RESPONSIBLE, OPEN SCIENCE CITIZENSHIP

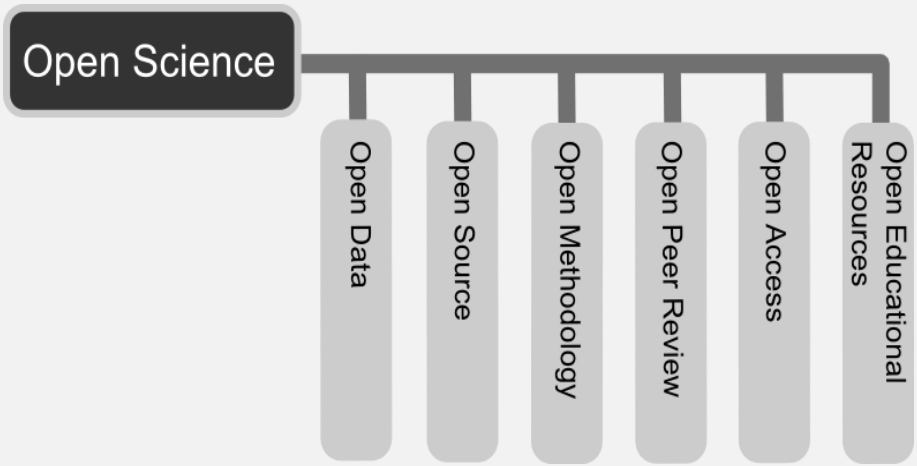
CODATA/RDA Research Data Science Summer School

Thursday/Friday, Kigali, October 2018

Louise Bezuidenhout

University of Oxford

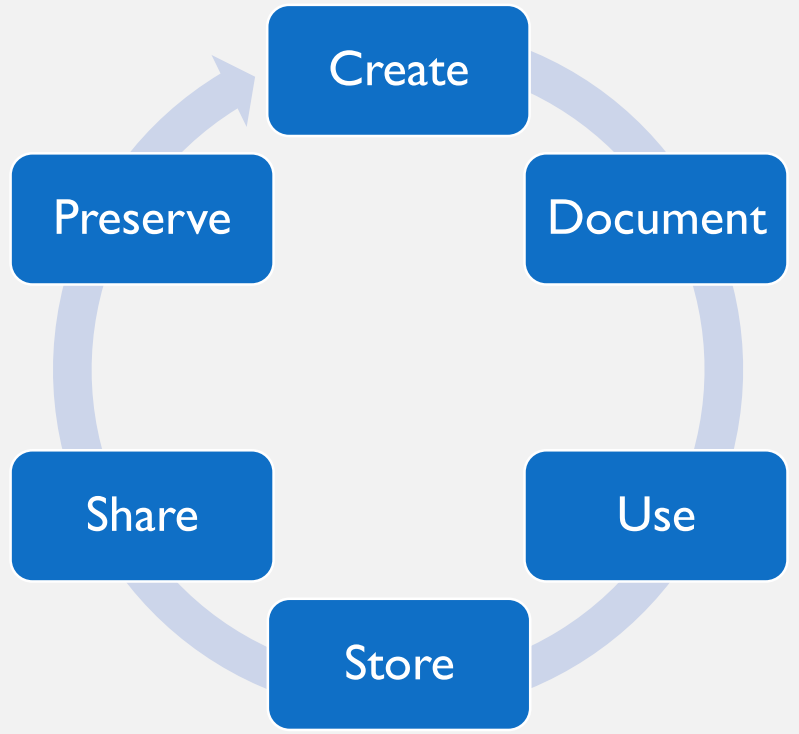
How to be a responsible and open data scientist



