

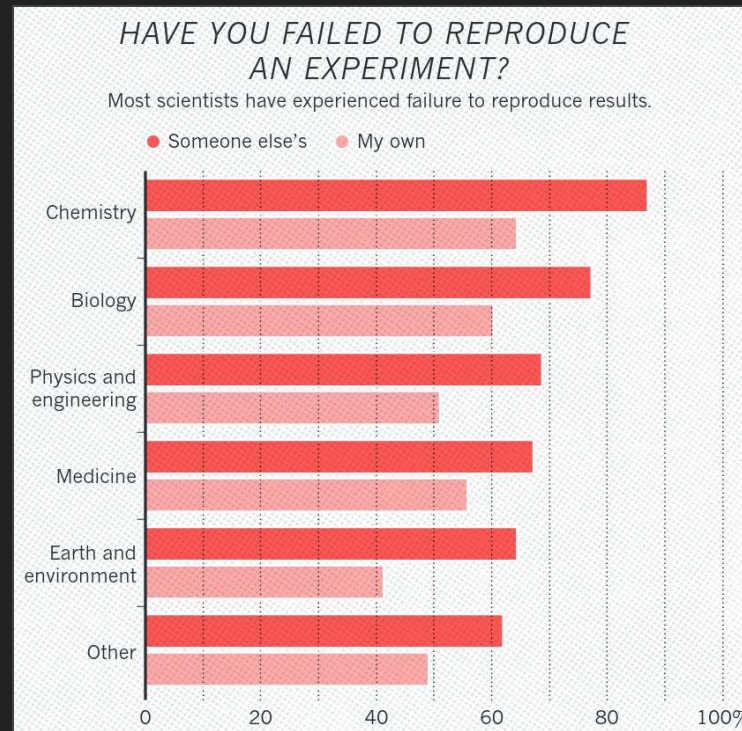
# Reproducible Data Preparation

Montreal, CA. Marc-Olivier Jodoin

“**Reproducibility** refers to the ability of a researcher to duplicate the **results** of a prior study using the same **materials** as were used by the original investigator.

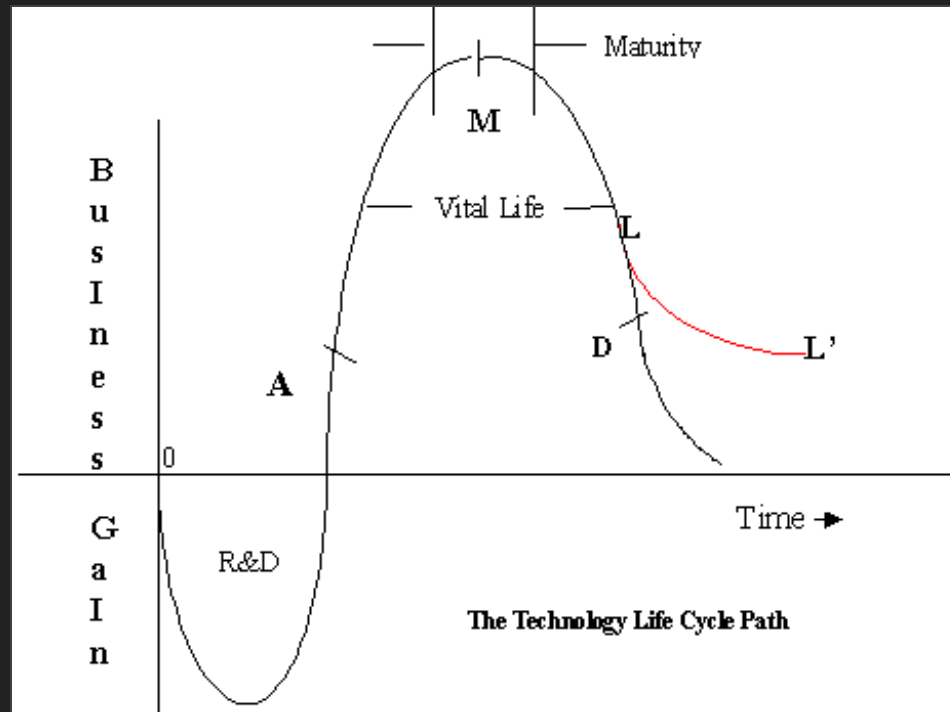
Reproducibility is a **minimum** necessary condition for a finding to be **believable** and **informative.**”

