When do open science practices lead to higher quality data?

Danka Purić, Iris Žeželj, Ljiljana B. Lazarević & Goran Knežević
Laboratory for Research of Individual Differences

PSSOH, October 2019

What should science be?

- Ideal-case scenario:
  - Researchers formulate a question (based on previous literature)
  - They design a study that can answer the posed question
  - Collect data on a sample of adequate size (adequately powered)
  - Publish their findings, regardless of whether they are “positive” or “negative”
  - Future research builds upon these findings (cumulative nature of science)
  - In time, we have a good idea of which findings are “solid” (replicated) and which are not (self-corrective nature of science)
  - We can make informed decisions, not just in the academic domain
- Is this how it really works?

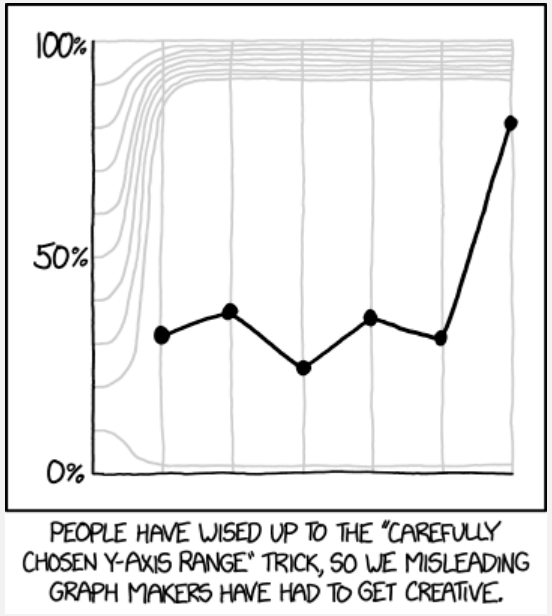
Problems with science

- "In 2011, German researchers in the drug company Bayer found in an extensive survey that more than 75% of the published findings could not be validated."

- "In 2012, scientists at the American drug company Amgen published the results of a study in which they selected 53 key papers deemed to be "landmark" studies and tried to reproduce them. Only 6 (11%) could be confirmed."

Why is this so?

- The interests of science and scientists are not always the same
- The scientific process is “hidden” from the public for the most part
- This leaves room for different questionable research practices (QRPs)
- What information do we actually have on the quality of data that we work with?



Open science practices

- A surge of initiatives supporting open science
- Almost undisputed acceptance
- The idea that open science can help prevent questionable research practices (QRPs)
- We will go through several mechanism by which open science practices could lead to better quality data
- And will also look into some (yet) unresolved issues

Registered reports

- Pre-registering a study before the data is collected

Registered reports

- Pre-registering a study before the data is collected

What is registered	What QRP are avoided

Registered reports

- Pre-registering a study before the data is collected

What is registered	What QRP are avoided
Hypotheses	HARK-ing (Hypothesizing After the Results are Known)

Registered reports

- Pre-registering a study before the data is collected

What is registered	What QRP are avoided
Hypotheses	HARK-ing (Hypothesizing After the Results are Known)
Analyses	p-hacking; selective reporting

Registered reports

- Pre-registering a study before the data is collected

What is registered	What QRP are avoided
Hypotheses	HARK-ing (Hypothesizing After the Results are Known)
Analyses	p-hacking; selective reporting
Sample size and structure	Underpowered studies; studies on inadequate samples