Tools for responsible and open data science

DMP ONLINE

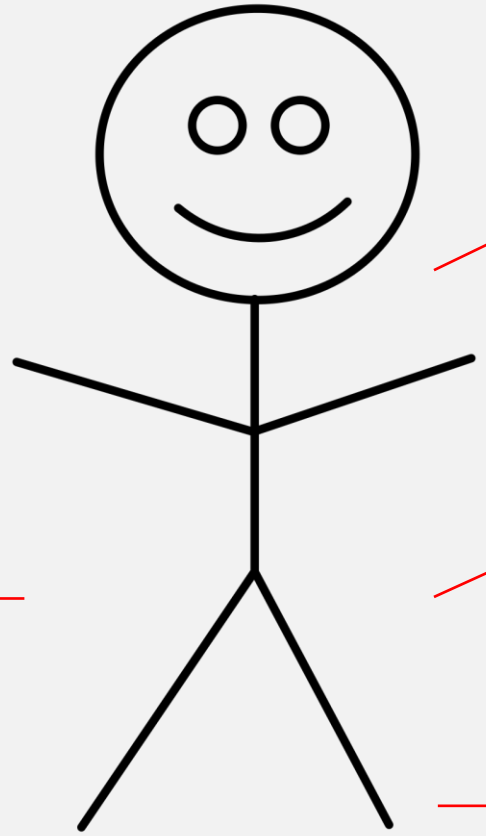
Doing responsible and open data science research



RESPONSIBLE, OPEN (DATA) SCIENCE CITIZENSHIP

Data science
- Practical skills

Data use
- Finding and using data
- Data re-use and attribution



Data management
- Norms and values (FAIR)
- Practical tools - RDM

Authorship
- ORCID
- Journals and licensing

Ethics and responsibility
- Multiple responsibilities
- Norms and values

ETHICAL CHALLENGES OF BEING A RESPONSIBLE, OPEN DATA SCIENTIST

- Data producer
- Data user and/or collaborator
- Author
- Employee
- Teacher/mentor
- Recipient of public funds
- Citizen/legally-obligated individual

Openness,

sharing,

justice,

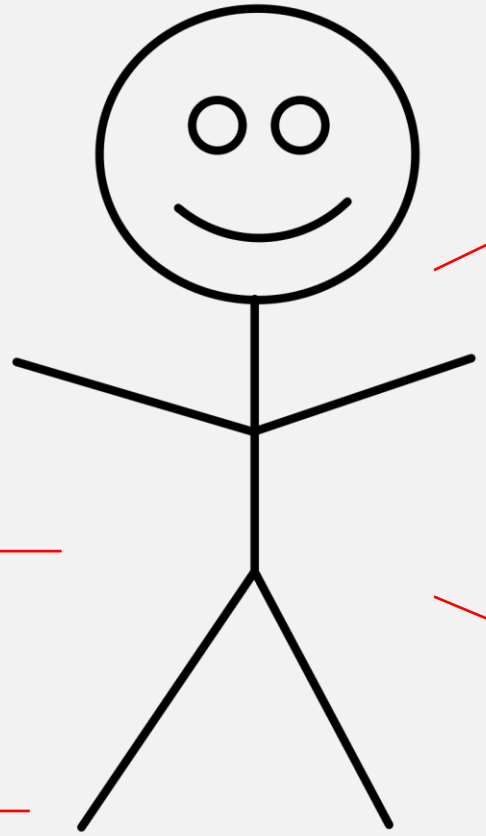
beneficence

ETHICAL CHALLENGES OF RESPONSIBLE, OPEN (DATA) SCIENCE CITIZENSHIP

Data science
- Biases in analysis
- FOSS vs commercial

Data use
- Inappropriate data use,
- Attribution, theft etc
- Commercialization issues

Ethics and responsibility
- Harms to society
- Present and future



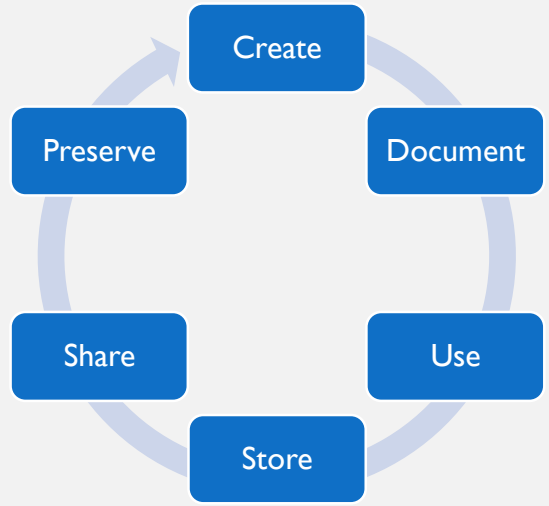
Data management
- Non-transparency
- Irreproducibility
- Waste of trust/resources
- OD

