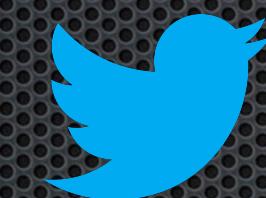


Understanding and practising open science



Angelos Krypotos
@AngelosVanKr

KU LEUVEN



Universiteit Utrecht



- What is your first name?
- Which part of the world are you now?
- Research interests?
- What do you want to get from this workshop?
- Favourite food?

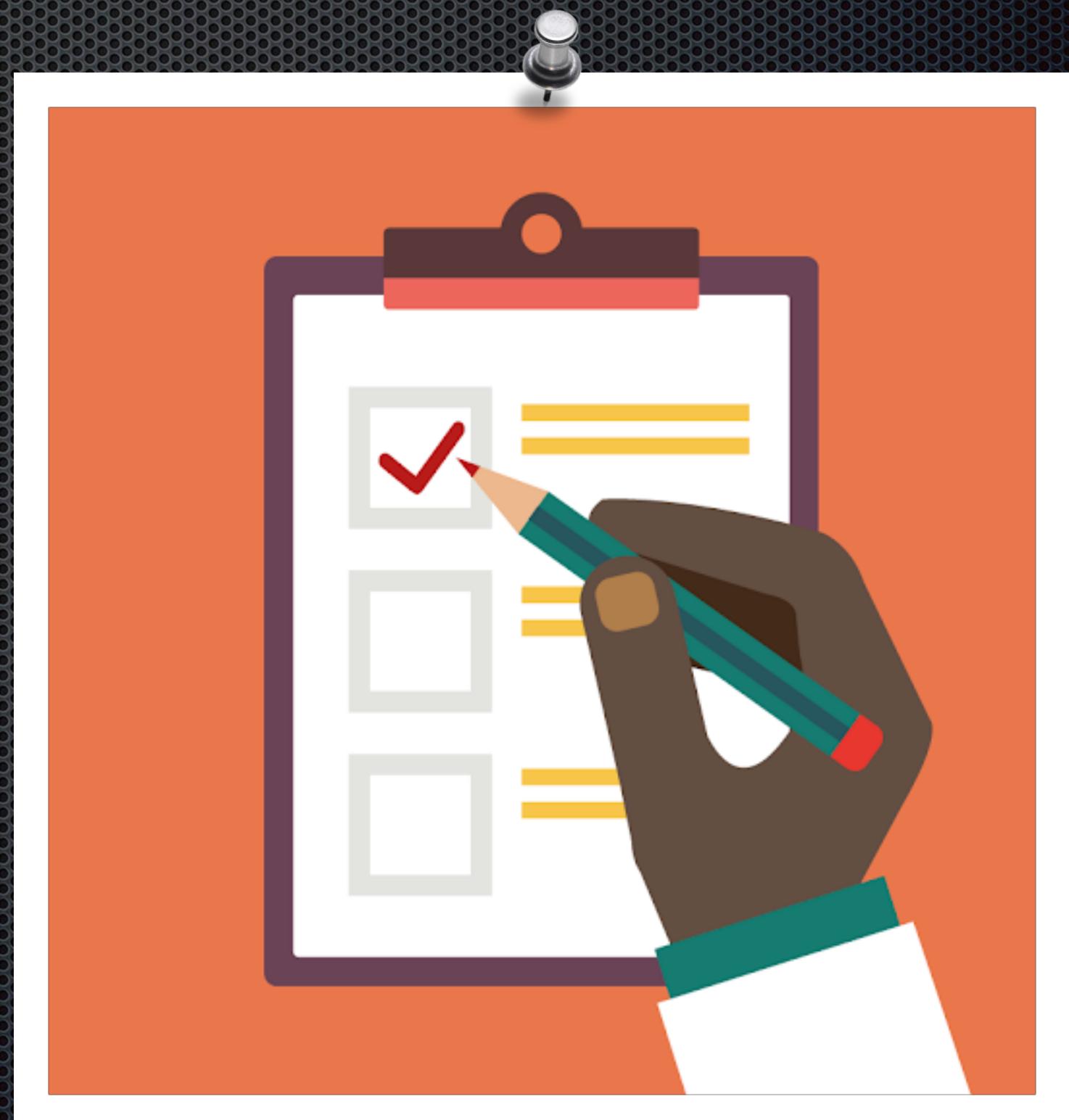
Conflict of interest statement

I am in not commercially affiliated to any of the software that I present here

Agenda

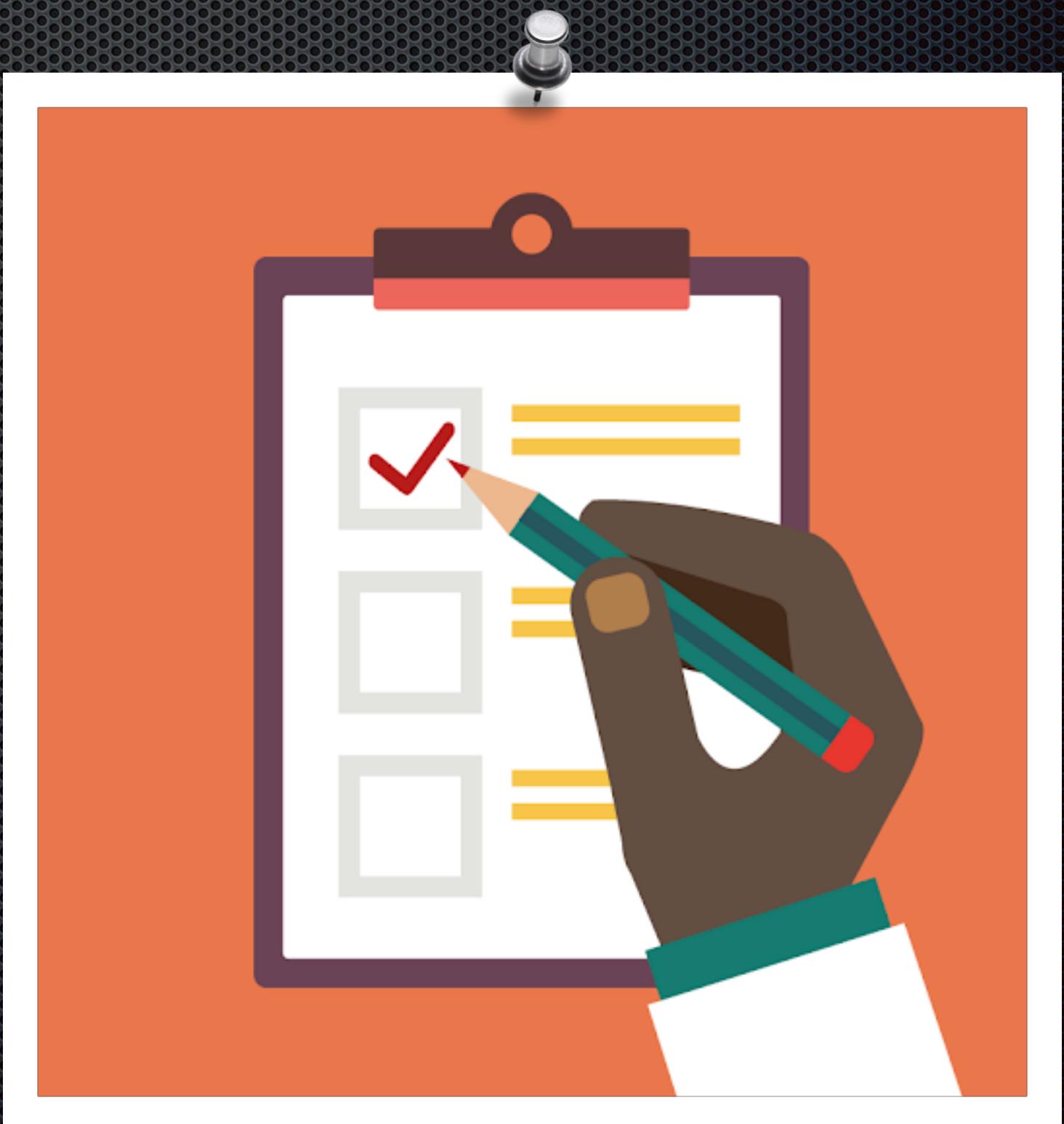
- Principles of open science
- Preregistration
- Open data
- Open code
- Preprints
- Wrapping up

