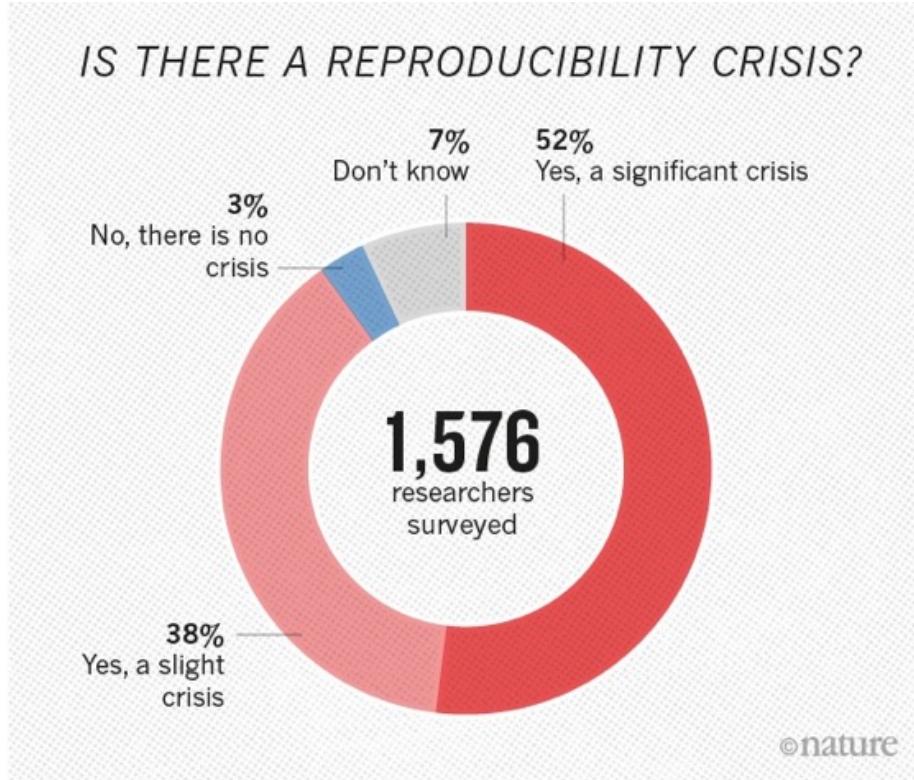


FIVETHIRTYEIGHT

SOURCE: BRIAN NOSEK ET AL.

- Nature survey of 1,576 researchers (Baker et al 2016)
- <https://psyarxiv.com/qkwst/>

A General Consensus on the Crisis



- Variability in analysis and methodology
- Incentives aligned towards publication not reproducibility
- Lack of transparency and access to data

Nature survey of 1,576 researchers
(Baker et al 2016)

When poll is active, respond at **PollEv.com/ucscstat5f22**

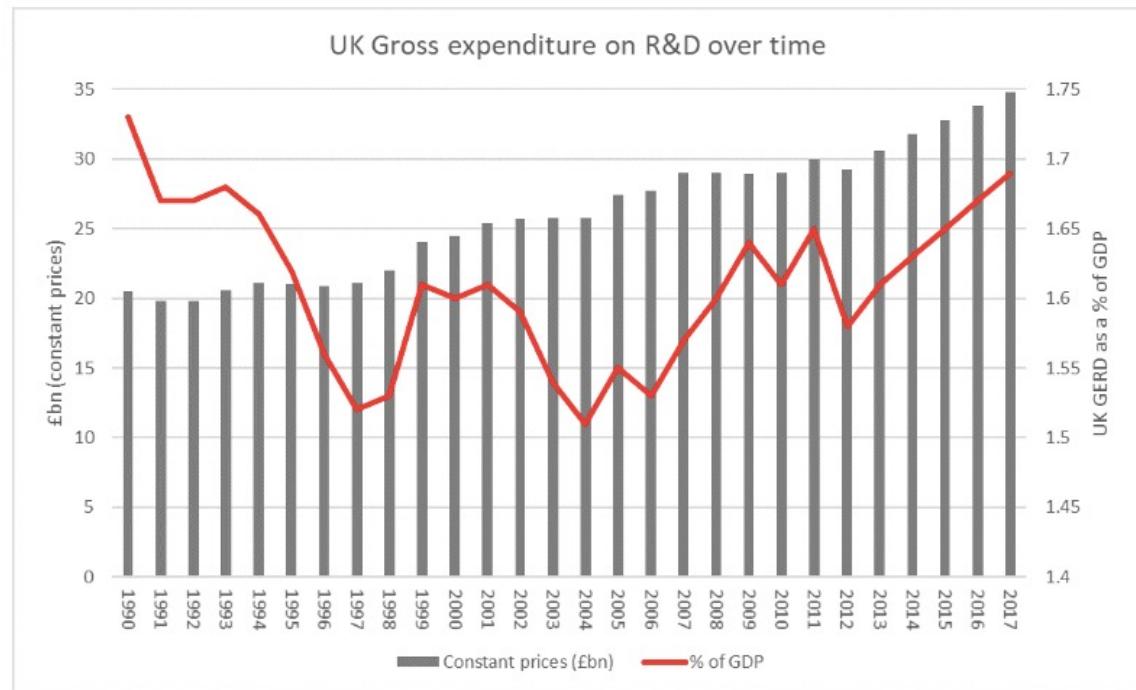
Text **UCSCSTAT5F22** to **37607** once to join

Have you tried to reproduce the findings of one study? either yours or from another research group

Yes
No
I don't know

Public Benefit: Returns on Public Investment

- Increasing scrutiny of public investment in research projects and infrastructures
- Changing models of university financing





SDGs, Research and Global Good



Increased support and funding for research that addresses SDGS and contributes to global development

New Challenges for Responsible Research Models

- Expanding scope of responsibility (led by SDGs)
- Digital systems and distributed research making individual role responsibilities complicated
- Inclusion of non-researchers in research process raises concerns about ethical practice
- Urgent need to (efficiently) use and re-use volume of data being created raises concerns about
 - Opportunity to share vs loss of control
 - Increase benefits of research for public vs possible harms
 - (Un)Intended marginalizations
 - Data recombination and re-use

New Challenges for Responsible Research Models

- How can the evolving power of digital technologies be harnessed to uphold the principles of responsible research and increase reproducibility?
- How can the culture of science be adapted to support this evolution?
- How can practices and structures of scientific research be adapted to ensure that research benefits the most number of people?

How do we ensure that we create presents and futures that uphold ethical principles and allow research with integrity?

Time For A New Approach?



Increase trust in science, don't waste public resources



Get constructive feedback

We need to start to effectively share



Be international and inclusive



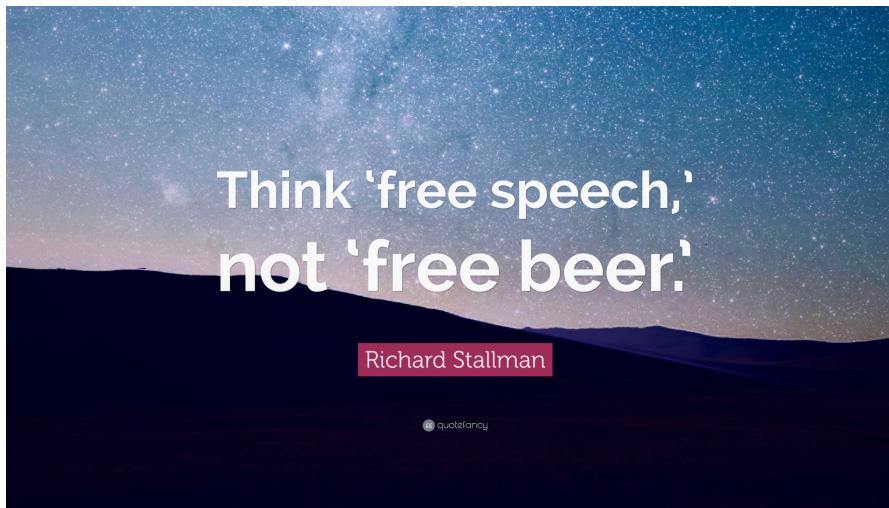
Increase the speed of discovery

Is Openness in Research the Answer?