Authorship
- Predatory publishing
- OA
- Inappropriate authorship
- Theft

BEING A RESPONSIBLE, OPEN (DATA) SCIENTIST AT HOME

- Institutional cultures
  - Promotion criteria
  - Incentivization schemes
  - Cultural specificities
- Institutional support
  - Facilities
  - Traditions

- Resources
  - Time
  - Money
  - Infrastructure
- Copyright, ownership and agency
  - IP
- Career pressures
  - Time
  - Being scooped



PERSONAL CONCERNS

- Lack of awareness and training
- Cultural inertia and misinformation
- Challenging the establishment
- Following status quo
- Lack of reward
- Publication bias towards novel findings
- Resources
- Fear – of being scooped, scrutinized, reduced scientific quality, risk to reputation

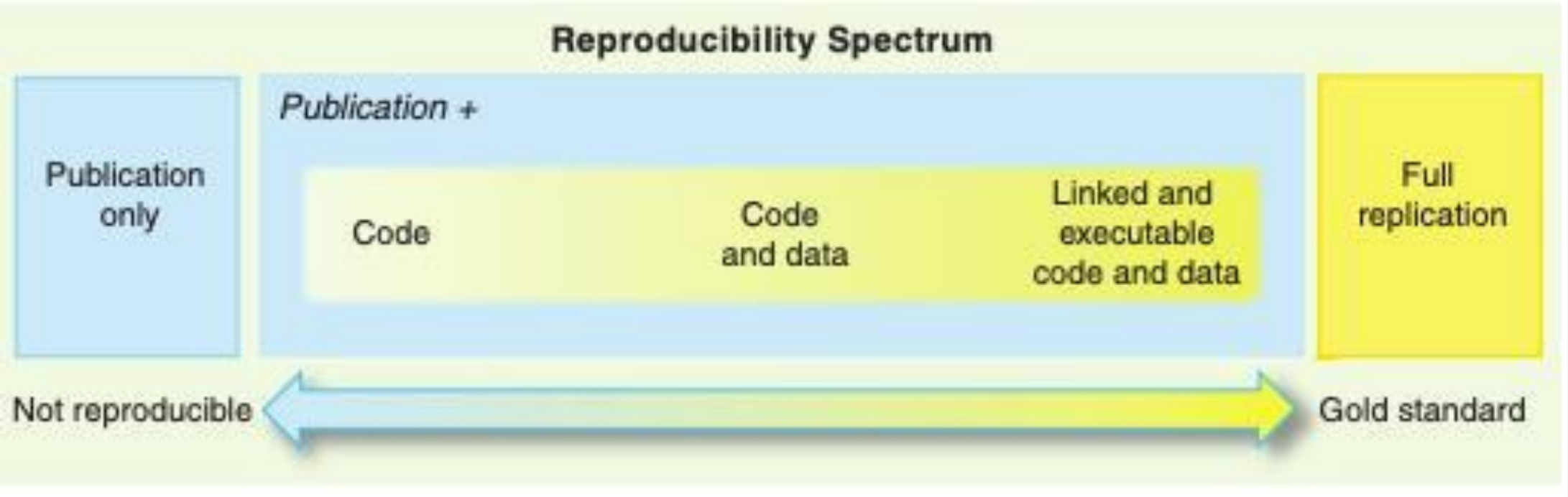
EXPERIENCING CHALLENGES IS NORMAL

Group discussion

1. What specific challenges do you anticipate encountering when you return home in terms of your data work?
    1. Using your hand-out sheets, go through the categories of RDM and RCR and think about challenges you will experience
    2. Discuss these specific, or general challenges in your groups
    3. Will this affect your ability to be responsible and open?

