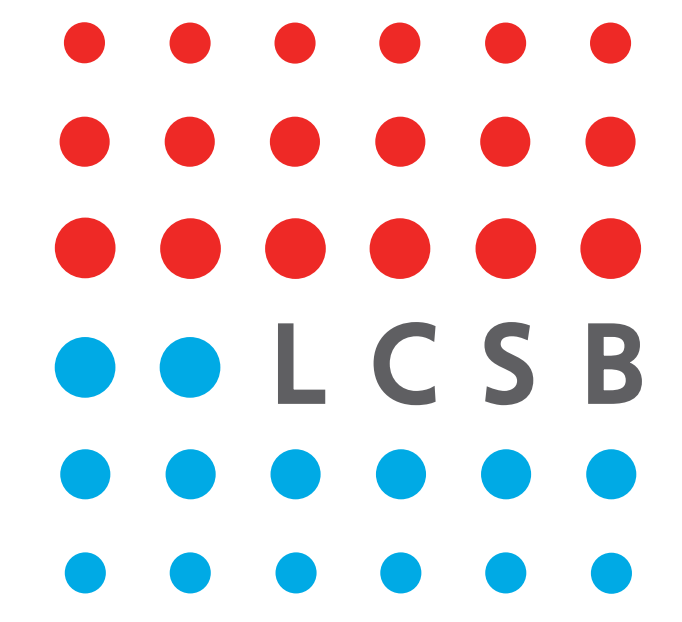


# REPRODUCIBLE RESEARCH RESULTS - R<sup>3</sup>

Luxembourg Centre for Systems Biomedicine



[Christophe Trefois](#)<sup>1</sup>, [Yohan Jarosz](#)<sup>1</sup>, [Wei Gu](#)<sup>1</sup>, [Venkata Satagopam](#)<sup>1</sup>, [Sergio Coronado](#)<sup>1</sup>, [Reinhard Schneider](#)<sup>1</sup>

<sup>1</sup>Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg (UL)

## Introduction

The Luxembourg Centre for Systems Biomedicine (LCSB) has a flagship research area around Parkinson's disease (PD). PD is a multi-factorial disease for which the etiology still remains to be elucidated. By the nature of its projects, the LCSB has tightly **integrated** collaborations between research groups and **accurate, shareable reproducible** results are required in order to tackle and understand the complex disease mechanisms. In the last decade, this need for increase of standardization has been rising in the whole research community.

With increasing **complexity** of generated data and associated analysis, there has been raising concerns in the scientific community about the reproducibility of findings and conclusions in many peer-reviewed scientific articles. In the past, experimental procedures and summary statistics were required in the publishing process. However, often, **workflows, computational environments** and **data sets** were not, leading to a global problem where many research findings cannot be reproduced based on the contents of a published article.