- There is an historical precedence for this argument:
 - Openness is a core value of science/research
 - Merton 1942
 - Communalism
 - Universalism
 - Disinterestedness
 - Organized skepticism
- Long tradition of sharing resources and scrutinizing research
- Transparency of data and sharing of resources key to addressing issues of reproducibility, networking and public trust
- Addresses need for research to be a common undertaking for the common good

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



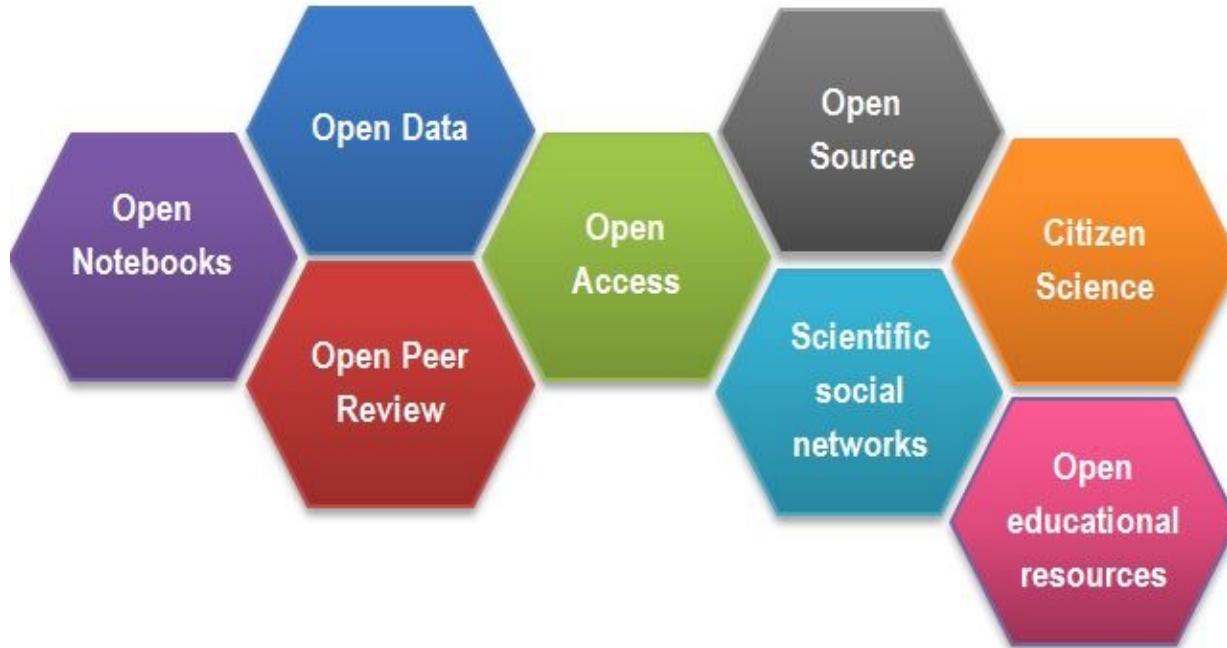
Gratis versus Libre

"Free" means there is no cost, where **libre** means "at liberty", referring to the freedom to modify source code. **Libre** doesn't mean **gratis**. **Libre** can mean available. **Libre** can mean without restriction

Open Science

- The movement to make scientific research, data and dissemination accessible to all levels of an inquiring society
- Scope:
 - 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

Open Science: an Umbrella of Many Activities



Open Science is Many Things

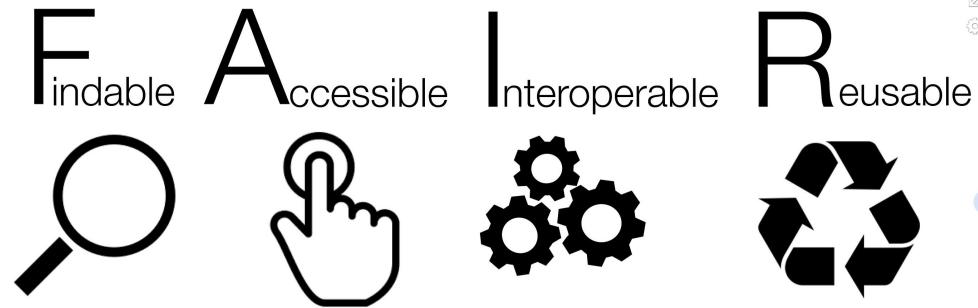
- Open Science includes activities that:
 - facilitate resource sharing
 - improve awareness of sharing
 - create linkages between resources
 - advocate for removal of financial barriers
 - advocate for just distribution of resources



Openness as an Aspiration not an Absolute

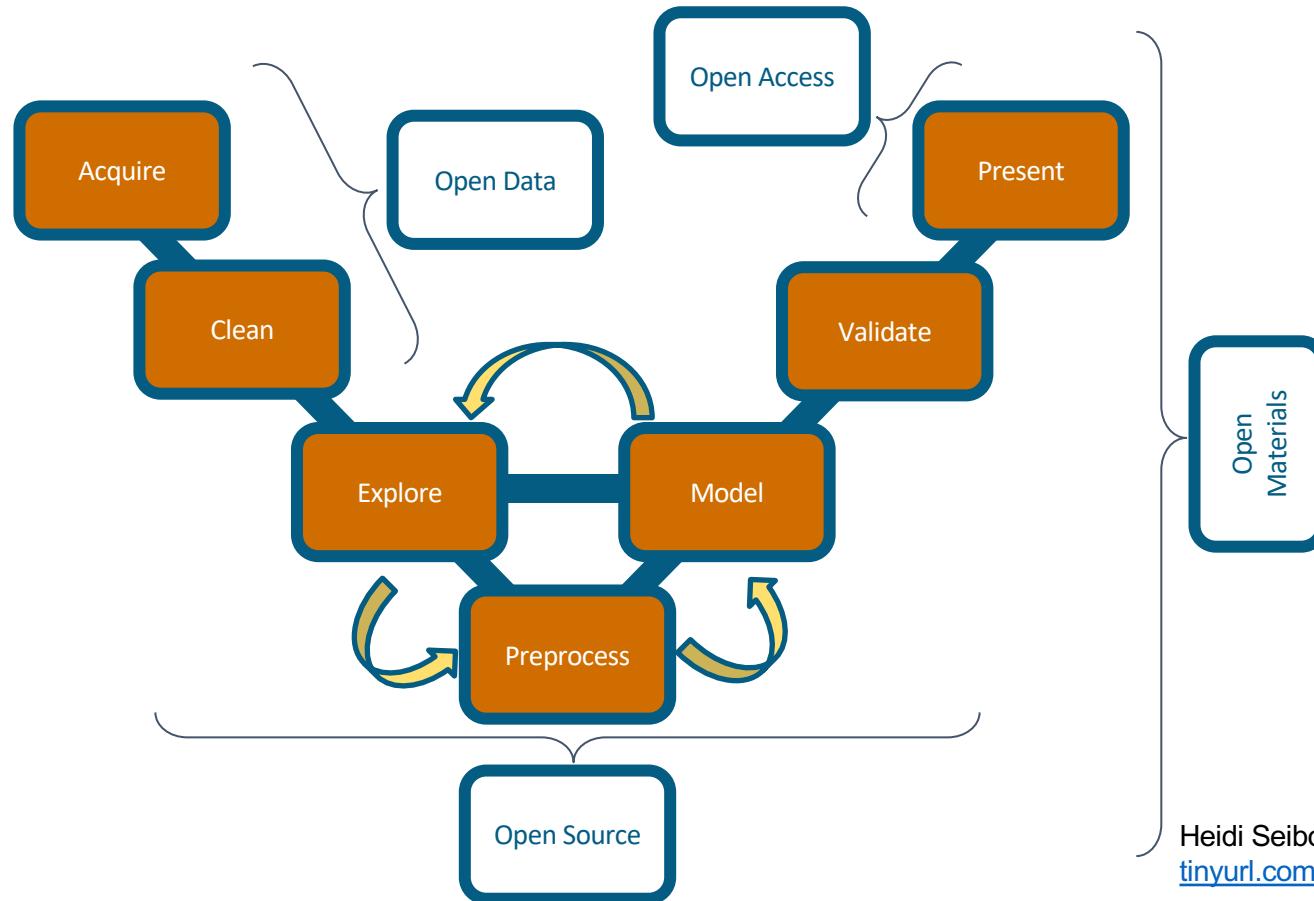
- Openness is not an absolute for research outputs – different research products require different solutions
 - Sensitive data
 - Commercial applications
 - National IP and traditional knowledge
- Open Science movement drives critical engagement in these areas to find solutions that are suitable
- Openness is not an absolute in the research process – different forms of openness foregrounded at different times