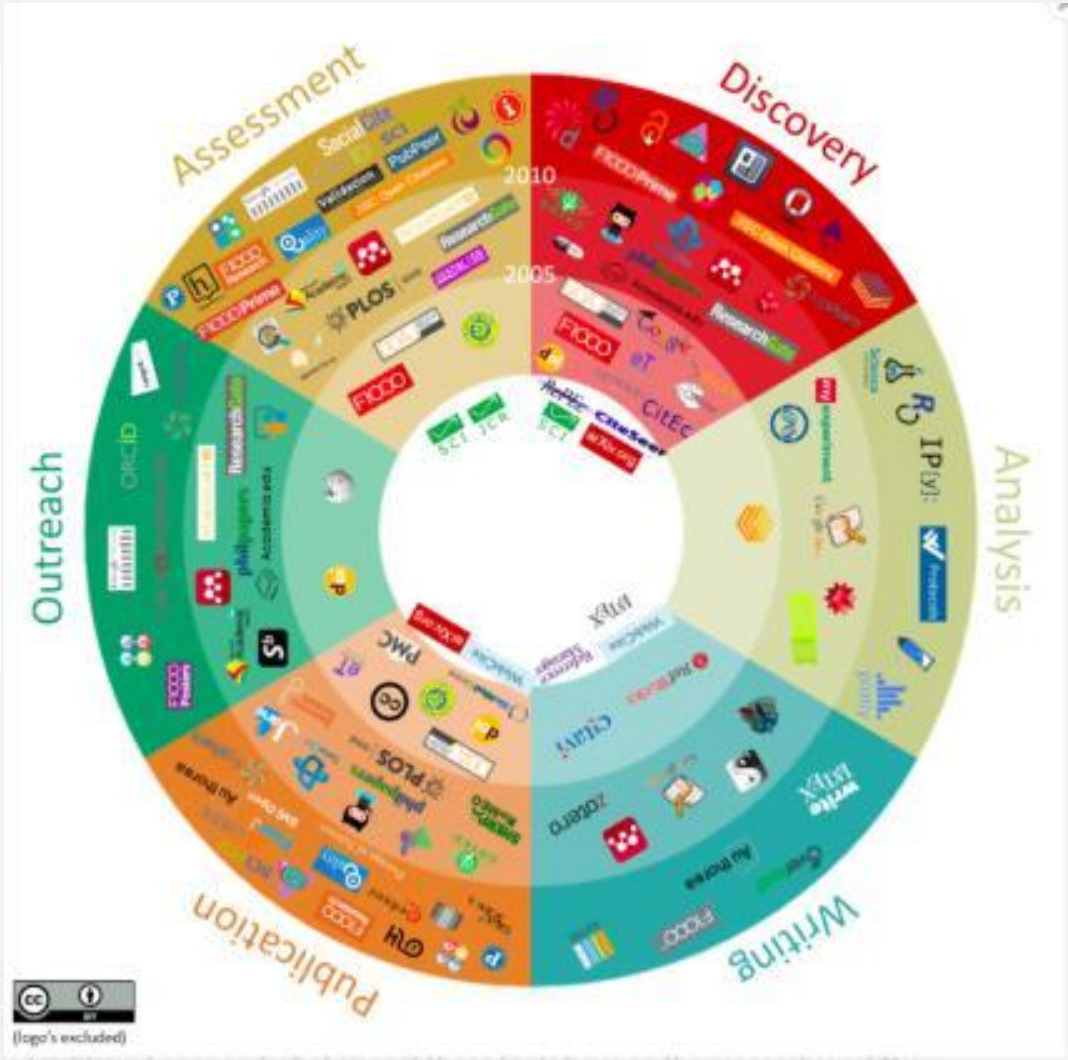
Working openly, responsibly and reproducibly ...

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



Peng 2011

RESOURCES TO CONSIDER



https://figshare.com/articles/101_Innovations_in_Scholarly_Communication_the_Changing_Research_Workflow/1286826

SOME USEFUL RESOURCES TO CONSIDER

A global network of researchers

AUTHORAID

Español

Search for... SEARCH

Sign in or Register

HOME ABOUT NEWS MENTORING COURSES FUNDING RESOURCES FORUM FAQ

AuthorAID is a free pioneering global network that provides support for researchers in low and middle income countries

Our online platform offers free online mentoring and collaboration, online courses and resources. There are currently over 17,000 members in our community from 175 countries, so why not join now?
Registering is free and easy

NEWS

Tip of the Week #401

The right tone

By *Bernard Appiah*, Aug. 4, 2018 | 0 Comments

Guest post: How can I choose the right conference to attend and present research?

Mohamed Mostafa introduces the Think. Check. Attend. Initiative and

RESOURCES

Presentation:Preparing Grant Proposals - Facilitation Kit

This kit consists of nine modules, each containing a PowerPoint presentation and a set of facilitator notes. The notes expand on the content of the slides and provide guidance in facilitating the modules, which combine presentation and small-group work.

Book:Research Ethics Commit-

FEATURED RESEARCHERS

Tonderayi Matsungo

Research interests: stunting, micronutrients, hidden hunger, biofortification, randomised controlled trials

Open to collaboration.