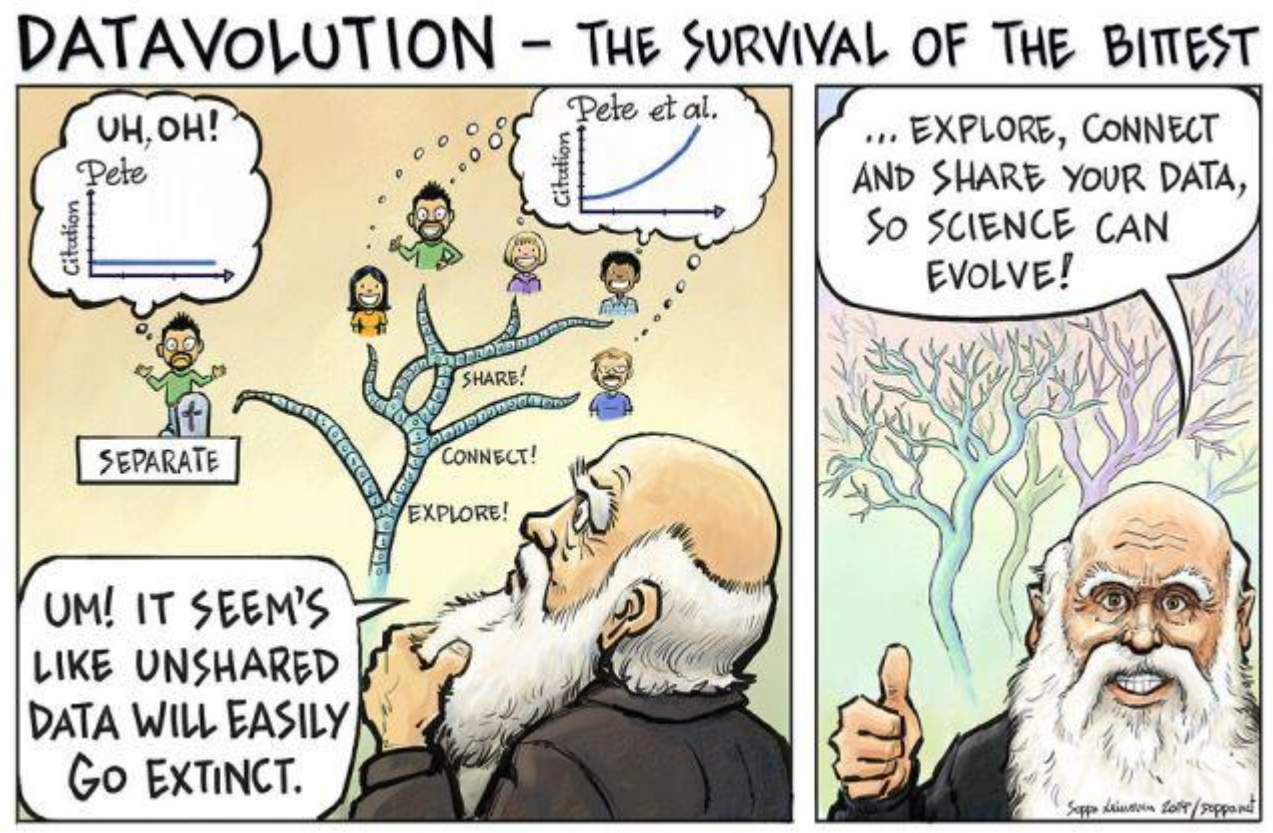
Registered reports

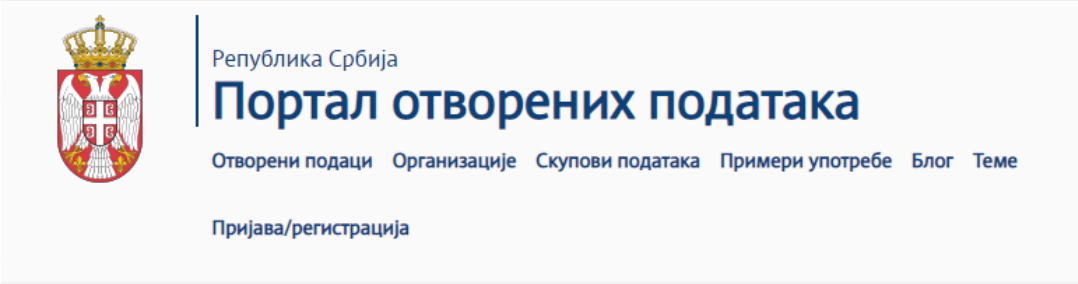
- Pre-registering a study before the data is collected