As Open As Possible, As Closed As Necessary

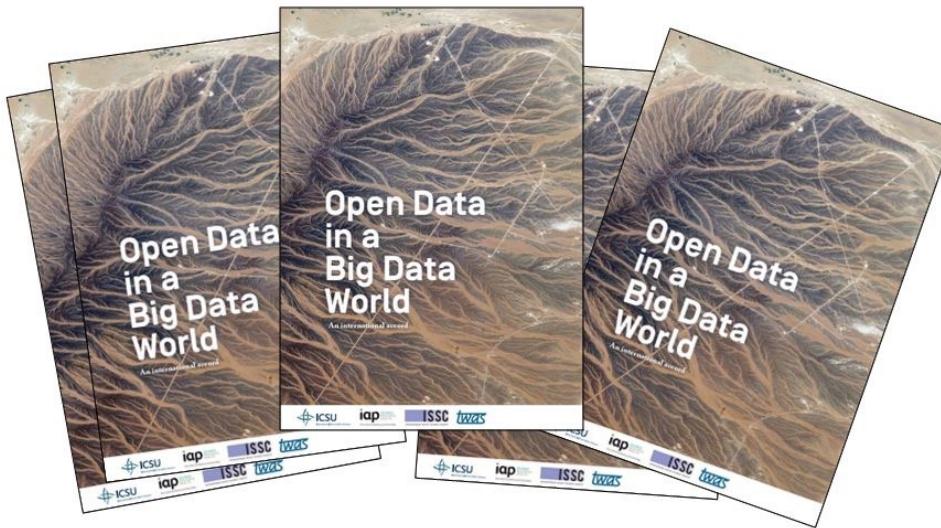


<https://www.timeshighereducation.com/blog/data-should-be-open-possible-and-closed-necessary>

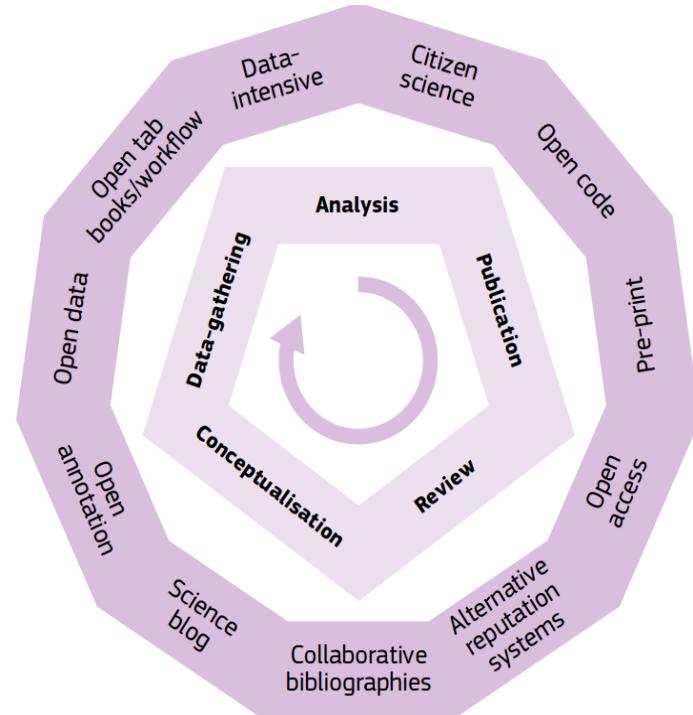
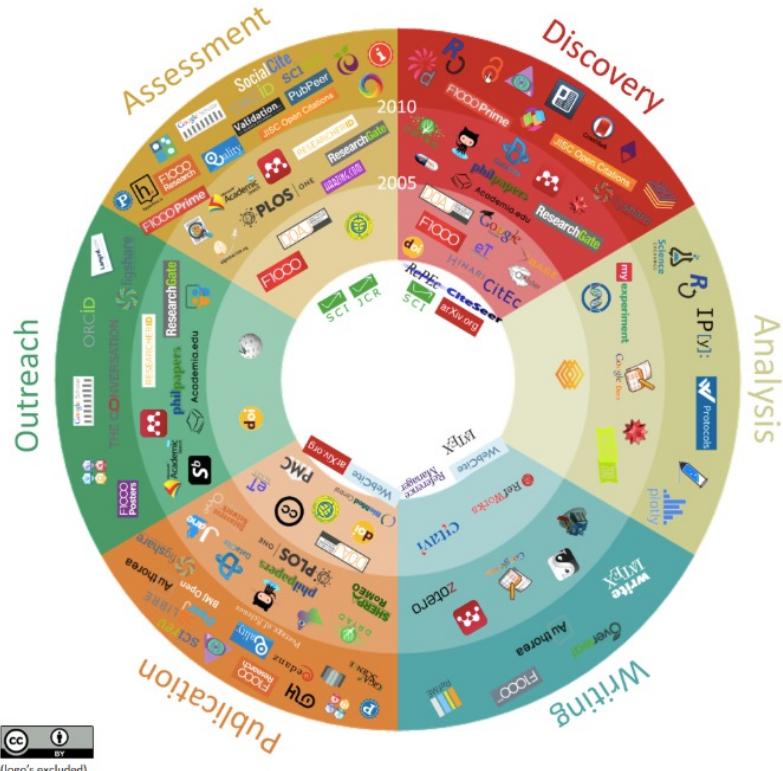
Openness Throughout The Research Lifecycle



Open Science: Changing How We Think About Research

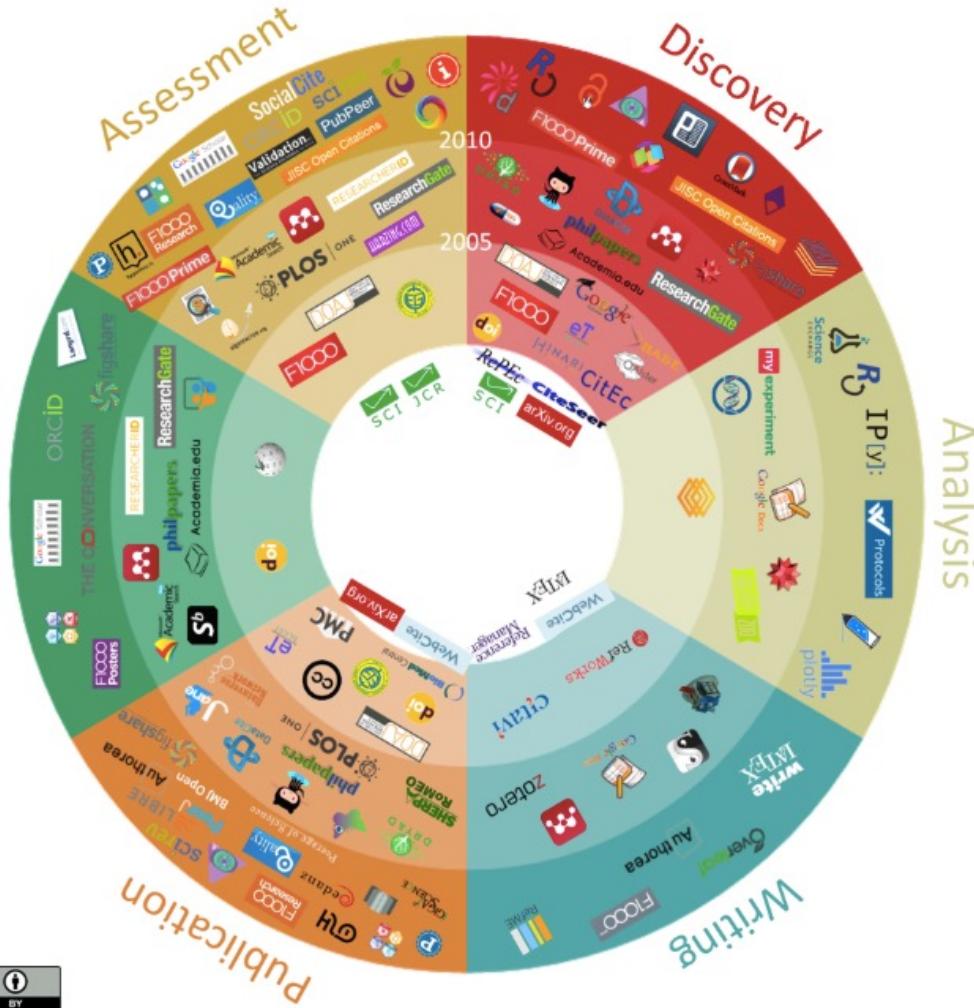


Open Science: Opening Up Research Potential



www.openphub.eu/component/k2/item/610-101-innovations-in-scholarly-communications

Outreach



Did you bring
your workflows?