HIIT blocks:
45 min on
15 min off

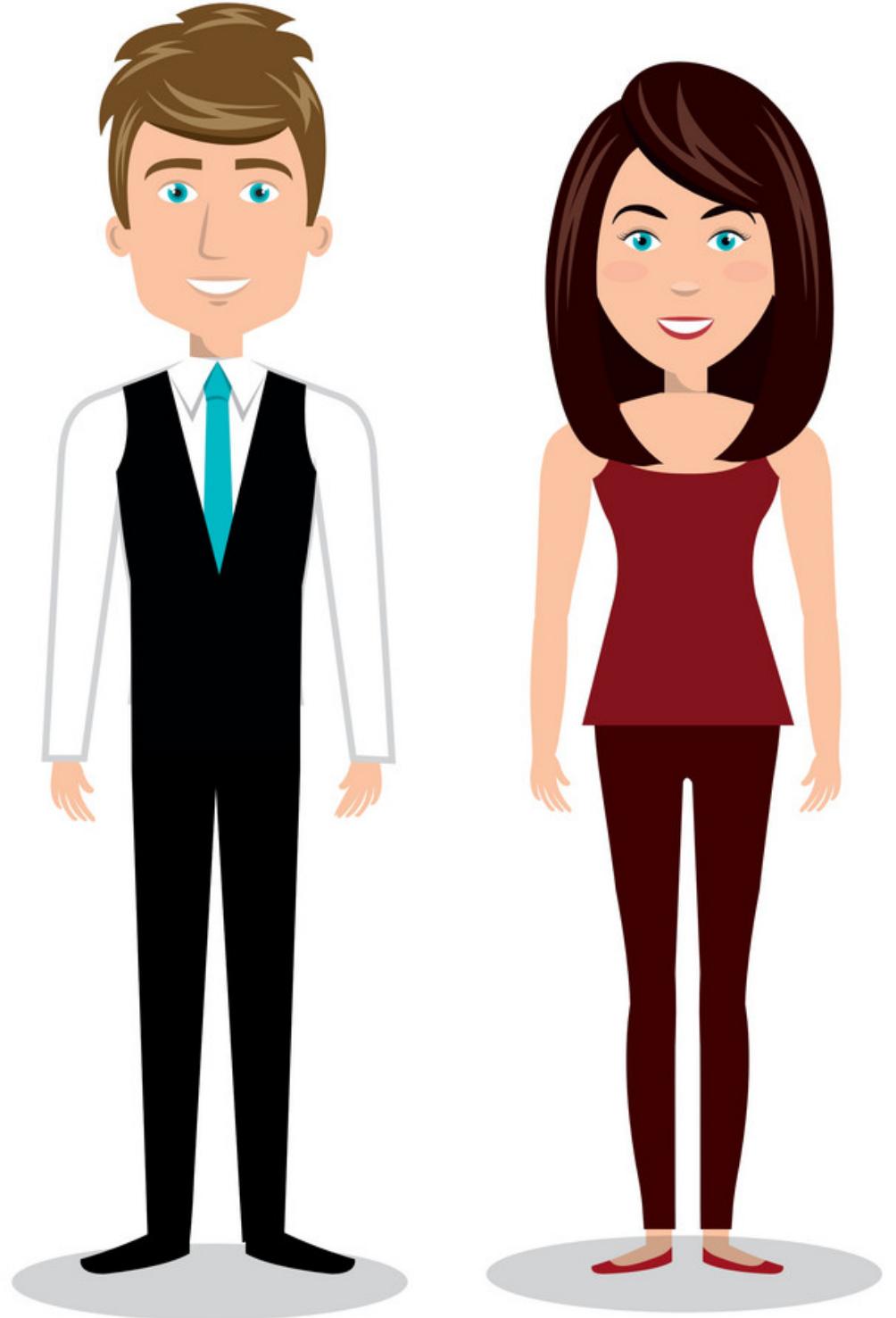


Agenda

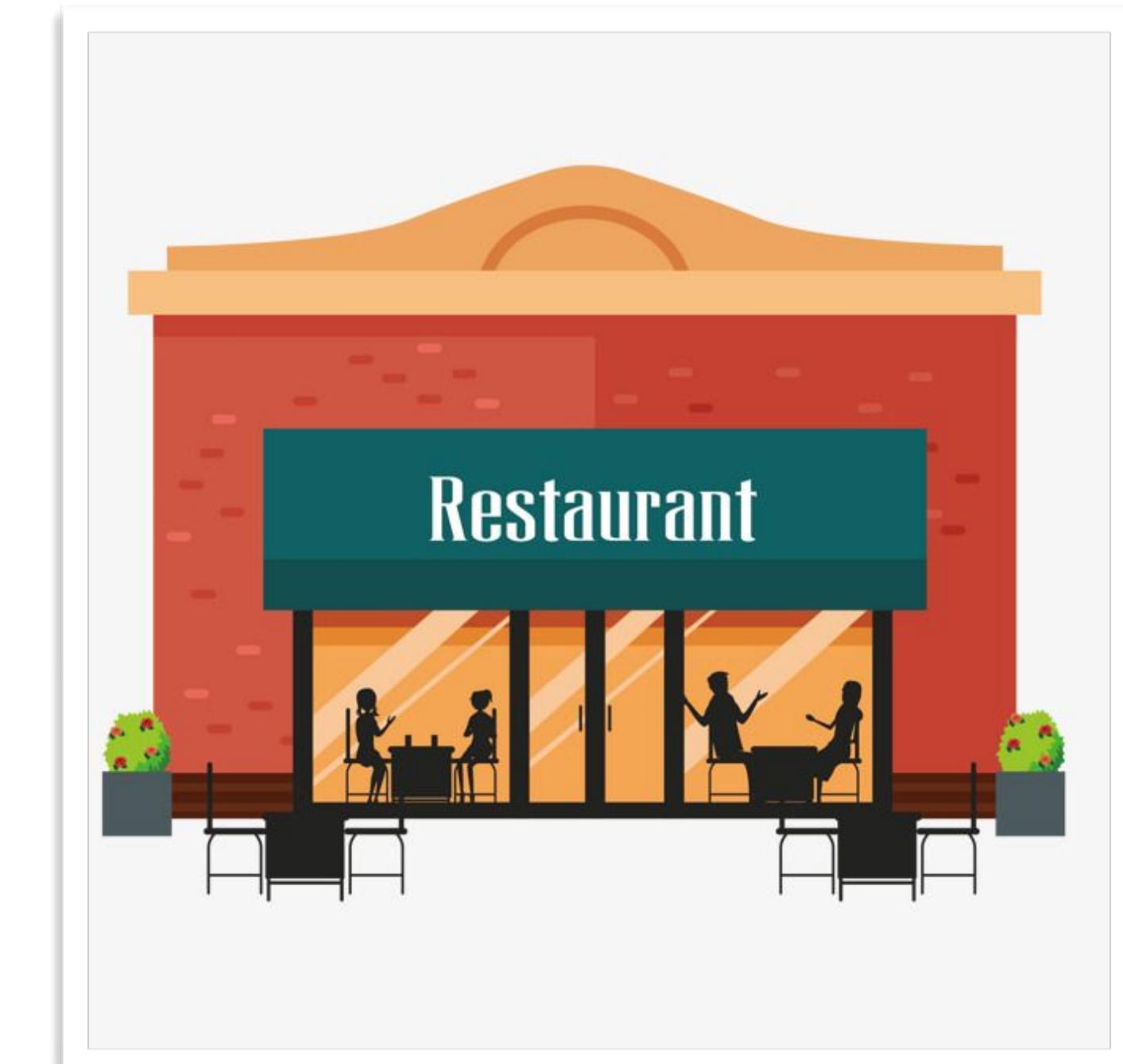
- Principles of open science
- Preregistration
- Open data
- Open code
- Preprints
- Wrapping up



Ben Barbara



Secrets



Open

Traditional Approach

- Carry out a study
- Share preliminary results with collaborators/conferences
- Publish the report of the paper
- Advantages: Many people know how to do it (but that does not mean they are right)