Godfrey Zari Rukundo

Research interests: suicidology,psychiatry,adolescents,

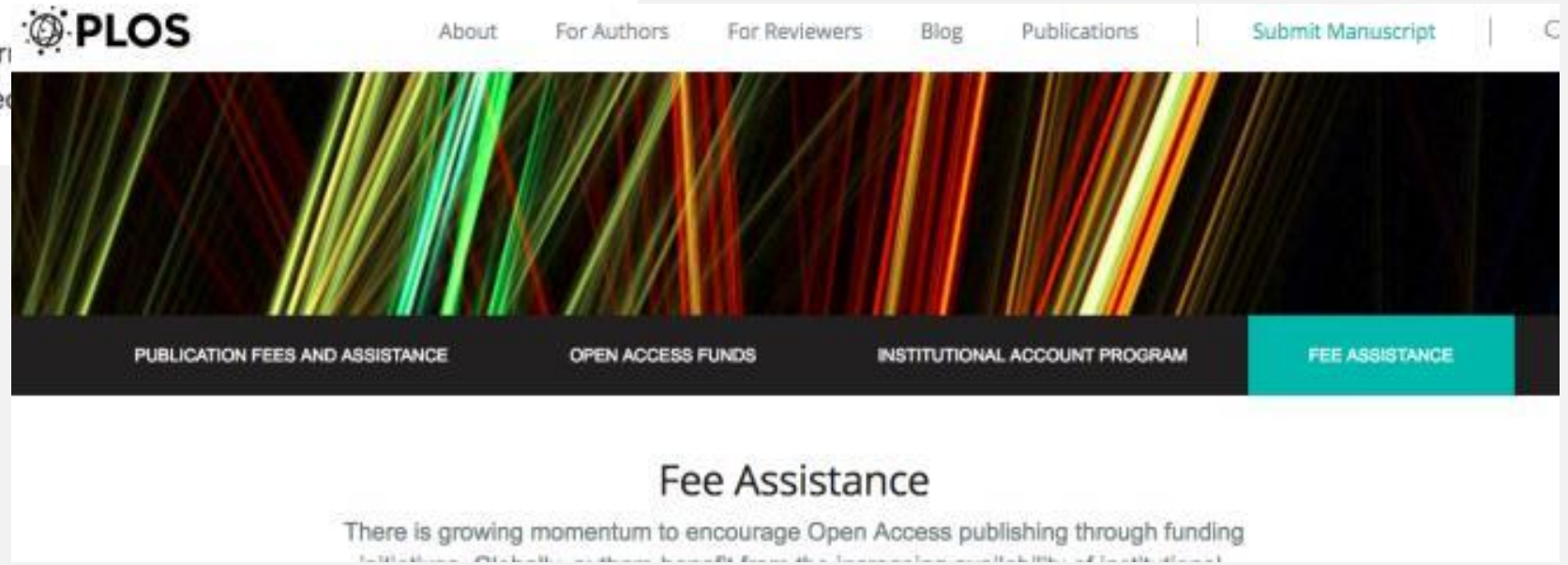
Looking for a mentor

APC WAIVERS

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 cor
classified by the World Bank as low-income e



ACCESS TO RESOURCES

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

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

SUPPORT NETWORKS

National research and education networks

NRENs are specialised internet service providers who support the needs of research and education communities within a country. They promote access to global educational resources and facilitate interaction at both national and regional levels among higher education and research institutions.

Major NRENs in Africa and South Asia

Africa

- WACREN - West and Central African Research and Education Network
- GARNET - Ghana
- NgREN - Nigeria
- ENREN - Egypt
- SudREN - Sudan
- SomaliREN - Somalia
- UbuntuNet Alliance for Research and Education Networking - the Alliance of NRENs of East and Southern Africa
- EthERNet - Ethiopia
- KENET - Kenya
- MAREN - Malawi
- RENU - Uganda
- RwEdNet - Rwanda
- TENET/SANReN - South Africa
- TERNET - Tanzania
- ZAMREN - Zambia

South Asia

- BDREN - Bangladesh
- ERNET - India
- NKN - India
- NREN - Nepal
- PERN - Pakistan
- LEARN - Sri Lanka

Thanks to Andy Nobes, INASP

EXPERIENCING CHALLENGES IS NORMAL

Group discussion

1. What tools and assistance can you utilize to address some of the problems you have listed?
    1. Using your hand-out sheets, go through the categories of RDM and RCR and think about how to overcome some of the challenges you have listed
    2. Discuss these specific, or general solutions in your groups
    3. How will these resources enhance your ability to be responsible and open?