What is registered	What QRP are avoided
Hypotheses	HARK-ing (Hypothesizing After the Results are Known)
Analyses	p-hacking; selective reporting
Sample size and structure	Underpowered studies; studies on inadequate samples
Study design	Paper rejection based on „unfavorable“ results

Sharing materials, databases and scripts

- Making the materials/databases/scripts publicly available to others





Sharing materials, databases and scripts

- Making the materials/databases/scripts publicly available to others
- Anyone can check for errors and the author has a hightened sense of responsibility for what is being shared

What is shared	What is achieved

Sharing materials, databases and scripts

- Making the materials/databases/scripts publicly available to others
- Anyone can check for errors and the author has a hightened sense of responsibility for what is being shared

What is shared	What is achieved
Data	Reduced probability of unintentional errors as well as QRPs; the data is made „readable“ and more likely to be (re)used; data can be aggregated (https://www.gapminder.org/)

Sharing materials, databases and scripts

- Making the materials/databases/scripts publicly available to others
- Anyone can check for errors and the author has a hightened sense of responsibility for what is being shared

What is shared	What is achieved
Data	Reduced probability of unintentional errors as well as QRPs; the data is made „readable“ and more likely to be (re)used; data can be aggregated (https://www.gapminder.org/)
Scripts	Data is more easily verifiable; also more likely to be used; it’s easier to build upon previous work (saving resources); encourages the use of free/open software

Sharing materials, databases and scripts

- Making the materials/databases/scripts publicly available to others
- Anyone can check for errors and the author has a hightened sense of responsibility for what is being shared

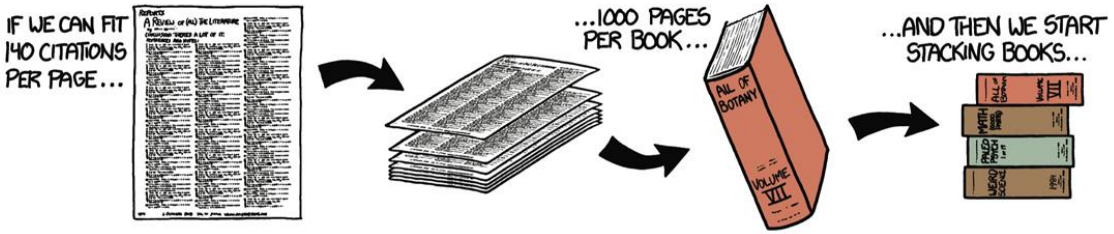
What is shared	What is achieved
Data	Reduced probability of unintentional errors as well as QRPs; the data is made „readable“ and more likely to be (re)used; data can be aggregated (https://www.gapminder.org/)
Scripts	Data is more easily verifiable; also more likely to be used; it’s easier to build upon previous work (saving resources); encourages the use of free/open software
Materials	Better understanding of the scope and generalizbility of findings; prerequisite for replications

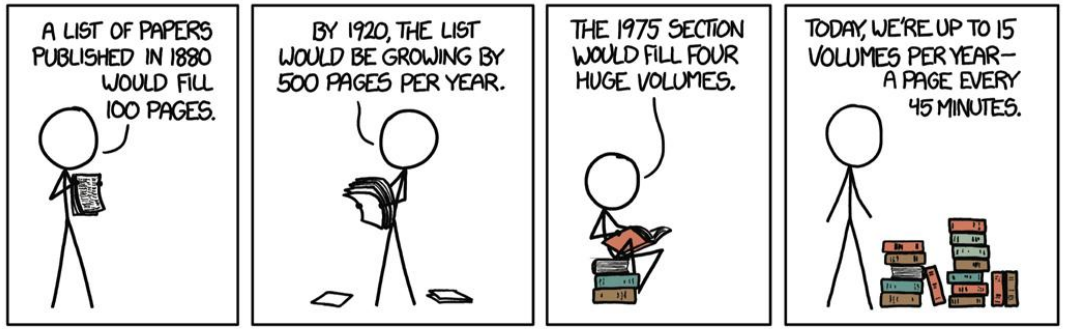
Opening access to published manuscripts

- Making manuscripts available to all – no paywall
- Reduces inequalities between researchers coming from different countries / institutions
- When everyone is up to date, global resources are put to better use
- New options for sharing pre-print versions of manuscripts (e.g., ArXiv, OSF, bioRxiv, PrePubMed) and receiving feedback from the scientific community

