

# Open Science (with R)

Alessandro Tomassini

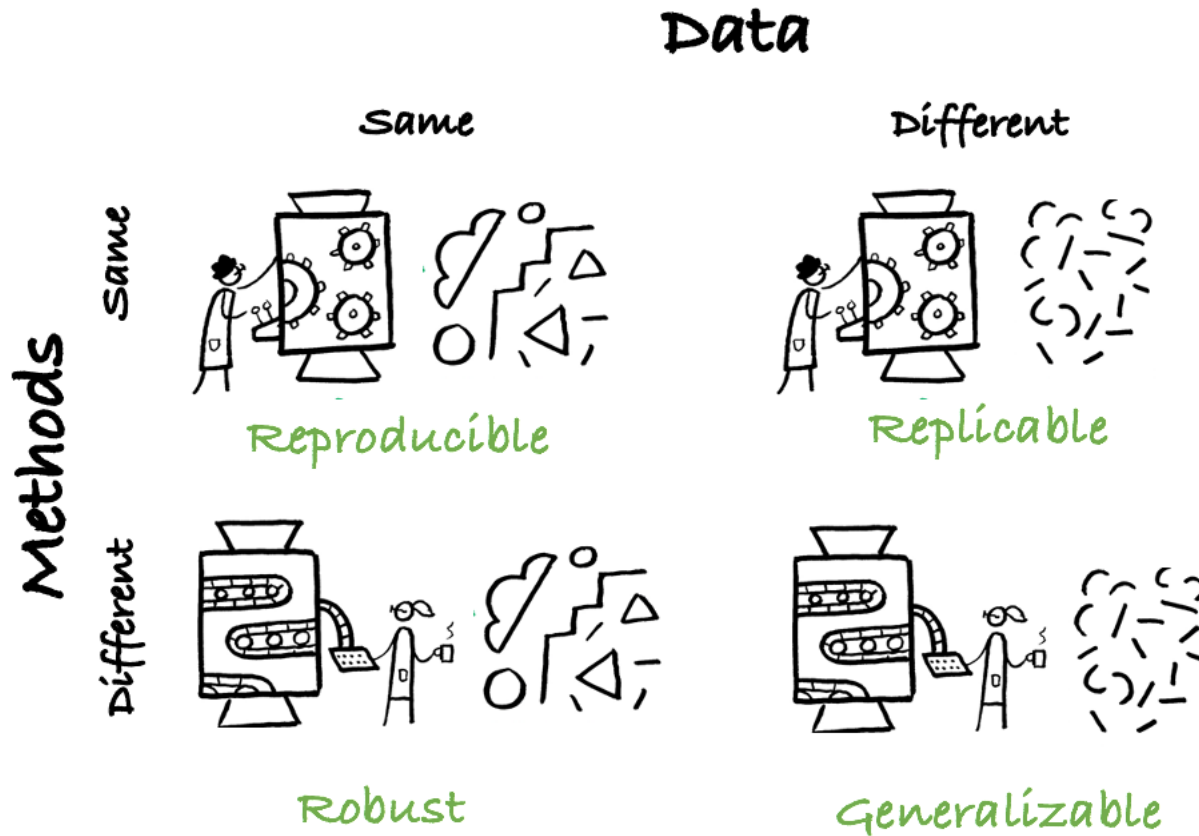
Lena Dorfschmidt

MPhil Basic and Translational Neuroscience

20.01.2021

# Science is NOT a Miracle!

# What are we talking about?



# Mismatch in Interest

## What's Good for Science



## What's Good for Scientists



# 5 Selfish Reasons to Work Reproducibly

---

1 2 3 4 5

Reproducibility helps avoid disaster

## U.S. student offers \$5,000 reward for PhD data stolen in Montreal

*With the smash of an SUV window on Saturday, a thief or thieves made off with years of Neil Kohan's work.*

**Montreal Gazette**

Nov 30, 2017 • Last Updated 3 years ago • 1 minute read

# *Apparent* Mismatch in Interest

What's Good for  
**Science**



What's **perceived as**  
Good for **Scientists**



# How to get started

Open research transforms science by aiming to make each part of the process (*data, software, hardware, results*):

- ...**publicly available**, so that anyone can benefit of it, without restrictions through paywalls and password.
- ...**transparent**, such that results and data can be accessed and reproduced with ease, ensuring time and effort goes into building on them.
- ...**reusable**, that is licensed appropriately, so that users know what they can and cannot do with it.