THINKING ABOUT NEXT WEEK: IMPACT OF DIGITAL SCIENCES

- technology affects communication, collaboration and knowledge exchange within scientific, work and home settings.
- people adapt to the rapid social changes brought by innovation and need to be assisted to use those innovations *more productively and safely.*
- need to consider the ways in which new technologies can be designed and developed to be *more responsive* to societal acceptability and desirability.

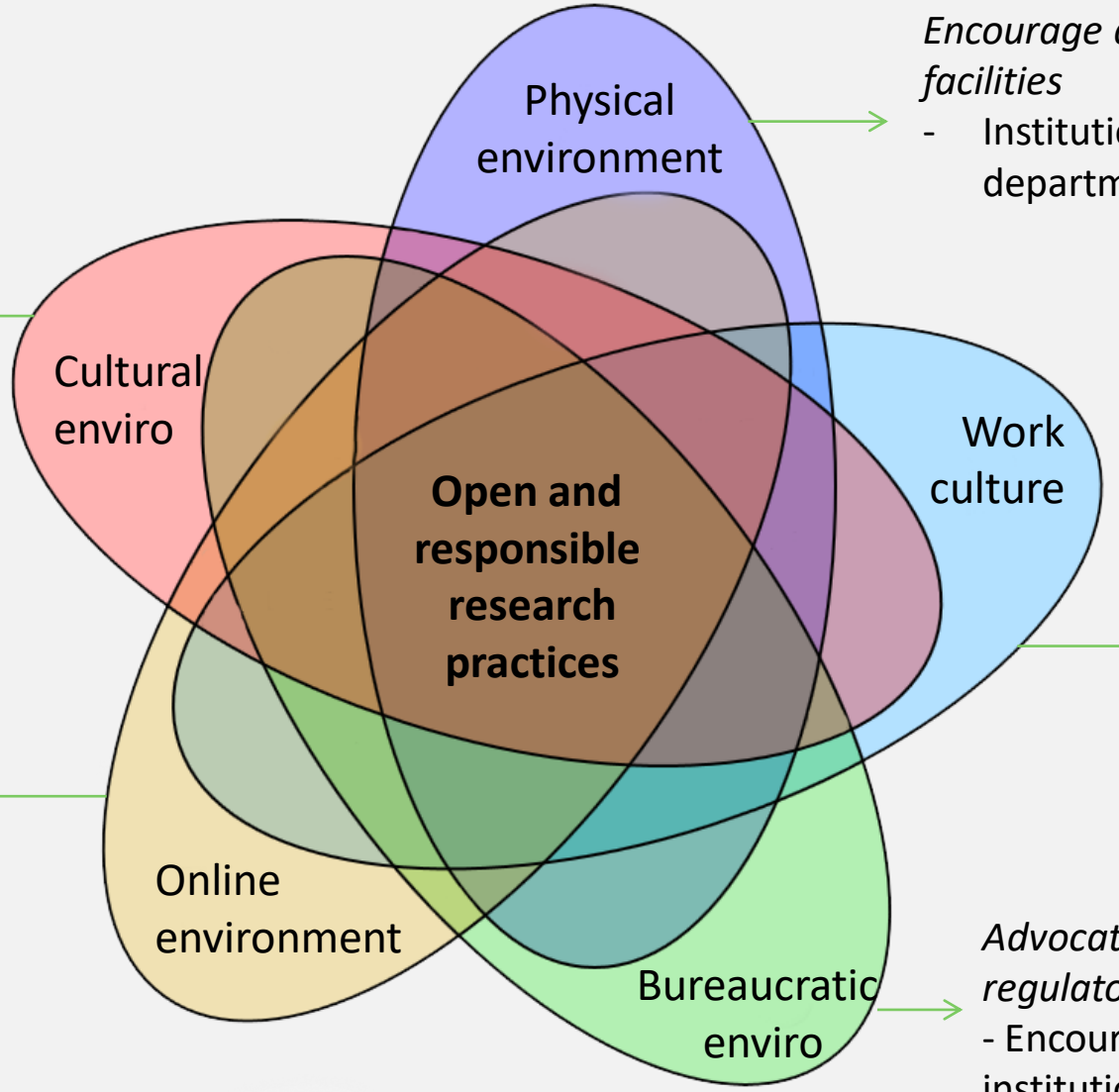
INDIVIDUAL ACTIVITIES … GLOBAL IMPACT

- Being a responsible and open science citizen involves more than just making sure that your own data practices are ethical
- Being part 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 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