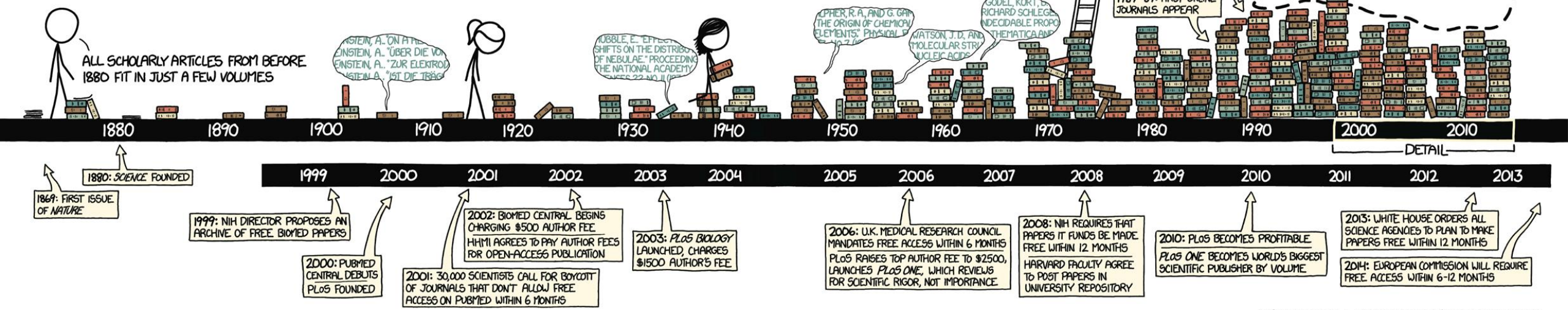
HOW MUCH SCIENCE IS THERE?

SCIENTIFIC PUBLISHING HAS BEEN ACCELERATING—A NEW PAPER IS NOW PUBLISHED ROUGHLY EVERY 20 SECONDS. LET'S IMAGINE A BIBLIOGRAPHY LISTING *EVERY* SCHOLARLY PAPER EVER WRITTEN. HOW LONG WOULD IT BE?