## Problem statement

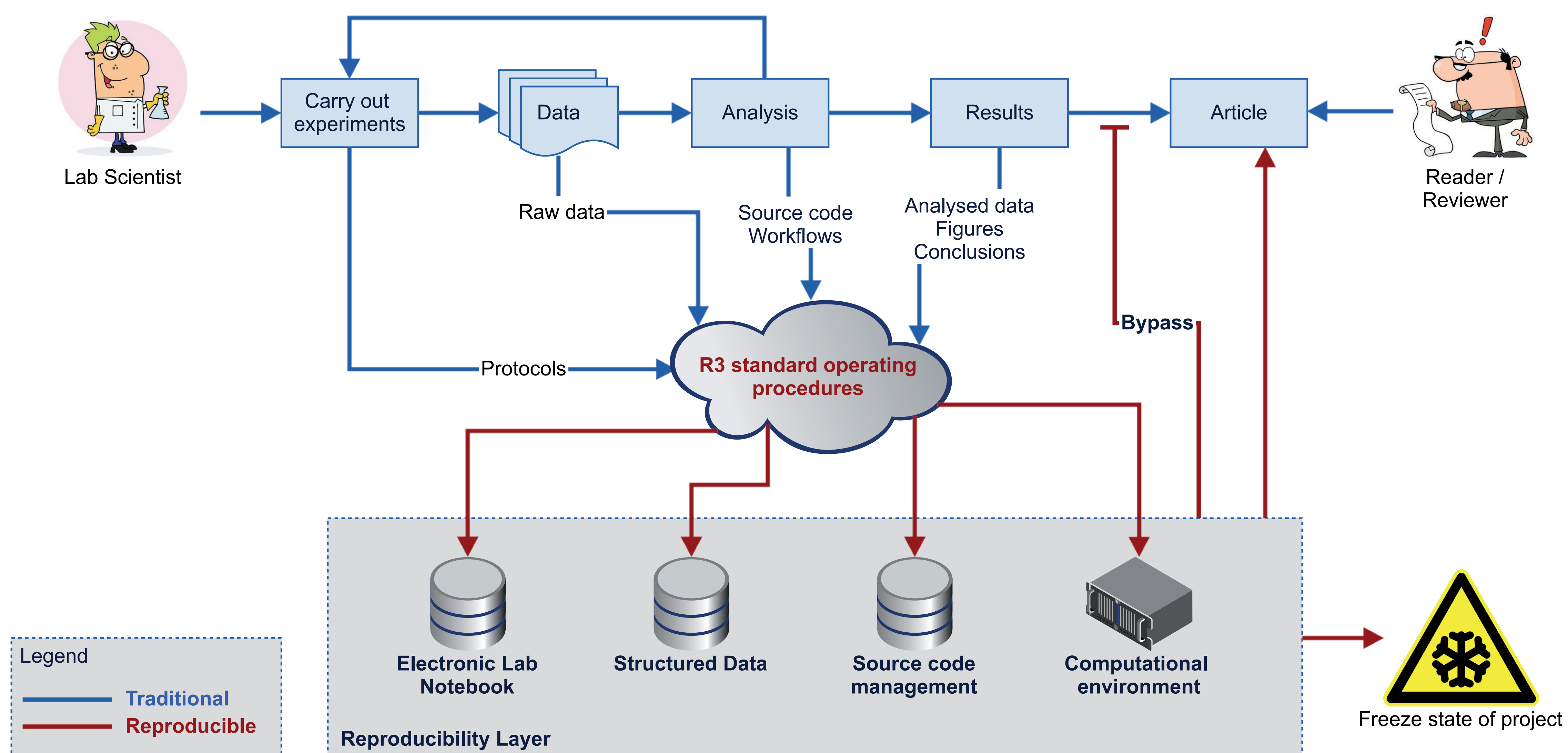
There is a constant increase of errors in published scientific output and it is largely not feasible to reproduce findings based on published material in peer-reviewed articles. We have identified a number of areas contributing to this decrease in scientific quality over time:

- Increasing **complexity** of research environment
- Lack of **availability** of data, algorithms, source code
- **Standardized operating procedures** for data handling & analytics
- Decreasing **persistence** of outdated technological tools

Due to these issues, there is an inherent need to **harmonize** IT infrastructures, data processes, tools, data handling and methods. The penultimate goal of these efforts is to **raise research quality** for the increasing complexity of research and increase overall reproducibility.

Therefore, we propose a **R<sup>3</sup> reproducibility layer** aiming to standardize all elements in the classical research publication workflow.

## Increase quality of research with the R<sup>3</sup> reproducibility layer



## Details of the R<sup>3</sup> reproducibility layer

### 1. Electronic Lab Notebook

- Holds protocols and experimental methods
- Contents can be locked and are accessible everywhere
- Contains metadata of experiments

### 2. Structured Data

- **CIA** triad
  - **Confidentiality:** Secure and selective access
  - **Integrity:** Prevent data modifications
  - **Availability:** Long-term access to raw and processed data
- Computerized access to available data

### 3. Source code management

- Version control
- Ownership of generated code

### 4. Computational environment

- Create workflows / pipelines to standardize data analysis
- Regroup all dependencies, programs and computational environment in one self-contained package

### 5. Freeze state of project

- Protocols, data, source code and environment cannot be changed anymore
- Project state needs to remain accessible / runnable over time

## Conclusions

In this project, we aim to implement the R<sup>3</sup> reproducibility layer at the LCSB across the different research groups and define a framework to help different academic institutions in achieving a similar environment. Finally, there is:

- Increasing need for **reproducible computational research**
- Need to establish a **reproducibility layer** to cover all processes from experiment to article publishing
- **Reliable** and **heterogeneous infrastructure** has to be established

Ultimately, the goal of reproducible science is to provide the **reader** with all the information required to **accurately reproduce** scientific results.

