
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.

That is, a second researcher might use the same raw **data** to build the same analysis **files** and implement the same statistical **analysis** in an attempt to yield the same results.

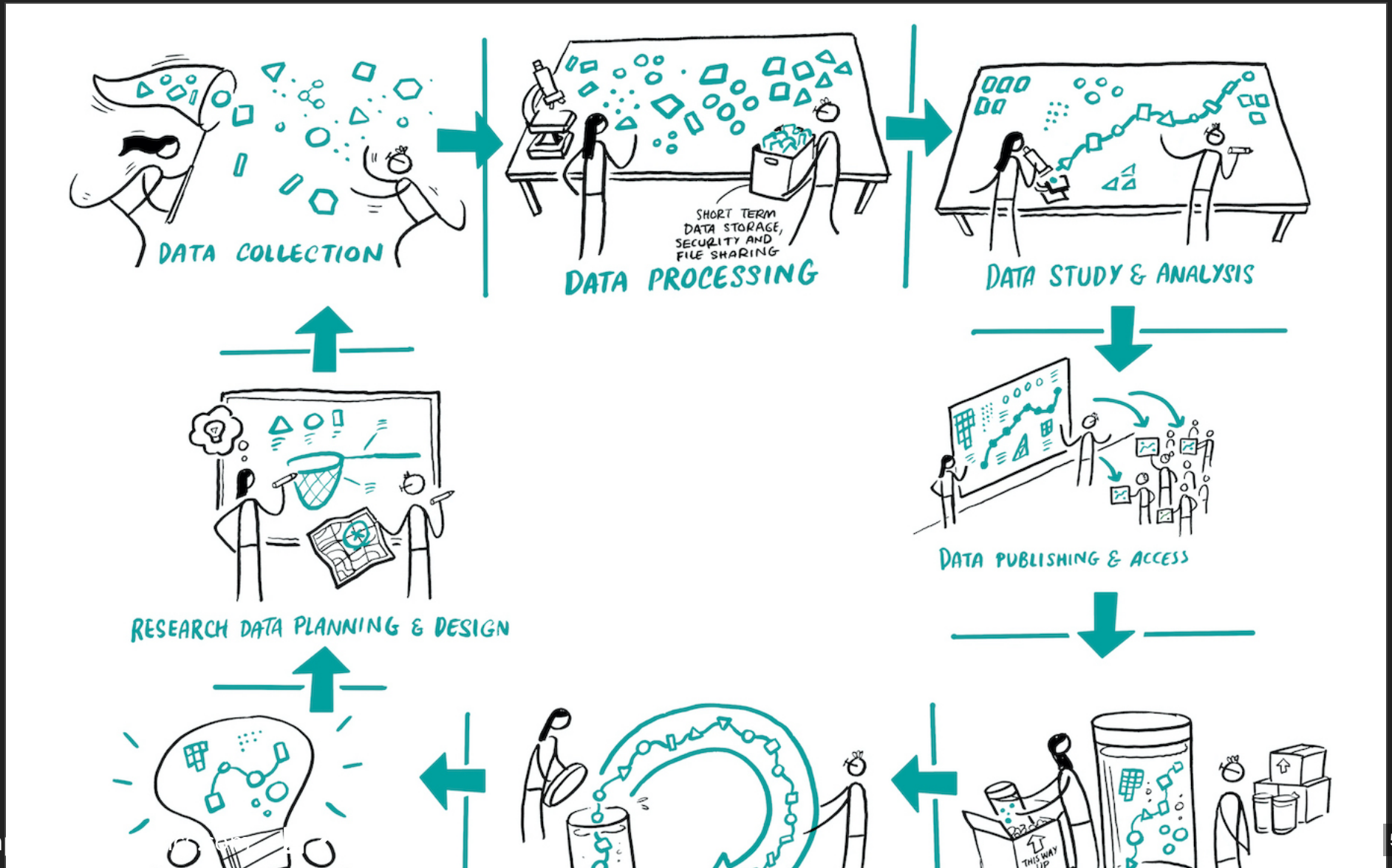
Reproducibility is a **minimum** necessary condition for a finding to be **believable** and