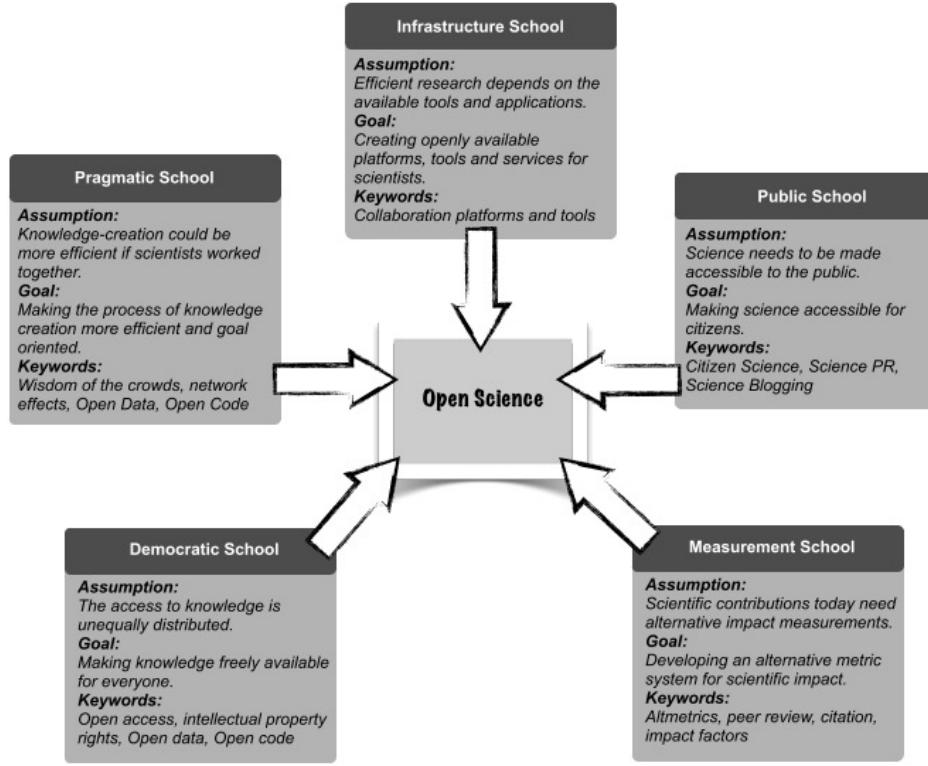
- If yes, try to think in pairs how to make it more open
- If not, then try to describe it first, and then think in pairs how to make it more open.
- 5 minutes
- If you feel comfortable, share it here:
<https://docs.google.com/document/d/1gfr5qUfOgcB6KUfYT6qFpNxHmrOgPADTWPWsyewZAw/edit?usp=sharing>

A Way of Thinking And Doing

- An ideology as well as a set of practical actions
 - Commitment to equity, excellence, responsibility and justice
 - Practice of research practices such as collaboration, sharing of research data, lab notes and other research processes with no restriction, promote reuse, redistribution and reproduction of the research and its underlying data and methods
- Open Science = practice-based ethics
- Like all ethics, it requires motivation, buy-in and commitment
- Changing the way we think about responsible research



Different Motivations, Same Response



(Fecher and Friesike, 2014)

Open Science: an Extension of RCR Values

- Openness can be thought of as an extension of RCR
- Allows researchers to uphold *integrity* and core ethical values underpinning RCR
- Allows researchers to engage in practical activities relating to RCR



Openness as an Extension of RCR

- 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

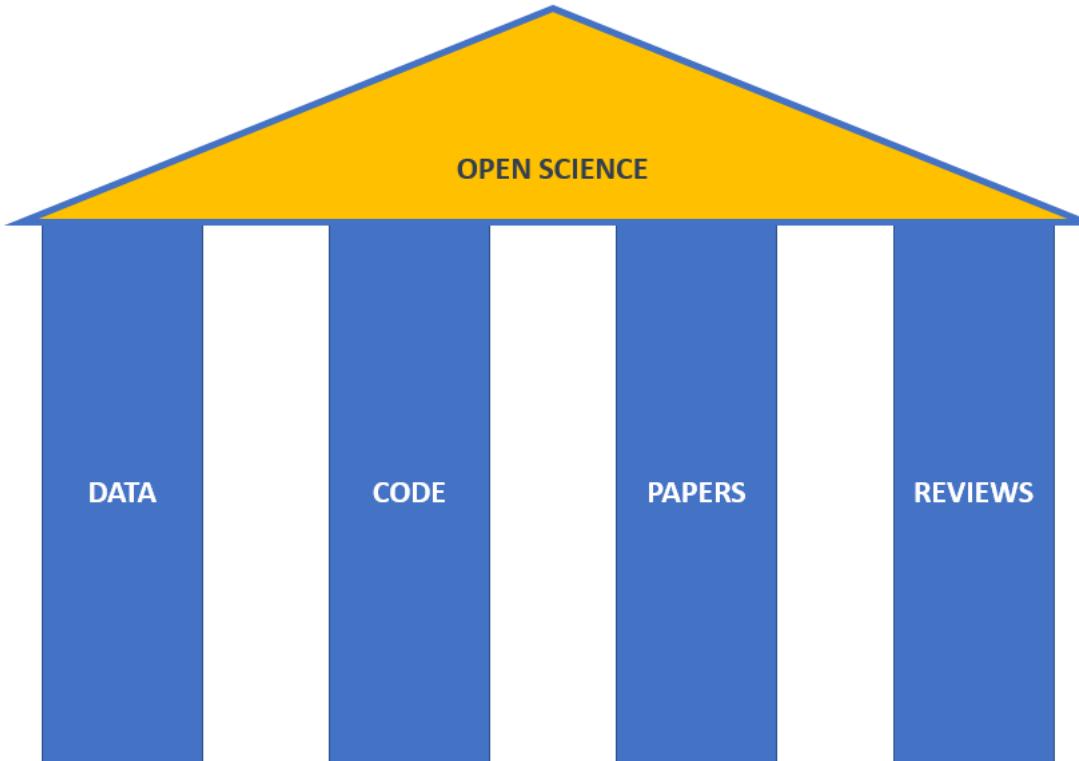
Openness in RCR



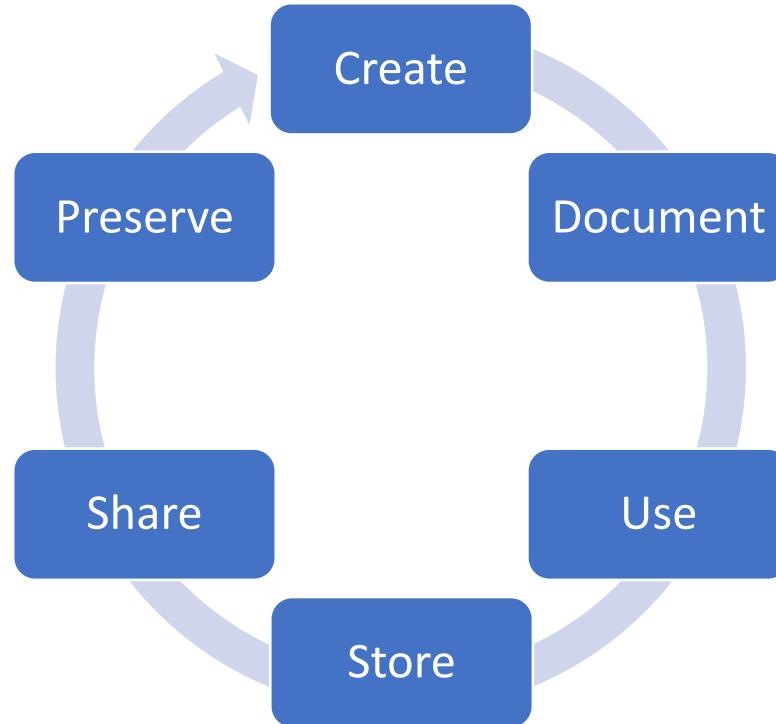
Openness as a Continuum and a Journey

- Open Science is both a cultural movement as well as a set of individual research decisions/practices
 - Commitment to openness is the start of a lifelong journey as an Open researcher – like anything, learning to be open takes time, motivation and mentorship/support
 - Individual researchers can show support for Open Science through the way they do their research
 - Can make changes to daily research practices to be more open in all aspects of research
-
- *There is no “hard and fast” rule – openness must work for you within your specific work context and with your daily challenges*

“My” Open Science



Different Actions at Different Stages



Individual Openness: More Than Open Access

Share your data - the research data that underpins publications should also be accessible to support validation and facilitate reuse. In cases where data sensitivities won't allow open access, be sure to provide details on how someone could request authorised access.

Share your code - many researchers now develop bespoke bits of code to help them analyse and/or visualise the data they have collected. Having access to this code is essential for supporting the validation of your findings and to help others to build upon your work.

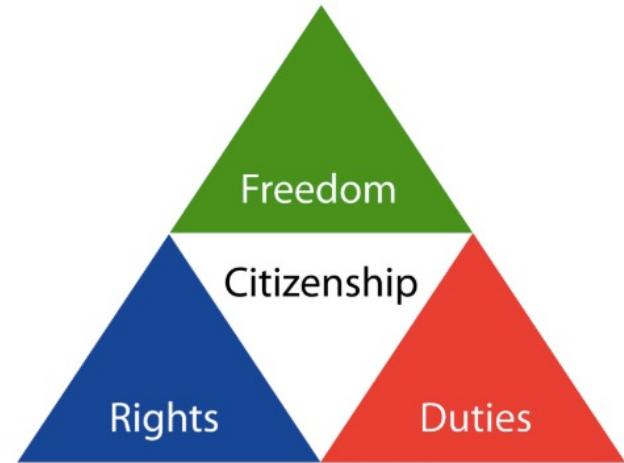
Share your workflows - without knowing what steps were taken to capture, process and analyse the data - and in what order - it can be virtually impossible to validate published findings. This has led to what some are calling the Reproducibility Crisis. Nature's special issue on Challenges in Irreproducible Research gives you a better sense of the scale of this problem.