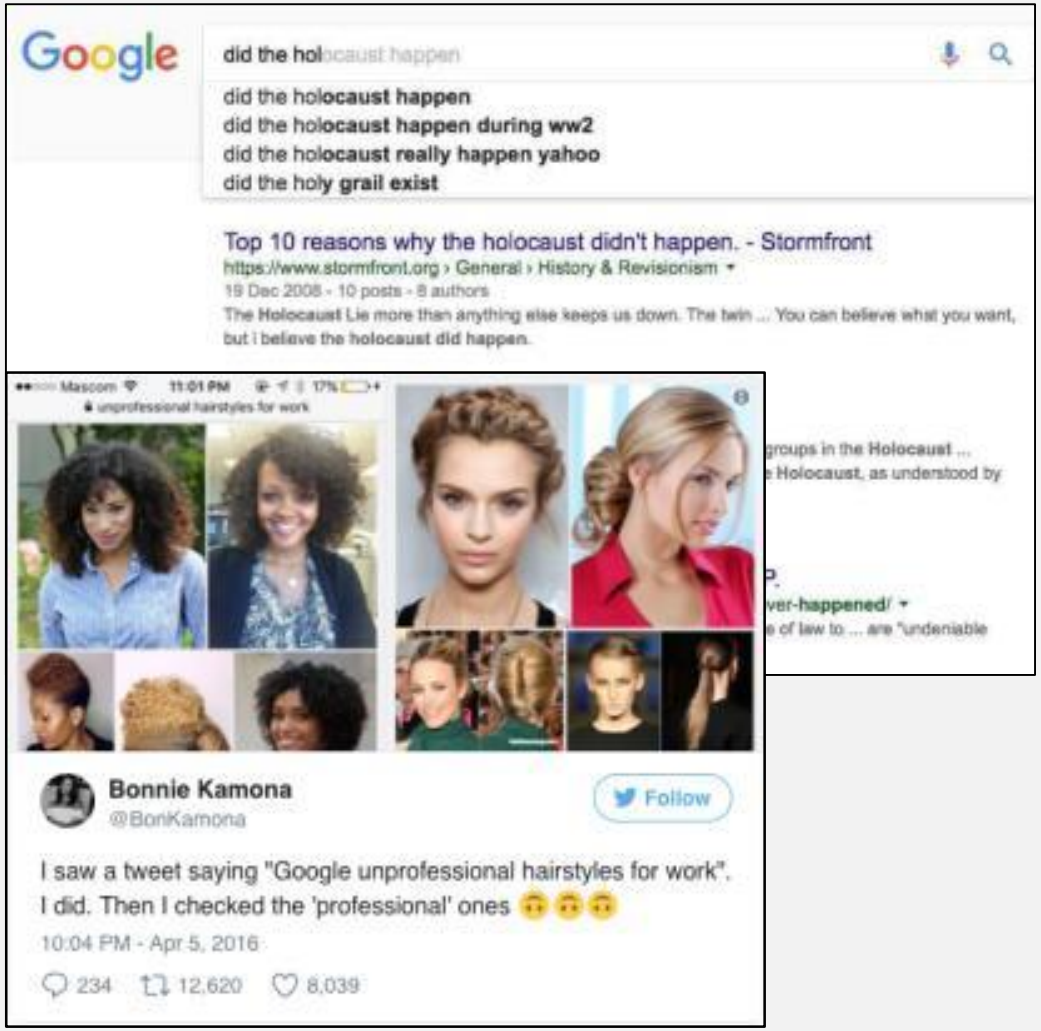
Improve openness and responsibility in your work environment
- Advocate for RDM practices
- Engage colleagues in discussions on RCR, RDM, FAIR
- Encourage data dissemination beyond disciplines

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

INFRAETHICS ...

CURRENT CHALLENGES



Bias

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

= inclination or prejudice 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.

"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)

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.

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 …

CHALLENGE FOR NEXT WEEK

Group discussion

1. 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