Reproducibility Crisis in Research

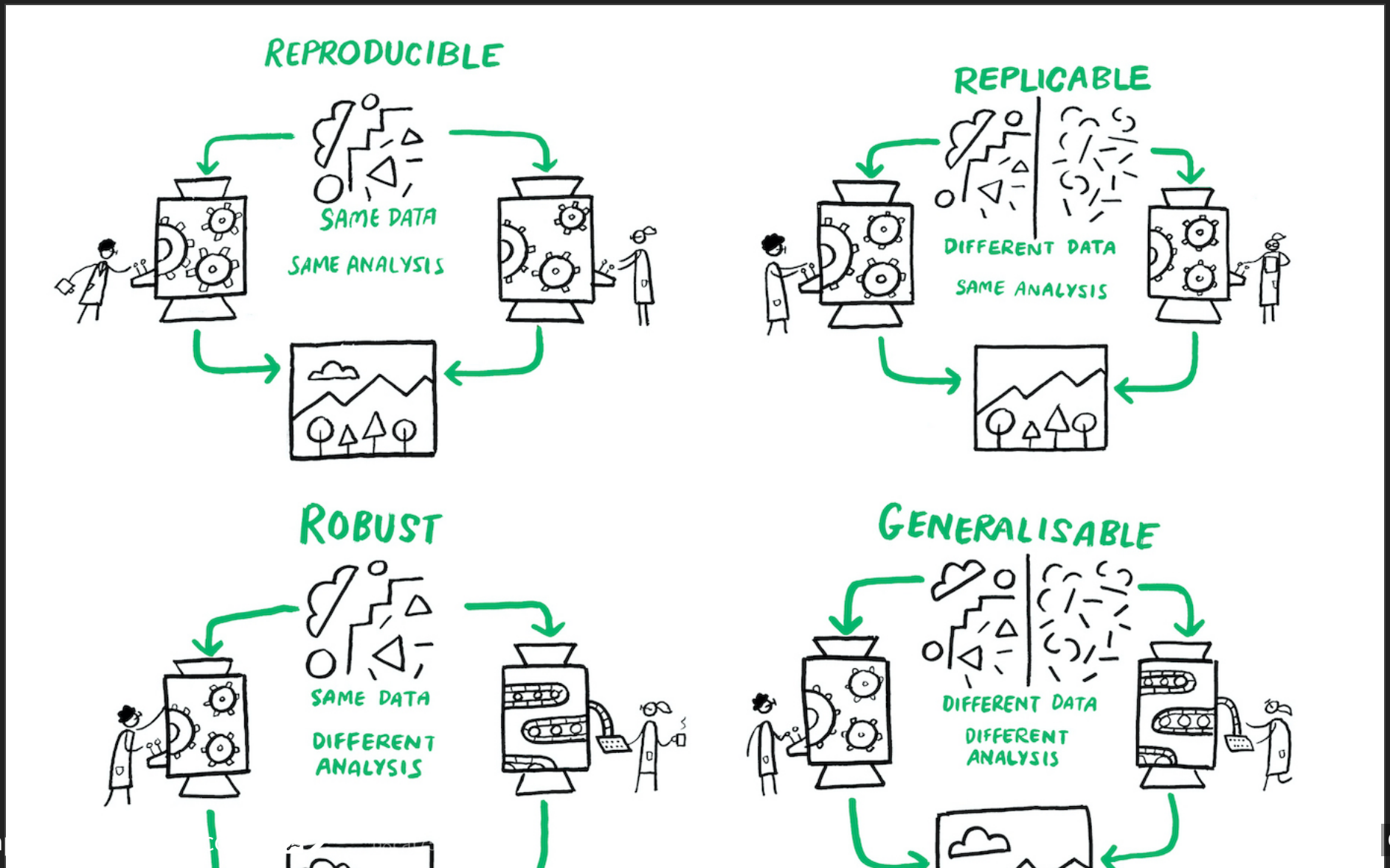
Technology Life Cycle

Technology life cycle

Research Cycle



Beyond Reproducibility



Software Engineering Principles

for research data analysis

- **Reproducibility**: automation, syntax
- **Transparency**: docs, test cases, version control
- **Modularity**: packages and functions
- **Generalisability**: compartmentalisation of dataset-specific quirks
  - multiple-select questions