Pragmatic Openness

- Openness in research is new ... And can be scary
 - It's ok to have concerns
 - It's ok to realize that not all the open practices will work for you
 - There can be legal, regulatory, ethical reasons not to share
- However, you benefit from the increasingly open research landscape (papers, data, software)
- So, how can one avoid "freeloading"?

Responsible and Open Research as Citizenship

- Citizenship:
 - ethical obligations arising out of social living
 - being part of a community requires the acceptance of civic responsibilities and contribution to the overall public good
- As a citizen you have duties and expected ways of acting
 - Follow rules
 - Participate in community activities
 - Protect the community and its resources from misuse



Responsible and Open (Data) Science Citizenship

- Research is a community endeavour
 - involves social actions such as resource sharing and communal practice
 - responsible researchers are “citizens” of the research community
- Citizenship is a give and take
 - Benefits to facilitate freedom of research
 - Structures to safeguard rights as researcher
 - Responsibilities to assume to protect culture
- Support and grow culture instead of just living in it

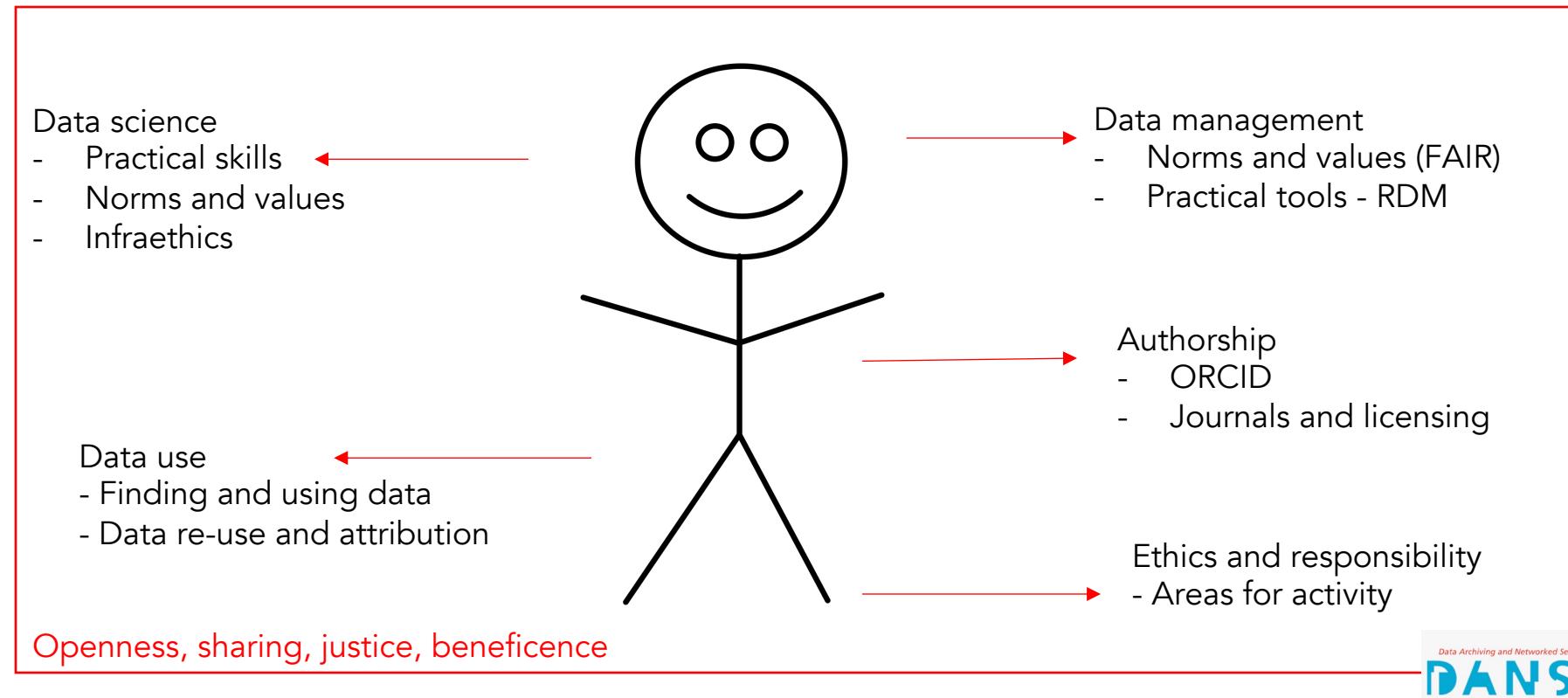
Responsible and Open (Data) Science Citizenship

- RCR and Open Science form the blueprint for a form of “science citizenship”
- Research relies on the use of “community resources”
 - data, papers and so forth
- As a “citizen” of the research community you therefore have responsibilities for these resources
 - follow community determined rules (such as citation, licensing and so forth)
 - Contribute to communal resources (data sharing)
 - Maximise good for the community by participating in civic service (reviewing, curating etc)

Responsible Research is a Life-Long Journey

- Responsible conduct in research will be a life-long journey.
Remember that everyone has to start somewhere
- Start small, with activities that you are comfortable with
- Ethical conduct is part of daily life and small tasks. It's not just about the big picture
- Ethical conduct benefits everyone. It will make you a better researcher and your research better. This is good for everyone!

Tools for Responsible, Open (Data) Science Citizenship





Data
Schools

Open and Responsible (Data) Science Citizenship 2

Material from Louise Bezuidenhout,
presented by Marcela Alfaro Córdoba

Plan for the second part of the morning

- Personal concerns and implementing OS practices at home
 - What have we learned during these weeks?
 - 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 marginalization

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. Take 10 minutes to write some ideas down on this shared doc:

<https://docs.google.com/document/d/1f4HzVPsjuEQcR5N88SMDkaNjURjOsakbkR4GrzV0Z8/edit?usp=sharing>

CHALLENGES OF OPEN+RESPONSIBLE RESEARCH AT HOME		
INSTITUTIONAL / CULTURAL	CHALLENGE	SOLUTIONS
	<ul style="list-style-type: none">• lack of support from supervisor / peers• peers not sharing• ownership / IP concerns• tradition of not being open• lack of interest• being naive about risks• traditions of non-transparency• lack of knowledge• no oversight on RDM	<ul style="list-style-type: none">• highlight that openness is not new• draw attention to funding requirements• openness = citations• use international policy• institutions (ie. COPE)• make (more) examples - ie. work flows• engage multiple stakeholders (incl. uni. admin, grad students, undergrads)• talk about FAIR + CARE• ENGAGE WITH CONFERENCES
INFRASTRUCTURE / RESOURCES	<ul style="list-style-type: none">• time• infrastructure• availability of data• repositories - access / lack• lack of training & opportunities• lack of RDM awareness• knowing the laws• processing power	<ul style="list-style-type: none">• use OS tools online - ie/101 tools• make use of available databases• check redata• ask for APC waivers• check out lists of resources• get involved in training - ie. Carpentries• join communities - ie. R/Rhythm• use support networks - ie. AuthorAID• get involved in OS discussions to reflect LMIC perspective
PERSONAL CONCERN(S)	<ul style="list-style-type: none">• not sure how to share• working with qual research• lack of knowledge of tools• not getting credit• working multidisciplinarily - different approaches to RDM	<ul style="list-style-type: none">• reach out to decision makers to raise concerns• know your rights - Creative Commons license• know community expected behavioural standards - ie. codes of conduct• "take it slowly" - increase openness incrementally in your own work• find people to model your work practices on• find trusted infrastructures (open, recommended basis)