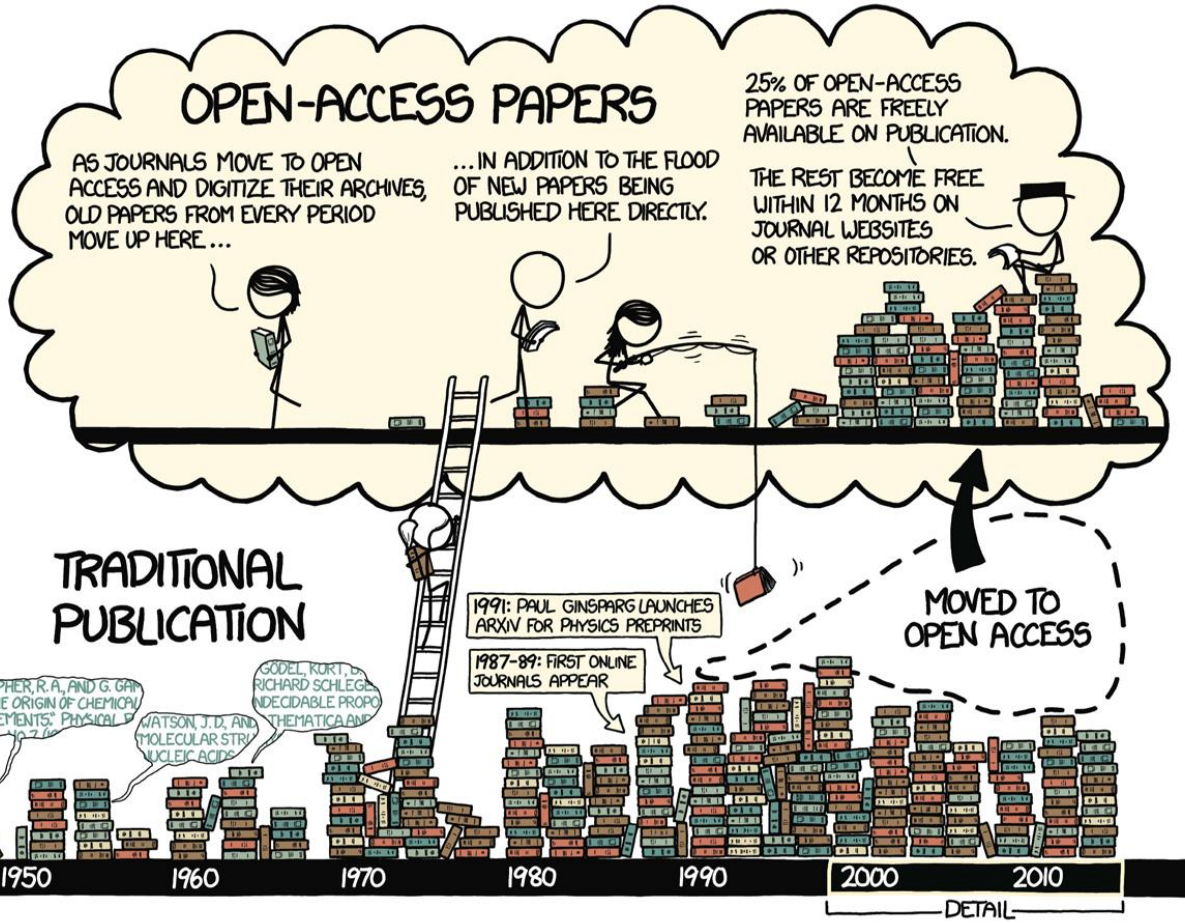


...THIS IS WHAT THE FULL LIST WOULD LOOK LIKE:



HOW OPEN IS IT?

SINCE THE ADVENT OF THE WEB, MUCH OF SCIENTIFIC PUBLISHING HAS BEEN MOVING TO *OPEN ACCESS*. ACCORDING TO SCIENCE-METRIX, OPEN ACCESS REACHED A "TIPPING POINT" AROUND 2011: MORE THAN 50% OF NEW RESEARCH IS NOW MADE AVAILABLE FREE ONLINE.



BY RANDALL MUNROE • REPORTING BY JOCELYN KAISER AND DAVID MALAKOFF

Scientific collaborations

- Increase the sheer quantity of data
- But also its quality:
  - Usually stringent in terms of study design and methodology
  - High power to detect effects of interest
  - Diverse (sub)samples
  - Higher academic impact
  - Reducing inequalities and strengthening all partners for future data collection

Open questions on open science practices

- Open science practices are relatively new, so a number of issues have yet to be resolved
  - Overuse of shared materials – lack of diversity in the data
  - Focusing only on replicated findings – we still need to do exploratory analyses
  - Anonymizing open data
  - Making sure non-experts do not make the wrong conclusions based on open data
  - Ownership of open data
  - Bureaucratic load on researchers practicing open science
  - …
- The most important thing is to be transparent about the desicions we make and aware of the limitations of our practices

Thank you for your attention!

Questions?

dpuric@f.bg.ac.rs, izezelj@f.bg.ac.rs, ljiljana.lazarevic@f.bg.ac.rs, gknezevi@f.bg.ac.rs

https://lira.f.bg.ac.rs/sr/