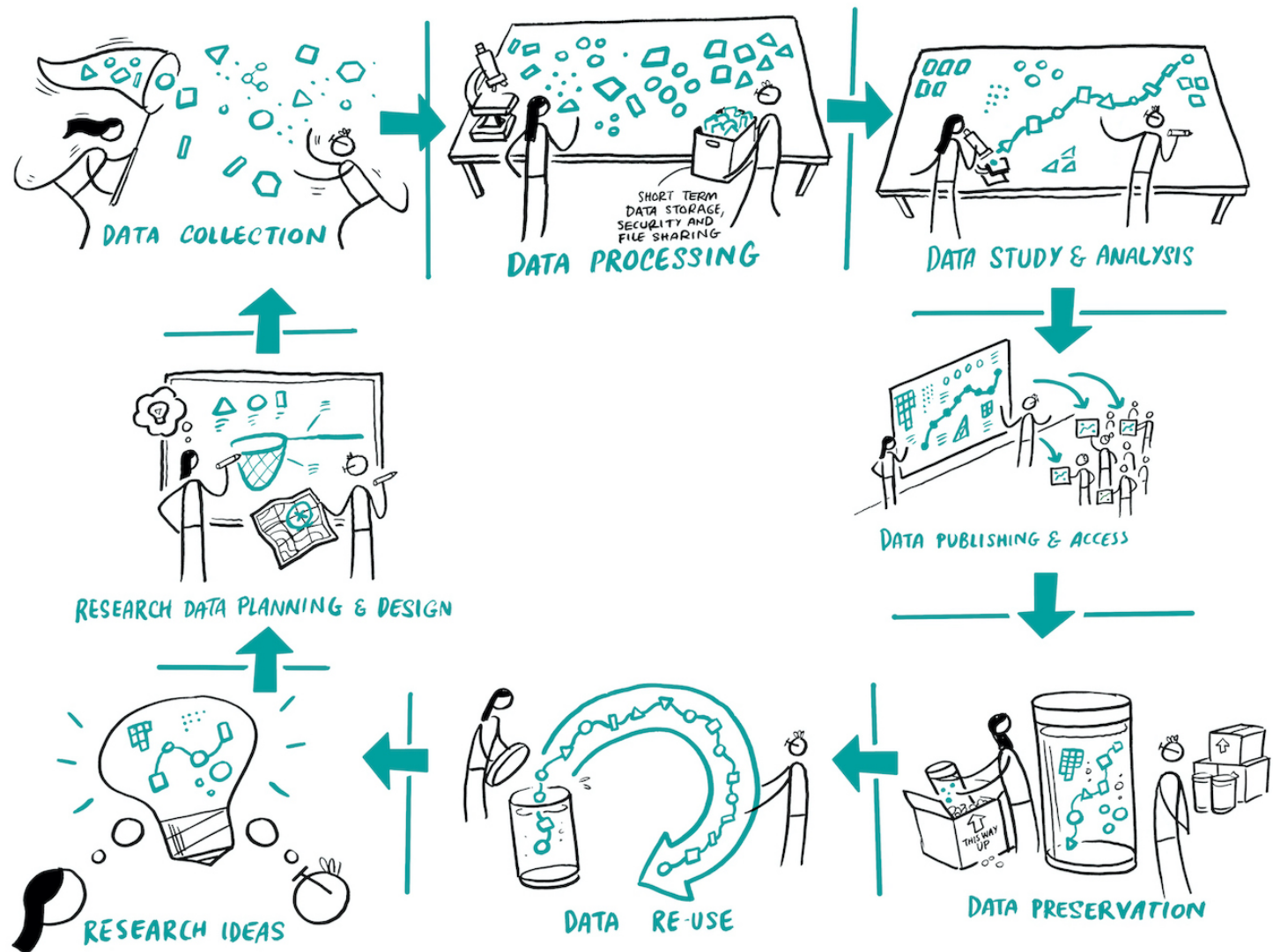
Report of the Subcommittee on Replicability in Science Advisory Committee, U.S. National Science Foundation, 2015



- Psychology: [Open Science Collaboration, Science \(2015\)](#)
- Cancer biology: [Begley, C. G. and Ellis, L. M., Nature 483, 531–533 \(2012\)](#)