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

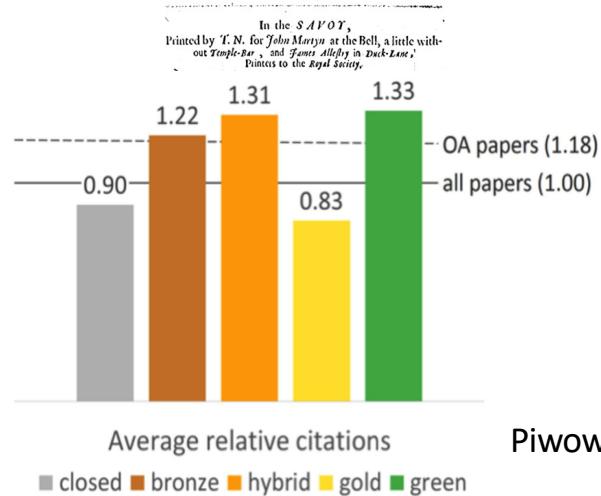
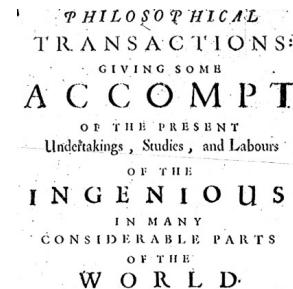
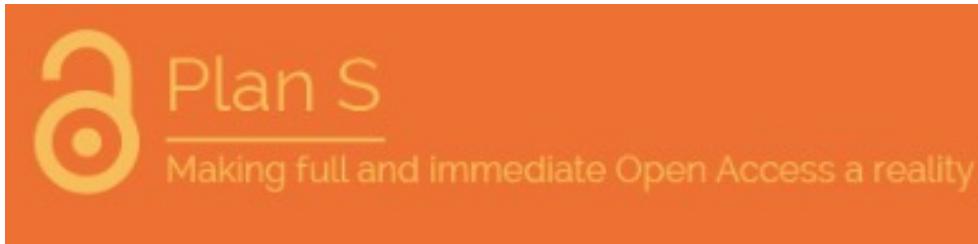
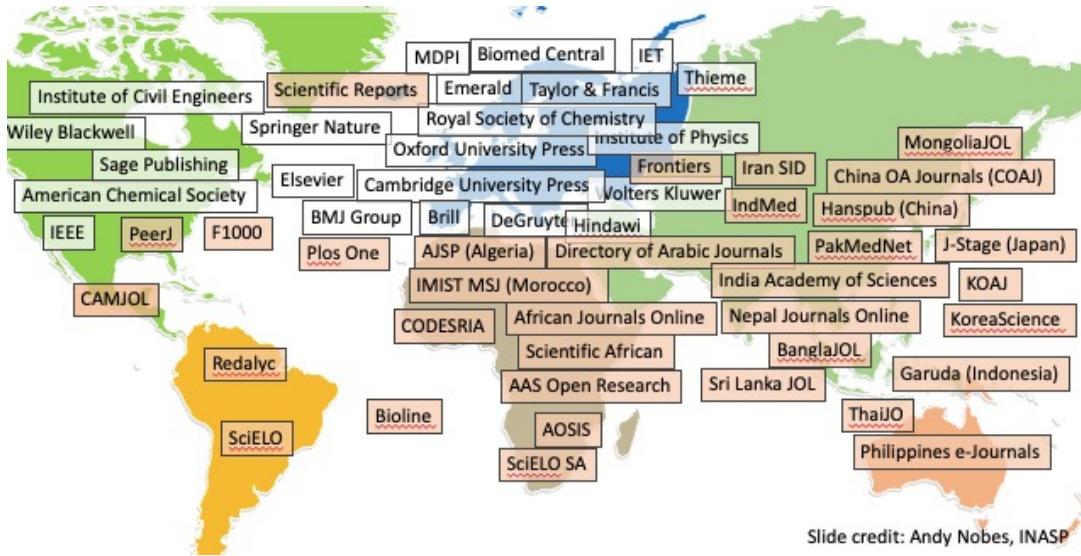


1. Cultural Resistance and Lack of Support

- 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



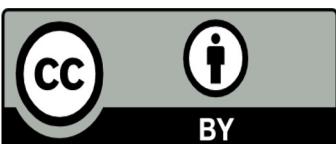
Piwowar 2018

1b. Getting Better Protection and Guidance



Data-sharing Agreements

! THINK ✓ CHECK > SUBMIT



1c. Getting Over Issues of Trust



Data-sharing Agreements

C O P E 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

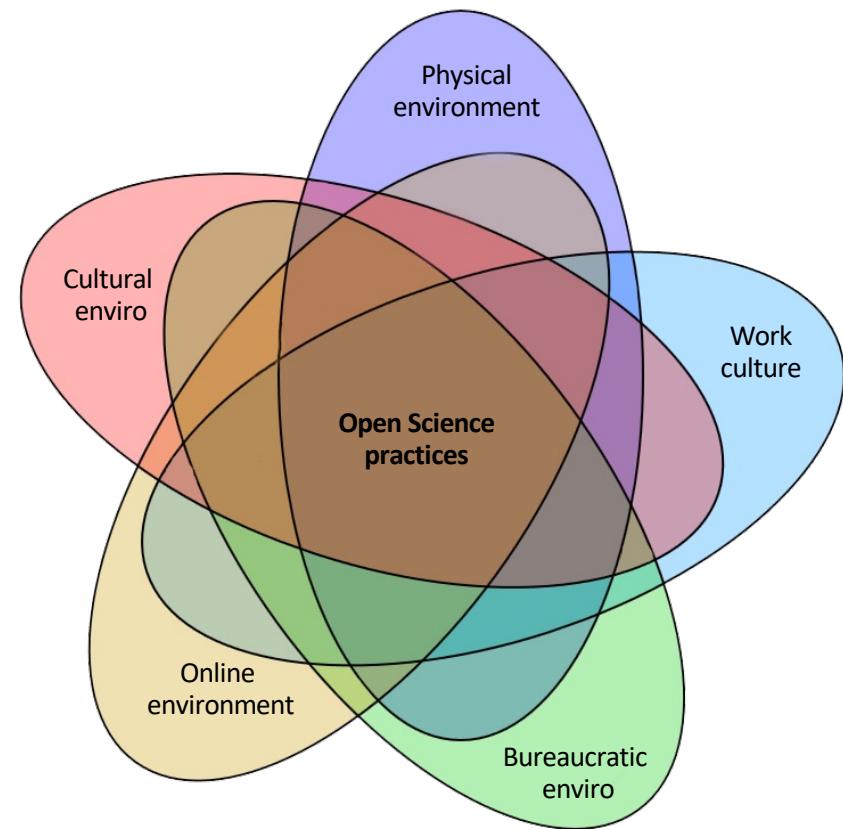
When poll is active, respond at PollEv.com/ucscstat5f22
Text UCSCSTAT5F22 to 37607 once to join

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

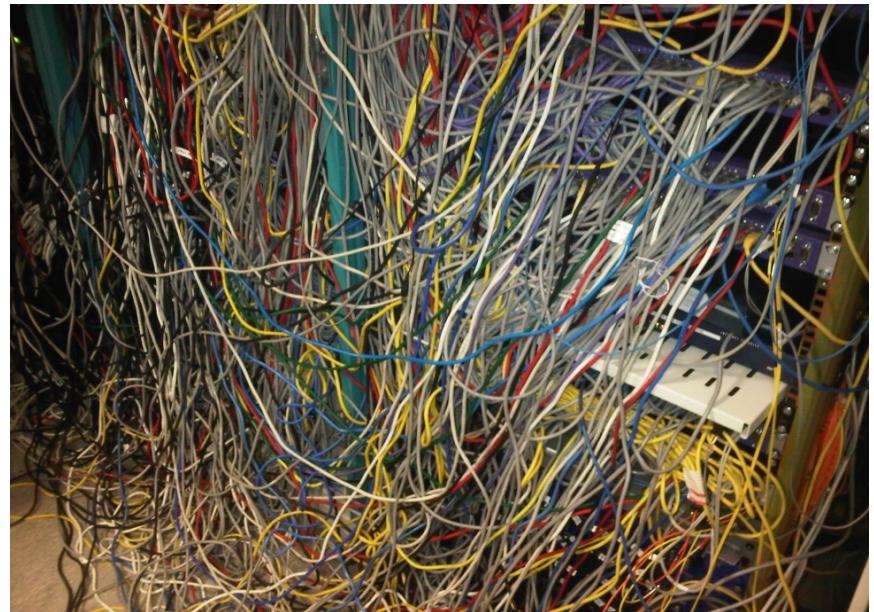


2. Infrastructures that Support Openness

- 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

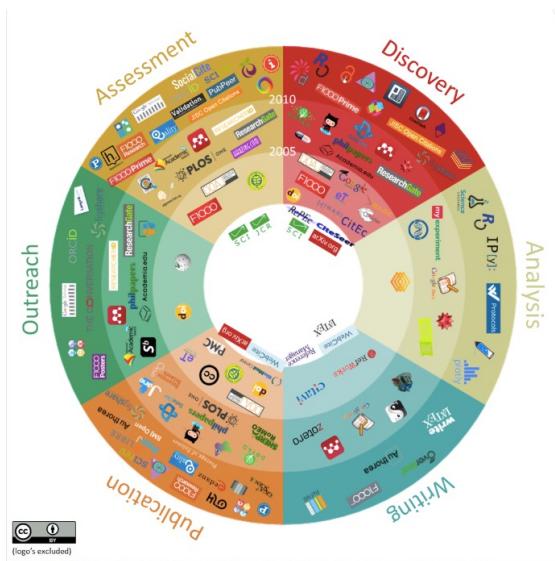


Just Because the Resources Are Online ...