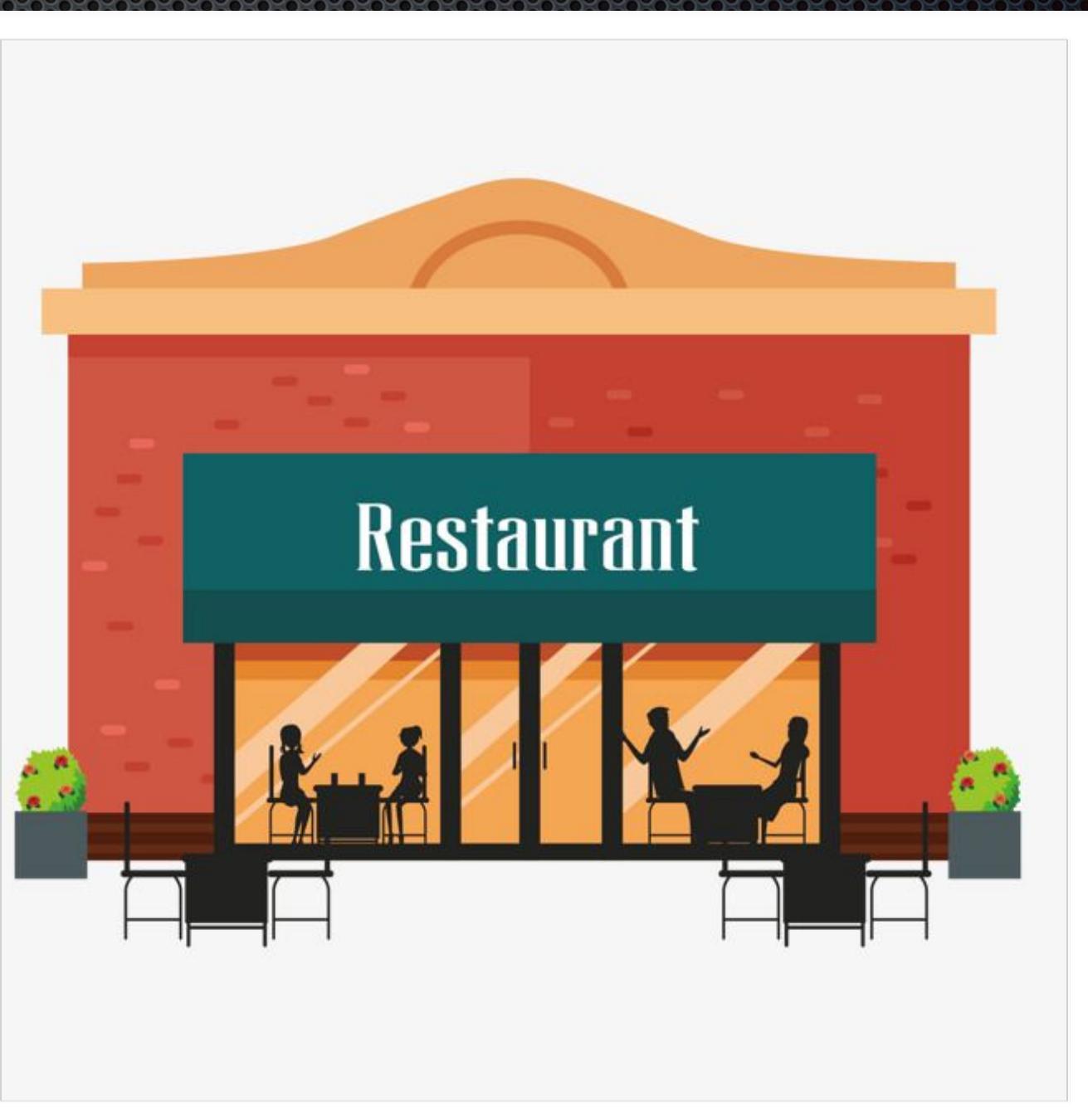


Disadvantages of current research practices

- Not transparent methods
- HARKing
- P-hacking
- Slow communication
- Often research not available to the public

The Open Science approach

- Share the design/analyses before data collection
- Carry out the planned analyses as soon as the data are in
- Deviations from the plan always explained
- Share data and material
- Advantages: Transparency, not loss of enthusiasm



Disadvantages of open science

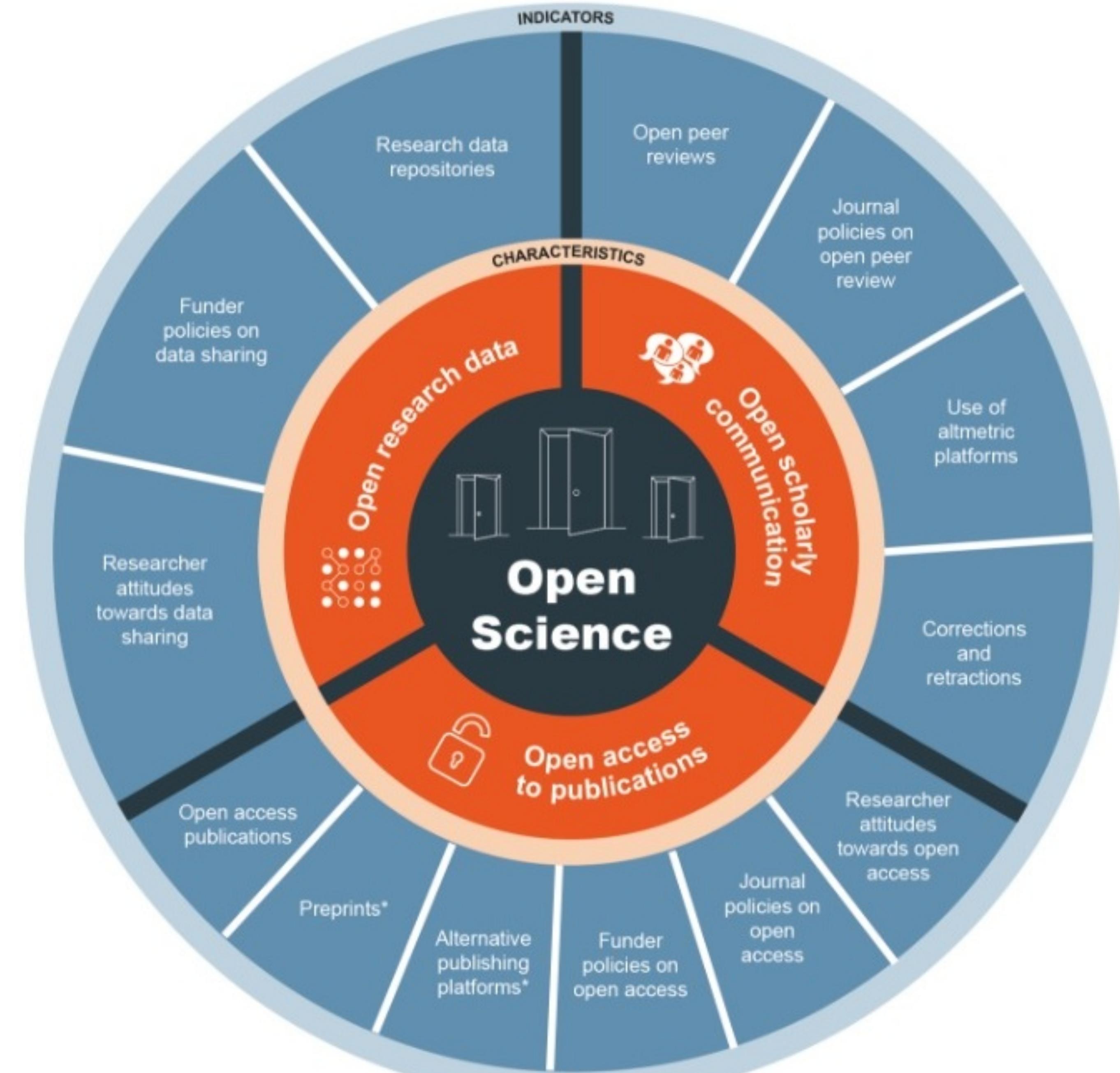
- You have to attend workshops to know how to do it

What is open science anyway?

Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process ([European Commission](#))

What is open science anyway?

Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The idea captures a systemic change to the way science and research have been carried out for the last fifty years: shifting from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process (European Commission)



Research circle/Open Science Practices

Hypotheses

Preregistration

Data collection

Open material
Logbook

Data analyses

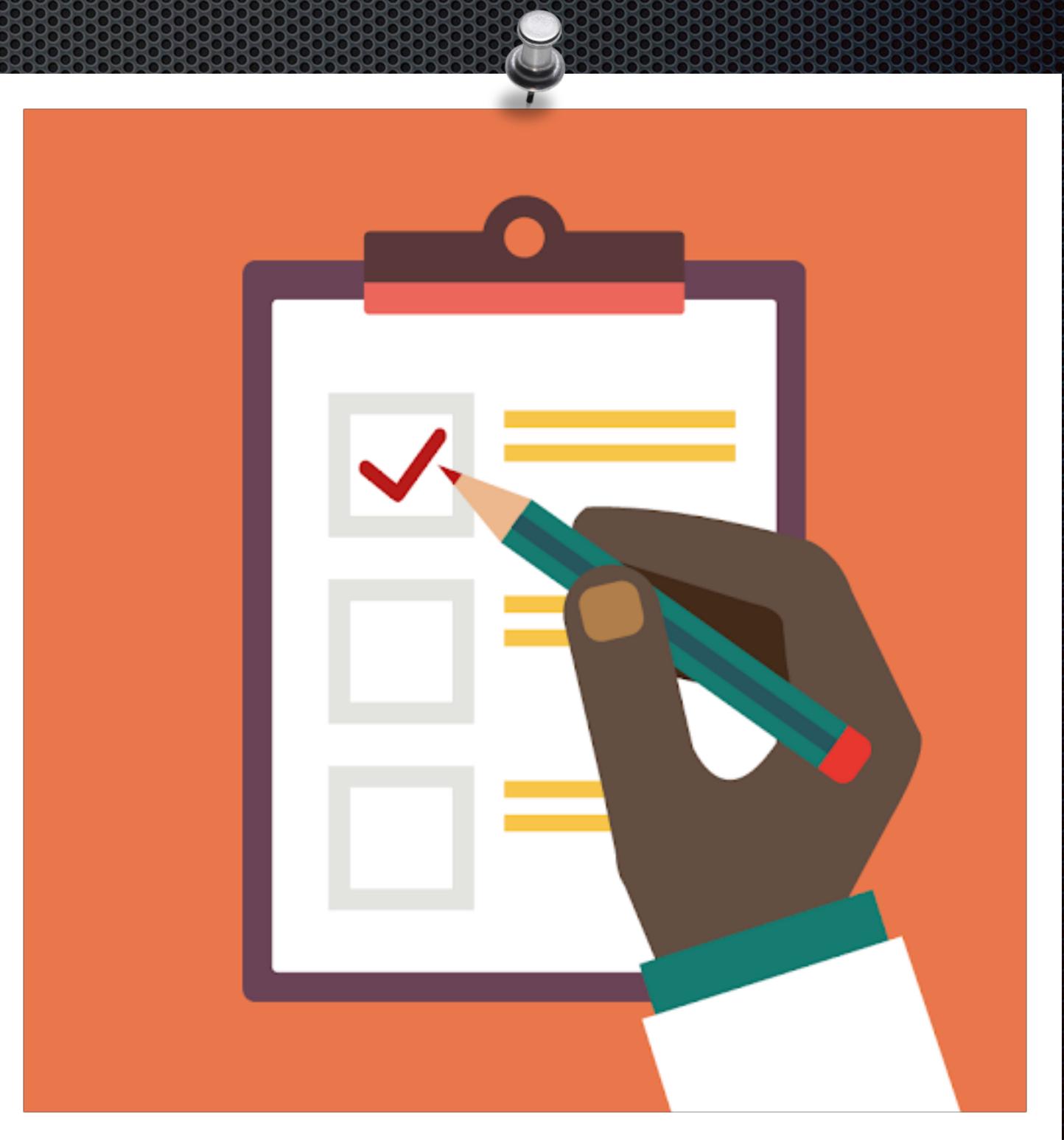
Open data
Open analyses

Publication

Preprint services

Agenda

- Principles of open science
- Preregistration
- Open data
- Open code
- Preprints
- Wrapping up

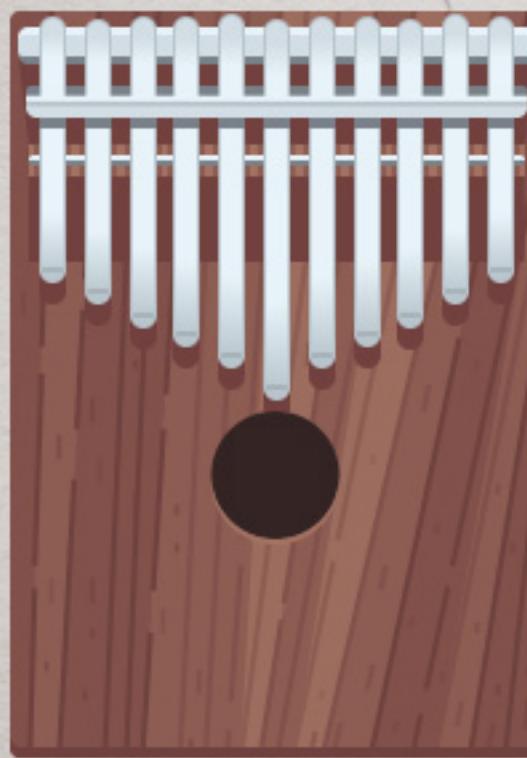


Preregistration

- Share a study's design, hypotheses, and statistical analyses prior to observing the outcomes of a study