## Data Availability Statement

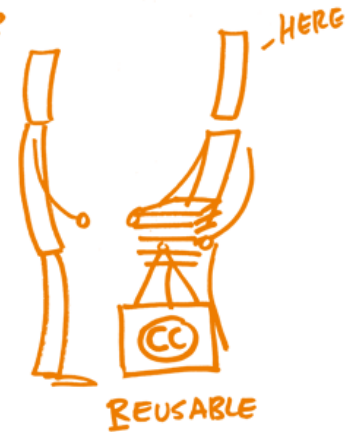
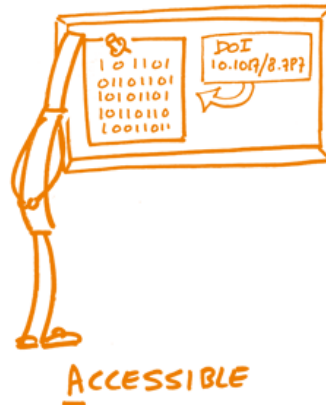
The datasets for this article are not publicly available because they have not been released yet to the public by the Indonesian Ministry of Health. Until that takes place requests to access the datasets should be directed to Dr Fatmawati, Head of the Microbiology Department at Sanglah Hospital, Denpasar, Bali. The request shall be sent via email to the following email address: [info@sanglah-hospitalbali.com](mailto:info@sanglah-hospitalbali.com).

## CODE AVAILABILITY

All processing and analysis code is available upon reasonable request.

# How to get started: Data

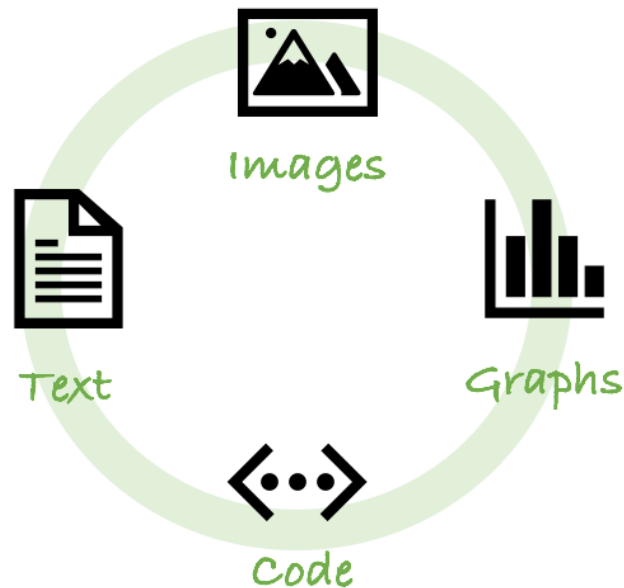
## FAIR DATA PRINCIPLES





# How to get started: Code

R Markdown allows you to share your work - code, text, images, graphs - all in one document.



# Further Resources

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. <https://doi.org/10.1038/533452a>

Markowetz, F. (2016). 5 selfish reasons to work reproducibly. Talk at scidata 2016. <https://www.youtube.com/watch?v=Is15CMVPHas&feature=youtu.be>

The Turing Way Community, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, ... Kirstie Whitaker. (2019, March 25). The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4). Zenodo. <http://doi.org/10.5281/zenodo.3233986>

Check out their [Youtube Channel](#) and [Website](#)

# Let's get started

# R Markdown

Reproducible papers: The idea that any paper can be obtained, its source material being handy, and then any change anyone likes can be made to it.

---

What is it?	Reproducible	Other benefits
-------------	--------------	----------------

---

- R Markdown is an add-on package to R
- It allows to write a single document that combines:
  - Text
  - Code
  - Plots
  - Images
  - Videos
  - Equations
- It is extremely versatile and can render the document in various formats:
  - PDF
  - MS Words
  - Web page
  - Book
  - Slides...actually these slides are based on R Markdown!

Version controls is an approach to record changes made in a file over time

---

What is it?

Reproducibility

How does it work?

Collaborative

- **Git** is a version control system to manage the evolution of a set of files - called a **repository**
- **GitHub** is a web service (like GitLab and Bitbucket) for storing git repositories online
- **Versioning** is the management of changes or revisions made in a file or project

# Open Science Framework

Open source project management tool: supports the whole project's lifecycle

---

Pre-registration

Data & code storage

Versioning

Preprints

*"The specification of a research design, hypotheses and analysis plan prior to observing the outcomes of a study"* Nosek & Lindsay (2018)

- Follows established templates
- Power sensitivity analysis
- Predefined hypotheses (*separating exploratory from confirmatory research*)
- Predefined analytical plan (*improves transparency*)
- Time-stamped
- Can be embargoed until acceptance of paper

# Questions?

**Now try your hands!**