.... doesn't mean they're accessible

2a. Lack of Resources



WILEY
HOME MY DASHBOARD AUTHORS - REVIEWERS - EDITORS - HELP - REGISTER LOGIN

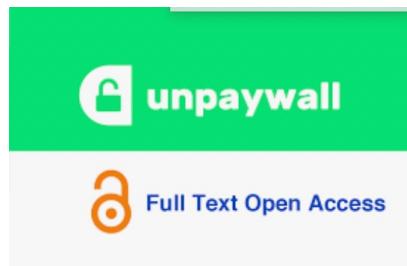
Open Science > Open Access > For Authors > Publication Charges > Waivers and Discounts

Open Access Open Data Open Practices Open Collaboration Open Recognition & Reward

Waivers and Discounts

For authors publishing in Wiley Open Access journals, Wiley offers waivers and discounts to authors based in developing countries. To ensure that editorial decisions are never influenced by ability to pay, it is Wiley policy that editors of open access journals are not involved in correspondence with authors regarding payment of Article Publication Charges (APCs). The automatic waiver system will be managed by administrative staff not involved in decisions regarding article acceptance. We ask authors not to discuss any issue concerning payment with editors.

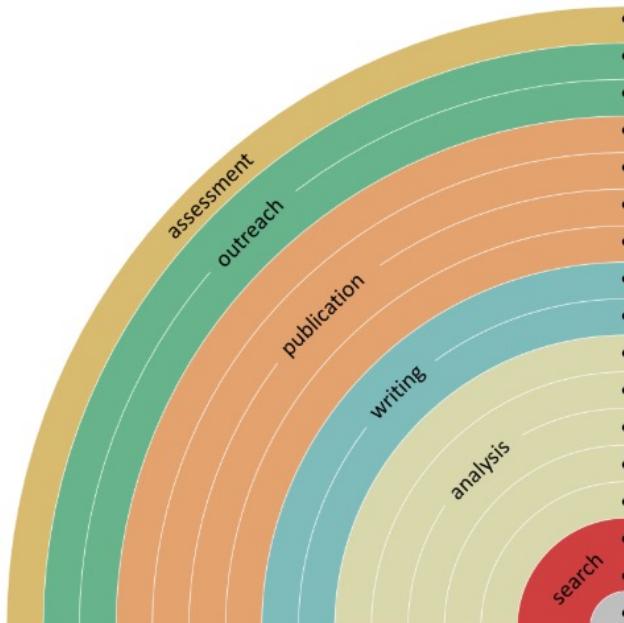
In addition to the Research4Life countries listed below, some journals may offer additional waiver initiatives. These additional waivers will be managed by the editors and may be discussed with them.



Welcome to Sherpa Romeo

Sherpa Romeo is an online resource that aggregates and analyses publisher open access policies from around the world and provides summaries of publisher copyright and open access archiving policies on a journal-by-journal basis.

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



Knowing Where to Look

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.iab.me/home/>
- See also the [Wikipedia list of academic databases and search engines](#)

2b. Lack of Expertise and Training



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/>

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

When poll is active, respond at PollEv.com/ucscstat5f22

Text UCSCSTAT5F22 to 37607 once to join

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

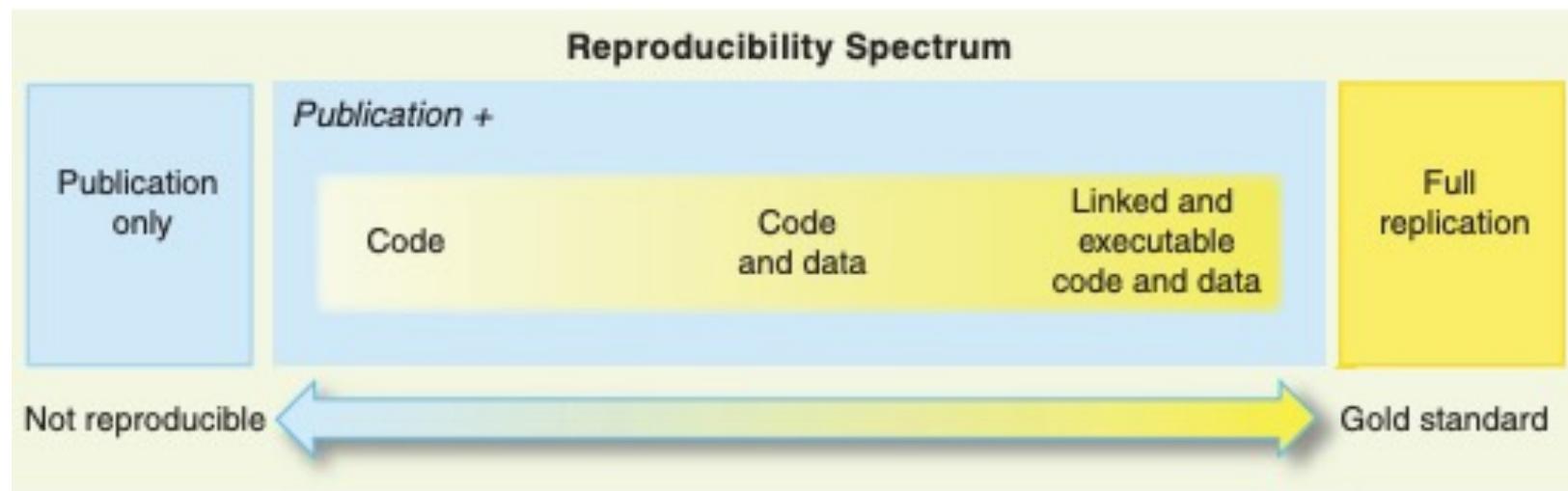


Free as in Freedom

ACM Code of Ethics and Professional Conduct

3b. Openness as a Continuum

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



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



Still Needed: More Evidence

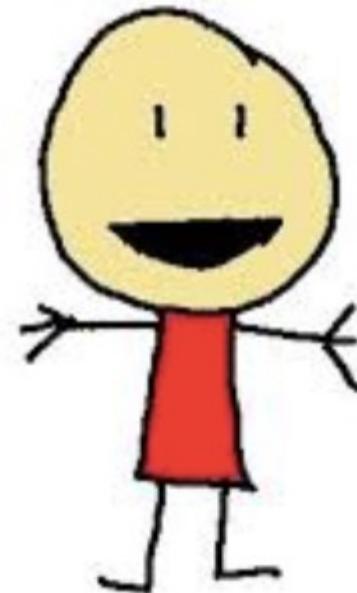
- Lack of evidence of LMIC concerns
- Tendency to treat LMICs concerns as “same as HIC but more”
- Need more evidence about what is working, what is preferred and what is still needed
- Creating, joining and interlinking networks of support is key to fostering Open Science

✉ When poll is active, respond at PollEv.com/ucscstat5f22

SMS Text **UCSCSTAT5F22** to **37607** once to join

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

Internet Hug



Please wrap arms around monitor now

Openness Is A Lifelong Journey

-  Publish Preprints
-  FAIRify data
-  Make code available
-  Publish Lab-Notebooks
-  Use version control
-  Preregister your project
-  Do science communication

Let's go back to our board and see what resources we can fill in:

<https://docs.google.com/document/d/1If4HzVPsjuEQcR5N88SMDkaNjURjOsakbkR4GrzV0Z8/edit?usp=sharing>

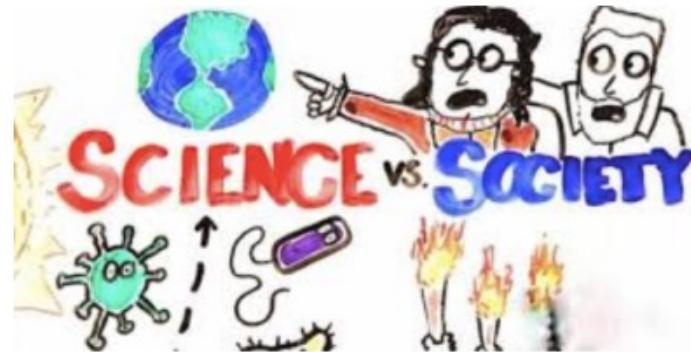
Learning to Look at
the Bigger Picture

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 Research Context

- 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

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

Discrimination

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

Value Laden 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)

Current Challenges

did the holocaust happen

did the holocaust happen

did the holocaust happen during ww2

did the holocaust really happen yahoo

did the holy grail exist

Top 10 reasons why the holocaust didn't happen. - Stormfront
<https://www.stormfront.org/> • General > History & Revisionism •
19 Dec 2008 • 10 posts • 8 authors
The Holocaust Lie more than anything else keeps us down. The twin ... You can believe what you want, but I believe the holocaust did happen.