Step number 1

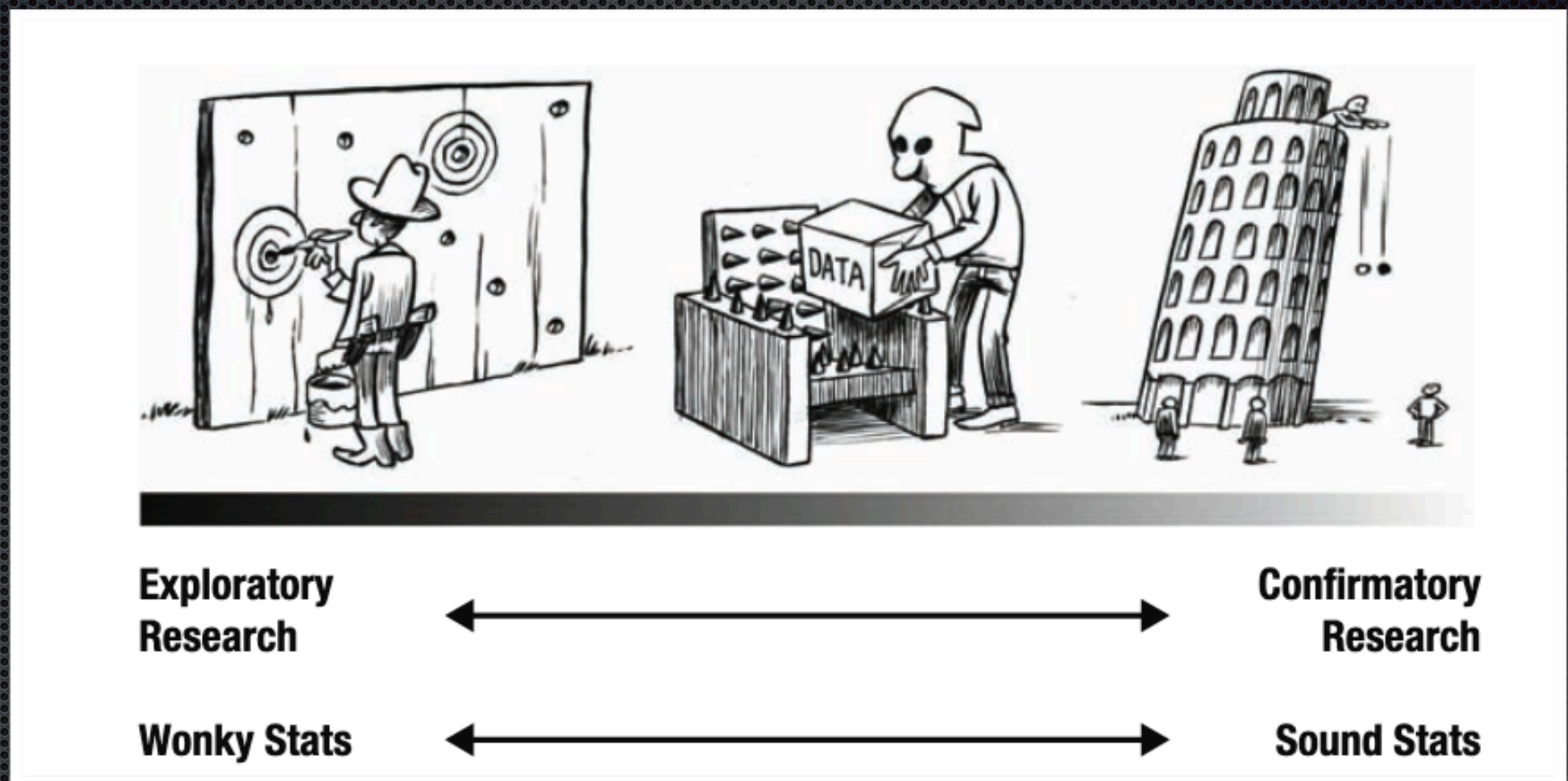
- Research idea
- Open science practices on their own do not make your idea valid



The Most
Important
Thing

Step number 2

- Type of study
 - Confirmatory
 - Exploratory



De Groot, 2014
Wagenmakers et al., 2012

Step number 3

- What data should I use?
 - Primary (original data)
 - Preexisting (yours or others)
- The type of data you use partially define the options you may have



Step number 3 (Original data)

- In case of original data
 - When are you gonna stop collecting data?
 - (e.g. Power analysis, optimal stopping)



