Open and Responsible Research 4 Looking at the Bigger Picture

Louise Bezuidenhout
Emanuele Ratti

CODATA-RDA Data Science School

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
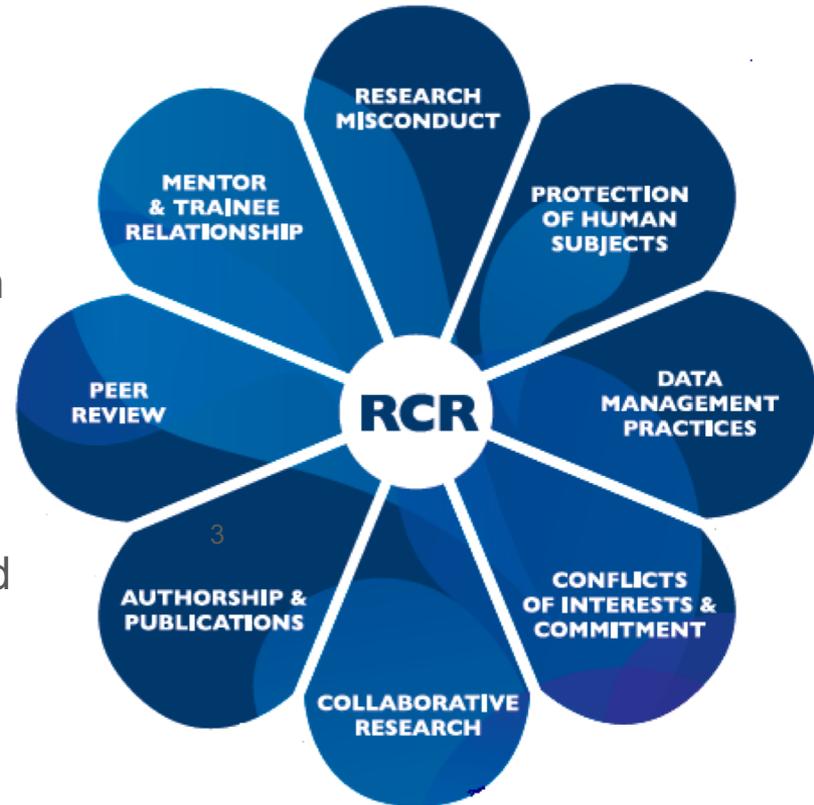
Looking at the Bigger Picture: Ethical Challenges of Data Science

Challenges Beyond the Lab

- Technology affects **communication**, **collaboration** and **knowledge exchange** within scientific, work and home settings:

1. help people to use those innovations **more productively and safely**

1. 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 <u>contain biases</u>
- The data that you select in your analyses can <u>produce biases</u>
- The algorithms that you design can <u>perpetuate biases</u>
- The websites, platforms, sharing pathways that you design, endorse or populate can perpetuate <u>discrimination</u>
- The data you generate can be re-used, re-combined, re-purposed in <u>unexpected ways</u>

6

(Algorithmic) Bias

(Fazelpour and Danks 2021)

- A bias is a **systematic deviation** from norms or standards

- Algorithmic bias is a deviation in **outputs, performance, or impact** relative to a defined norms/standards