# Technology Life Cycle

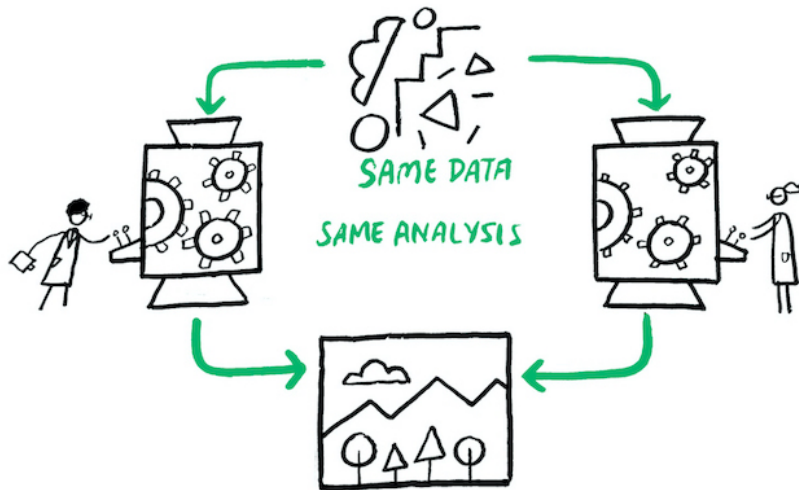




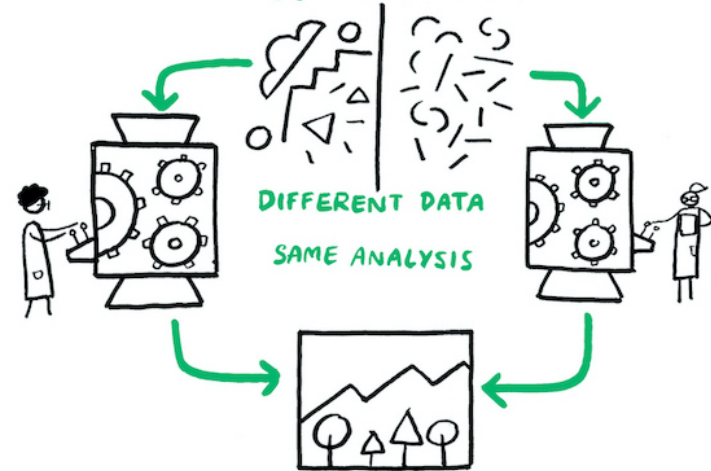
Scriberia 



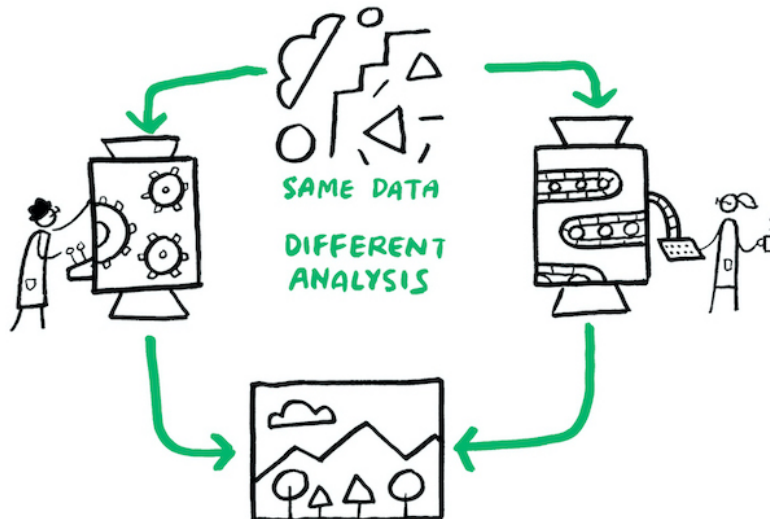
## REPRODUCIBLE



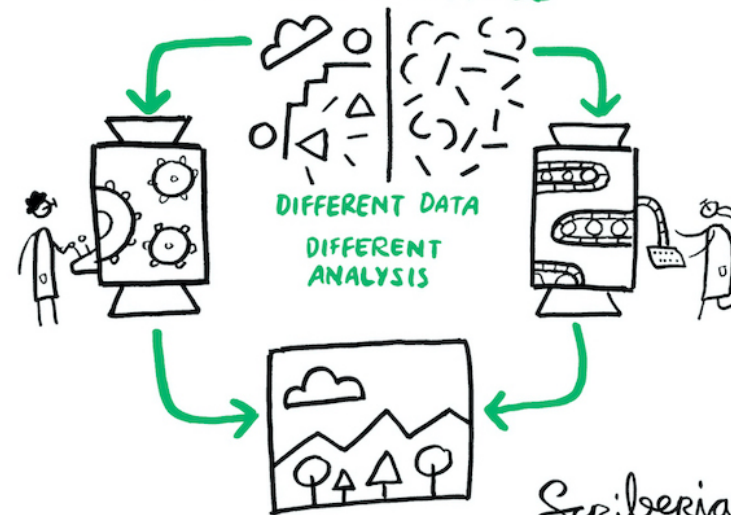
## REPLICABLE



## ROBUST



## GENERALISABLE



Scriberia 

# Software Engineering

## for research data analysis

- **Reproducibility**: automation, syntax
- **Transparency**: docs, tests, rev control
- **Modularity**: packages and functions
- **Generalisability**: compartmentalise quirks
  - multiple-select questions
  - non-standard user-defined missing
  - erroneous coding of responses
  - reordering and reverse coding of response levels