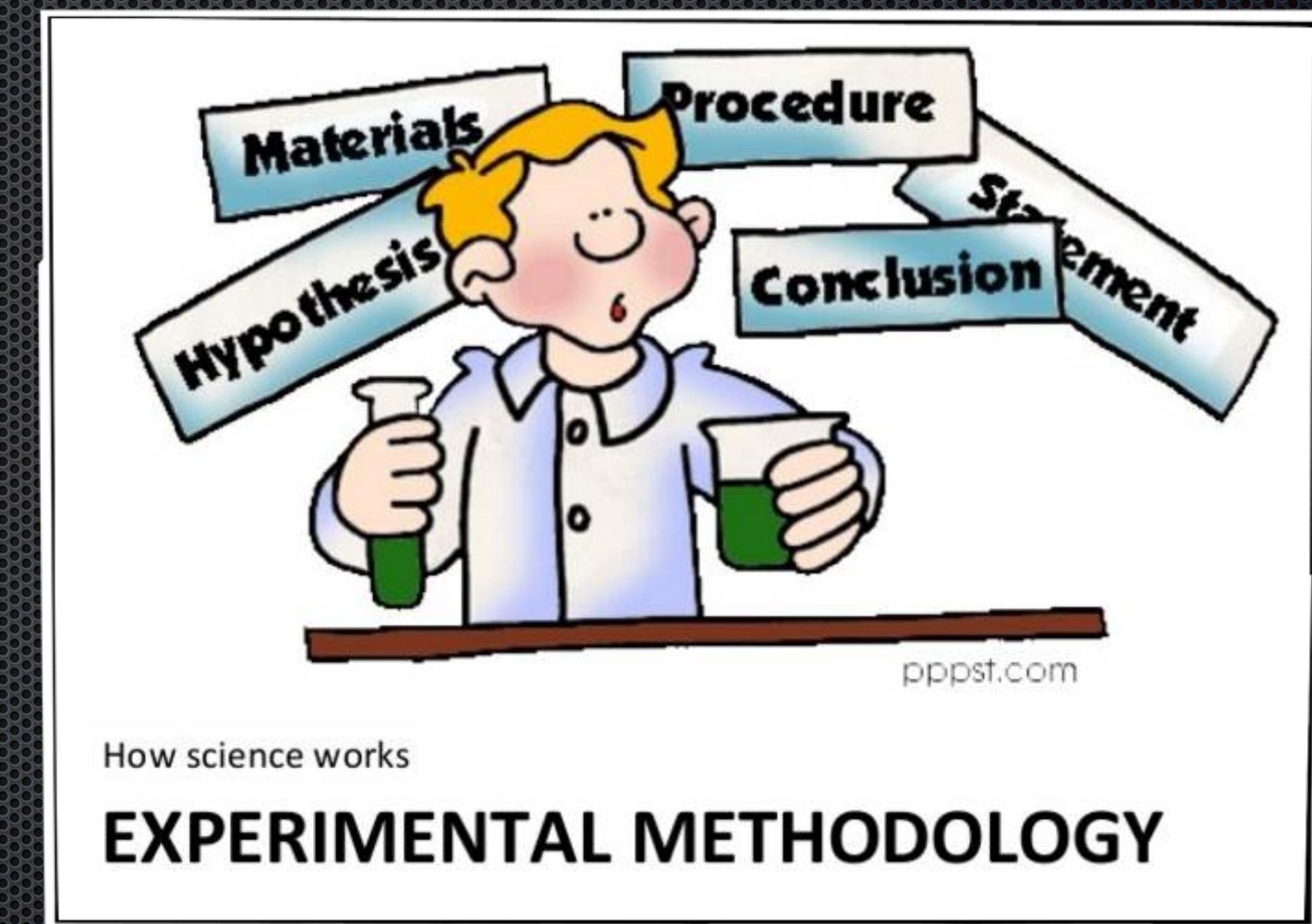
Step number 3 (Preexisting data)

- In case of preexisting data
 - What data you are going to use?
 - In case of a model: Validation/test data set



Steps number 4-5

- Method and analyses
 - Define everything (confirmatory)
 - Or not... (exploratory)



Benefits/shortcomings of preregistration

- Benefits
 - Think in advance about hypotheses/analyses
 - Take credit about your predictions
 - In some cases you are required to preregister (e.g., RCTs)
 - Just write up the discussion in the end and you are done!