Holocaust denial - Wikipedia
https://en.wikipedia.org/wiki/Holocaust_denial •
Holocaust denial is the act of denying the genocide of Jews and other groups in the Holocaust ...
denial movement bases its approach on the predetermined idea that the Holocaust, as understood by mainstream historiography, did not occur.



Mascom 11:01 PM 17% ↗ unprofessional hairstyles for work

D | E.T.P.
hoax-it-never-happened! ▾
require force of law to ... are "undeniable"

Follow

Bonnie Kamona
@BonKamona

I saw a tweet saying "Google unprofessional hairstyles for work". I did. Then I checked the 'professional' ones 😊😊😊

10:04 PM - Apr 5, 2016

234 12,620 8,039

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



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



Example 1: Algorithmic Decisions on Bail

The screenshot shows the ProPublica website with the title "Machine Bias" in large white letters. Below the title is a subtitle: "There's software used across the country to predict future criminals. And it's biased against blacks." The text is attributed to "by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica May 23, 2016". The ProPublica logo is at the top left, and social media sharing icons are at the top right.

Northpointe and COMPAS

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

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 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

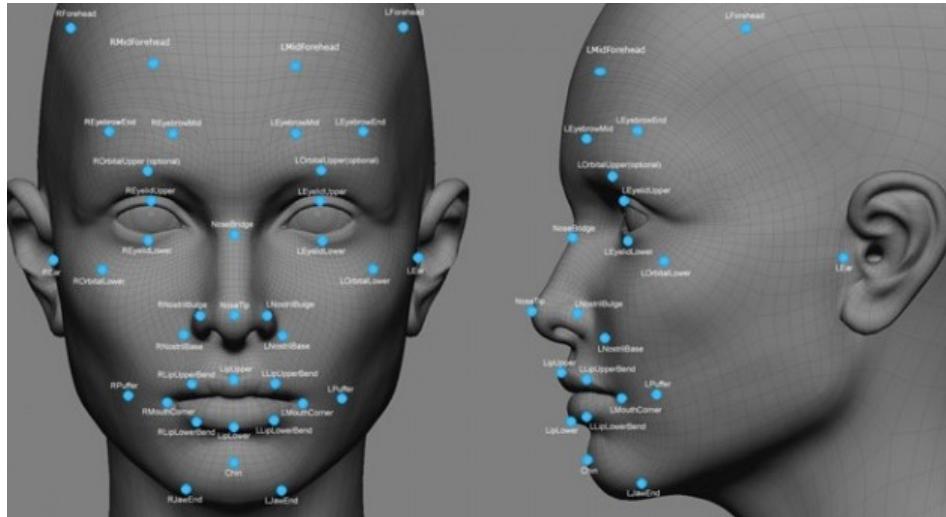
Campaigning for Justice: ProPublica

- 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



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

The screenshot shows the Algorithmic Justice League website. At the top, there is a navigation bar with links to Home, Safe Face Pledge, Gender Shades, and The Coded Gaze. To the right of these links is a logo featuring a shield with the letters "AJL". Further to the right are links for "AI, Ain't I A Woman", "Fight Bias", "Newsletter", and a "GET INVOLVED" button. Below the navigation bar, the text "ALGORITHMIC JUSTICE LEAGUE" is prominently displayed in large, bold, white capital letters. To the left of this text is a photograph showing a person's hand holding a smartphone displaying a facial recognition application. A green hexagonal icon with a camera lens and a checkmark is overlaid on the phone screen. To the right of the phone is a white rectangular box containing the text "SAFE FACE PLEDGE LAUNCHED" in bold capital letters, followed by the subtitle "In partnership with the Center On Privacy & Technology". Below this, a paragraph explains the purpose of the Safe Face Pledge, mentioning its goals of mitigating facial analysis technology abuse and prohibiting its lethal and lawless use. It also encourages transparency and human rights. At the bottom of this box is a red link to the website: www.safefacepledge.org.



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 that 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

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs



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...

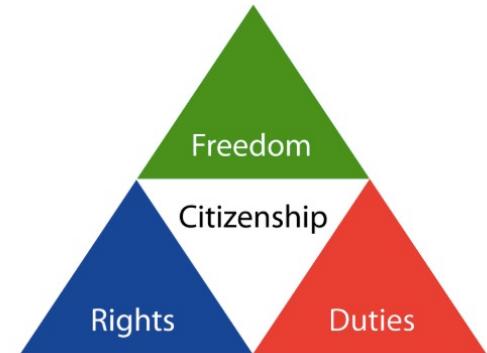
What Can I Do?

- 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



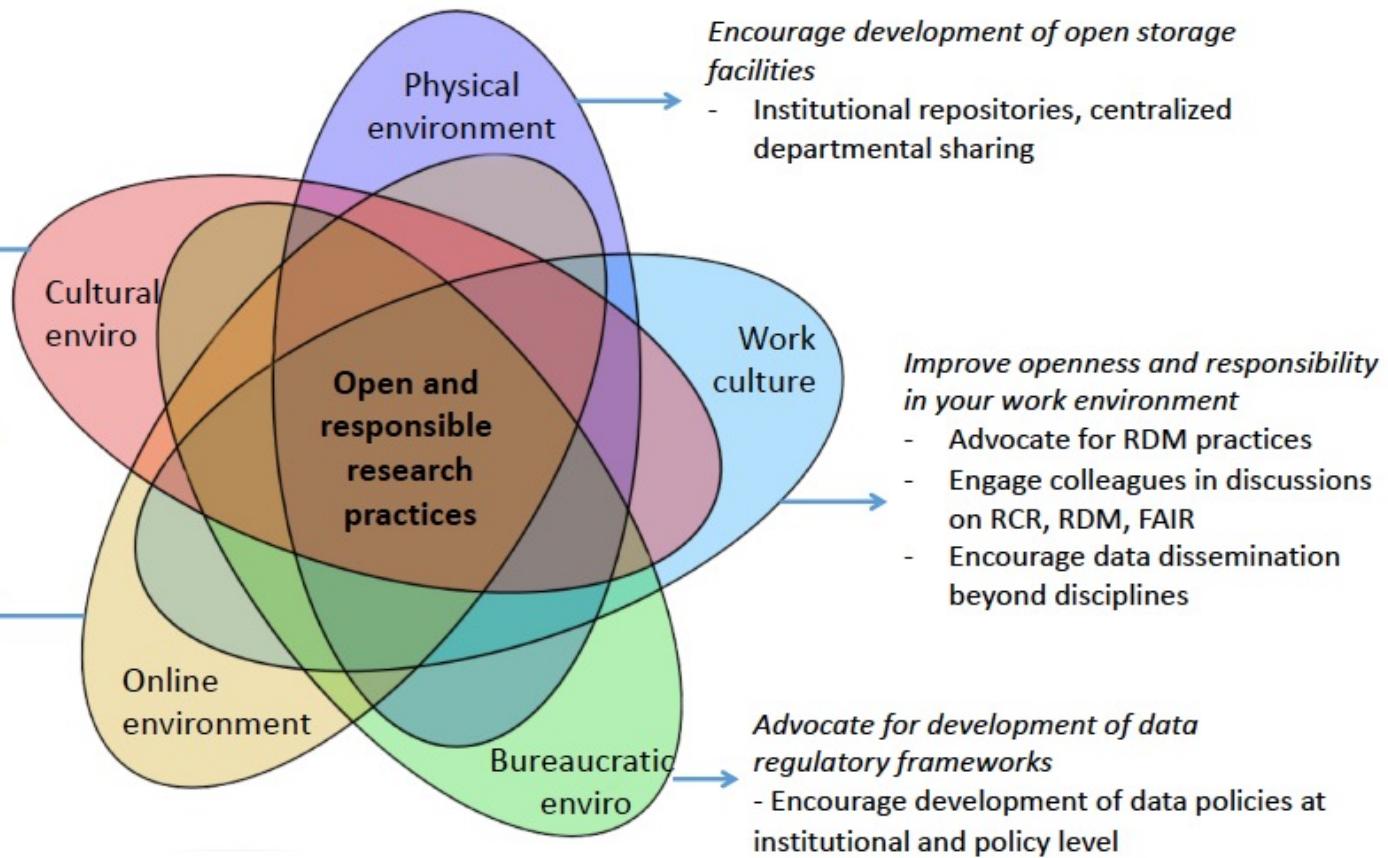
Extending Data Science Citizenship Responsibilities

Engage with public on data-related issues

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

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



Outline for Next 4 Weeks

Over the course of the next weeks, reflect on the rest of 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@dans.knaw.nl