- <u>Moral, statistical, or social</u>

- Example: algorithms and loans; algorithms and self-driving cars

Random error   Systematic error
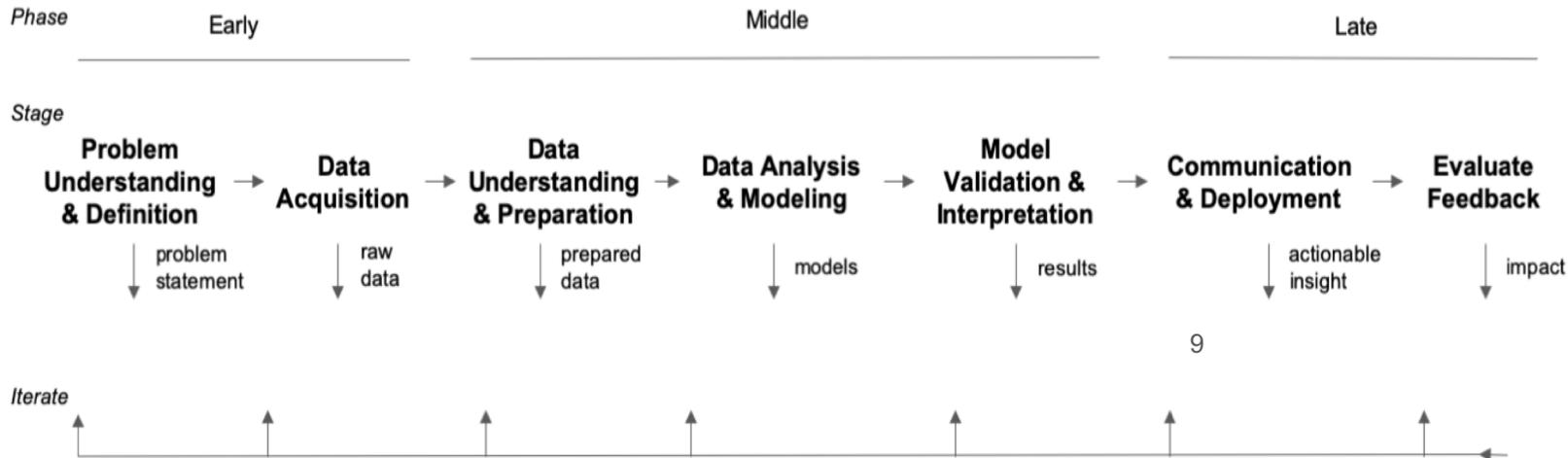
7

(Algorithmic) Bias

(Fazelpour and Danks 2021; O'Neill 2016)

- Algorithms are **value-laden**
- They are **morally loaded** not just because you can use algorithms for harms
1. The design of tools such as machine learning algorithms depend on **people's judgement**, which is replete with **value-judgement** (including metrics)
2. Algorithms are designed to **promote** certain kinds of **decisions**, which are deemed to be **instrumentally valuable**

(Algorithmic) Bias

- <u>How do biases arise?</u> (**Fazelpour and Danks 2021**)

Phase	Early		Middle			Late	
Stage	**Problem Understanding & Definition** →	**Data Acquisition** →	**Data Understanding & Preparation** →	**Data Analysis & Modeling** →	**Model Validation & Interpretation** →	**Communication & Deployment** →	**Evaluate Feedback**
	↓ problem statement	↓ raw data	↓ prepared data	↓ models	↓ results	↓ actionable insight	↓ impact

9

Iterate

(Ratti and Graves, *under review)*

Current Challenges

Google — 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.

Why Google Search Results Favor Democrats

It's not because the company is biased—it's more complicated.

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

10

Bonnie Kamona
@BonKamona   🐦 Follow

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

(Algorithmic) Bias

(Fazelpour and Danks 2021)

- There are **several solutions** to biases

1. Fairness through unawareness

2. Debiasing strategies

3. Etc

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
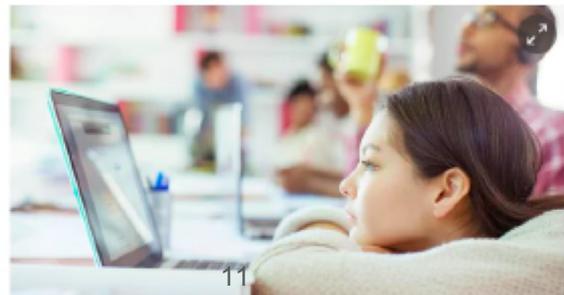
What Can I Do?

- 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 "<u>why do I think that way</u>"?

- 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

12

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

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 s**cientific community, public** and **future**

- **Not just** about responsible and critical use of data, **also about** scrutinizing evolving systems
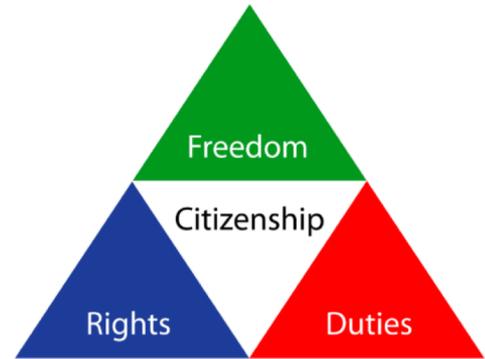
Extending Data Science Citizenship Responsibilities

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

Encourage development of open storage facilities
- Institutional repositories, centralized departmental sharing

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

Advocate for development of data regulatory frameworks
- Encourage development of data policies at institutional and policy level

Physical environment

Cultural enviro

Work culture

Online environment

Bureaucratic enviro

Open and responsible research practices

14