Where to preregister?

- Open Science Framework (osf.io)
- As Predicted (aspredicted.org)
- Clinical trials (ClinicalTrials.gov)

This is the point where I show you how
to perform the preregistration at OSF.IO

Exercise 1

- Download preregistration template ([https://github.com/AngelosPsy/pssr/
blob/master/inst/rmarkdown/templates/pss_prereg/skeleton/skeleton.Rmd](https://github.com/AngelosPsy/pssr/blob/master/inst/rmarkdown/templates/pss_prereg/skeleton/skeleton.Rmd))
- Alternatively, just fill in the osf template.
- Nice examples can be found here: [https://osf.io/e6auq/wiki/
Example%20Preregistration](https://osf.io/e6auq/wiki/Example%20Preregistration)

Exercise 2

- Let's preregister the study at osf.io (<https://help.osf.io/hc/en-us/articles/360019738834-Create-a-Preregistration>)
- BUT NOT SUBMIT IT!

One last thing...



- Register reports — the best thing to happen against publication bias
- <https://www.cos.io/our-services/registered-reports>

Open data



- Sharing of all data that will, at least, reproduce the results of the main paper
- Enables: reproducibility
- Possible concerns: privacy (the General Data Protection Regulation; GDPR)

Your data should be:



Findable (read by humans and machines)

Accessible (USB vs. The cloud vs. OSF)

Interoperable (can be read in the future)

Reusable (a script could reproduce the results)

What data should I share?

- Data that will reproduce your results
- No identifiable data (e.g., names, date of birth)
- For multiple waves, use code where you will have the key - no openly available

What data should I share?

- Data could be shared: restricted number of individuals, individuals who request access, everyone
- Whether you can do anything from that, need to be consulted with the relevant departments at your institution
- Open science: unrestricted access usually by storing the data in the cloud

Data anonymization

- Remove non-relevant variables
- Remove personal identification (e.g., names)
- Bin variables (see k-anonymity)
- You can also simulate data (e.g., see synth pop R package)
- Also: <https://osf.io/26y8j/>



A screenshot of a spreadsheet application with a dark theme. The ribbon menu includes Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, and Acrobat. The toolbar includes standard options like Paste, Cut, Copy, Font (Calibri), Size (12), Bold, Italic, Underline, Alignment, and Wrap Text. The formula bar shows the cell reference E3 and the date 01/08/2000. The table has 5 rows and 7 columns. Column A contains 'participant name'. Column B contains 'Performance'. Column C contains 'Sex'. Column D contains 'Political opinions'. Column E contains 'Date of birth'. The data is as follows:

| | A | B | C | D | E | F | G |
|---|------------------|-------------|--------|--------------------|---------------|---|---|
| 1 | participant name | Performance | Sex | Political opinions | Date of birth | | |
| 2 | John Doe | bad | Female | Right | 01/08/2000 | | |
| 3 | John Lennon | bad | Male | Left | 01/08/2000 | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |

Open analysis/code

- Replicability vs. Reproducibilitiy
- A scripting language is best
- However, point and click programmes are also OK-ish

README

- Explain clearly what each column is
- Explain clearly what each row is
- Explain how some columns were created (e.g., sums of two columns)

JASP

- Advantages
 - Reproducible analyses
 - Really easy to use
- Disadvantages
 - No code available (at least for now)
 - Not so easy with data reduction

JASP Example (follow along)

R/Rstudio



- Programming language
- It allows to easily conduct your analyses and share your code
- Everything should be done with a button press

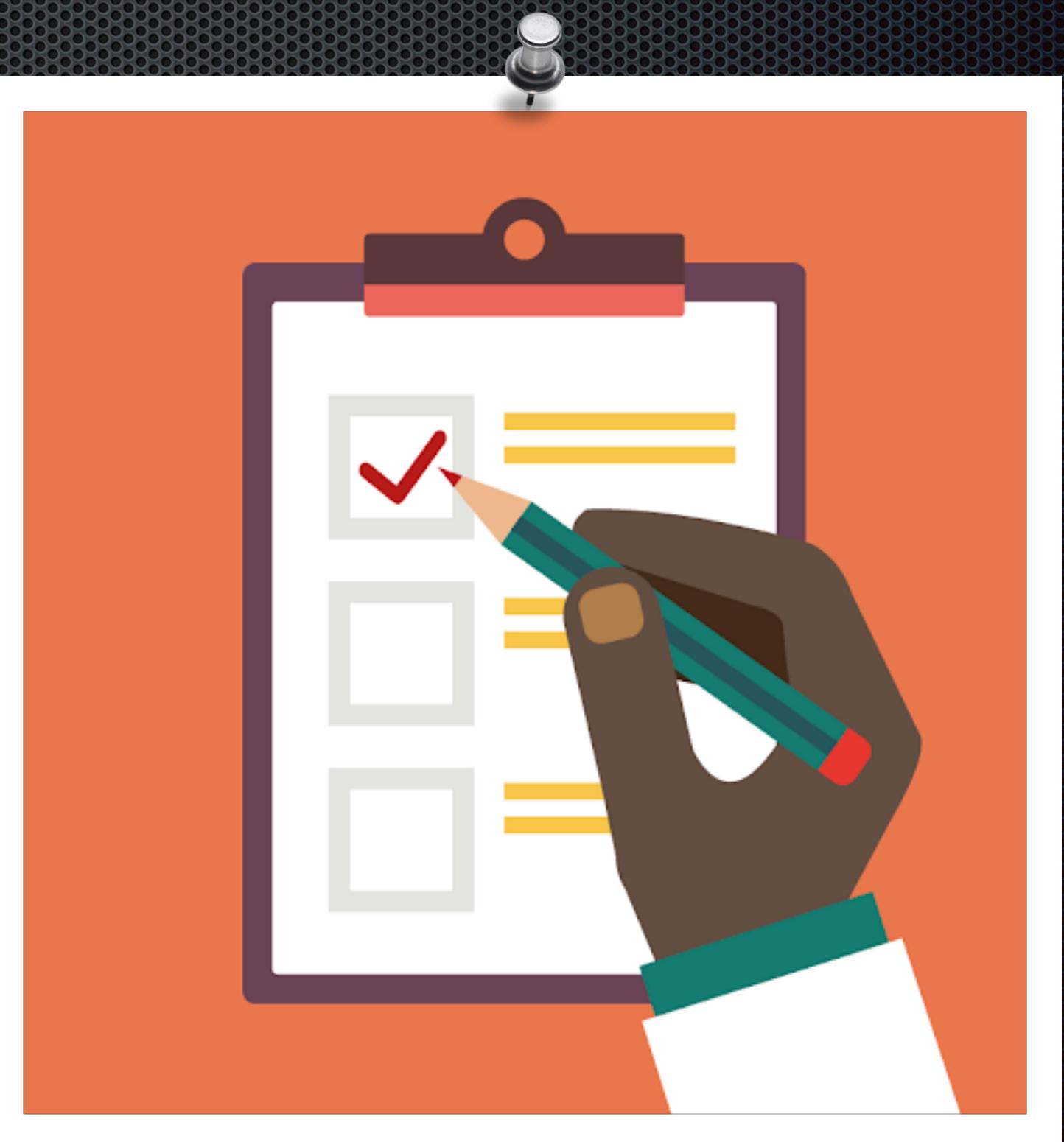
Git

- Version control system
- You can commit changes and comment on these changes
- It works regarding code, if you do not code, it does not do much

R/git Example (follow along)

Agenda

- Principles of open science
- Preregistration
- Open data
- Open code
- Preprints
- Wrapping up



Preprints



Pre-print/post-print

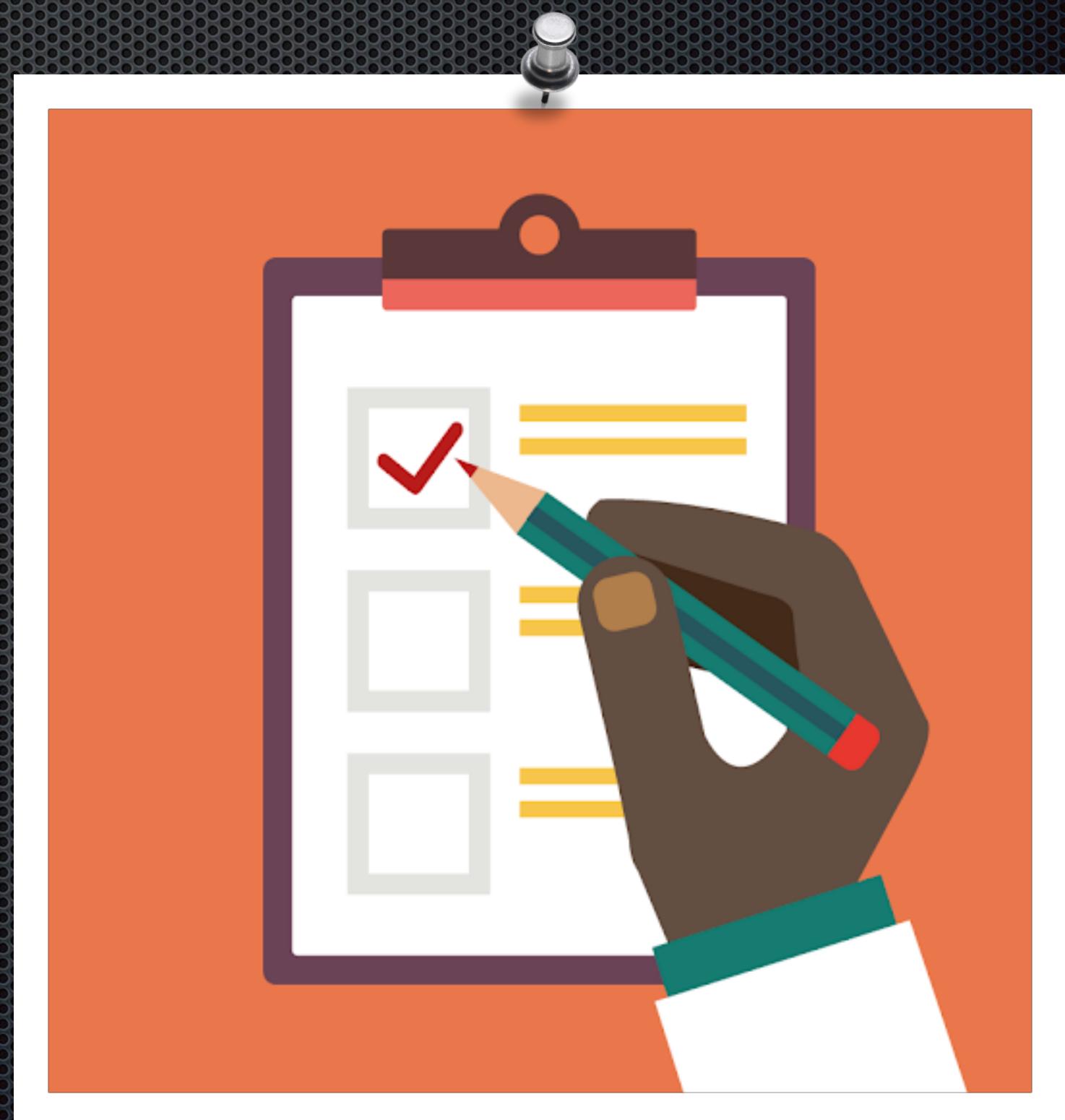
- Share your work prior to publication to a journal
- Receive comments and critique (+ maybe citations)
- (Maybe never publish it)
- Popular archives
 - arXiv
 - <https://psyarxiv.com>

Exercise (again...)

- Let's see how we can submit a fake paper
- But we will not actually submit it!

Agenda

- Principles of open science
- Preregistration
- Open data
- Open code
- Preprints
- Wrapping up



What did we learn?

- Open science can help you, science, society
- It may seem hard the first time – like almost everything
- In the end you save time, rather than waste time
- Open science should be just science

Don't fool yourself
— and you are the easiest person to
fool

It's a wrap

akrypotos.com



@AngelosVanKr