CREATIS | Medical Imaging Research Laboratory
www.creatis.insa-lyon.fr

UNIVERSITÉ DE LYON

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Inserm

cnrs

Lyon 1

UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE

PRIMES UNIVERSITE DE LYON

Virtual Imaging Platform

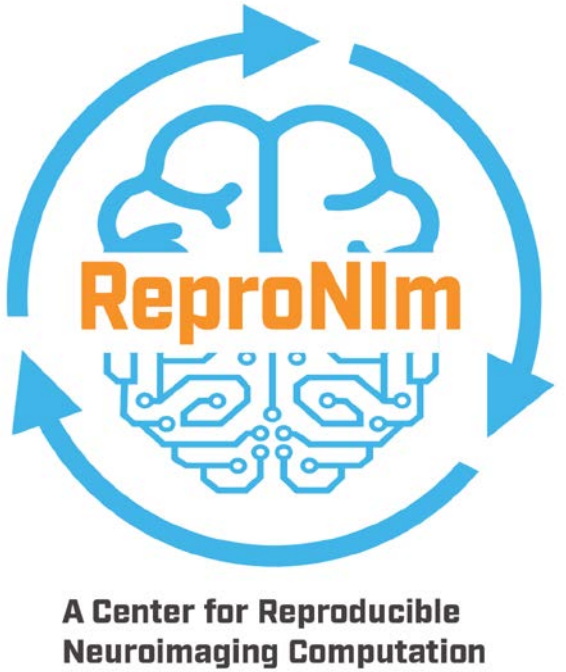
Reproducible Science

Reproducible Science

- Multiple terms, different meanings: Repeat, Reproduce, Reuse, Replicate…
  - https://www.frontiersin.org/articles/10.3389/fninf.2017.00069/full
  - Re-executability (publication-level replication): The exact same data, operated on by the exact same analysis should yield the exact same result (https://repronim.wordpress.com)
- Reproducibility crisis
  - https://www.biorxiv.org/content/10.1101/843193v1 – the analysis of a single functional neuroimaging dataset by 70 independent analysis teams revealed substantial variability in reported binary results
- A few tips for improving reproducibility
  - Document your work/practices both on data (protocols) and code (use versioning tools)
  - Publicly share data and analysis code, to enable others to run their own analysis with the same data or validate the code used
  - https://hal.archives-ouvertes.fr/hal-02144142v3

ReproNim

- Goal: improve the reproducibility of neuroimaging science
  - https://www.repronim.org/
- Development and delivery of software tools
  - data and software discovery
  - implementation of standardized description of data, results and workflows
  - …
- Training materials to facilitate and support re-executability practices
  - Excellent and multiple training materials



Examples of ReproNim Tools

- **NeuroDocker/ReproEnv** – Simplifying ***creating containers***
  - https://github.com/kaczmarj/neurodocker
  - A command-line program that generates custom Dockerfiles and Singularity recipes for neuroimaging and minifies existing containers
- **DataLad** – ***Version control*** and ***publication*** of ***data*** and ***containers***
  - https://www.datalad.org/
  - A data version control system; builds on top of git-annex and extends it with an intuitive command-line interface
- **ReproMan (& DataLad)** – Managing ***running of containers*** locally or on distributed resources
  - https://github.com/ReproNim/reproman
  - Simplify creation and management of computing environments in Neuroimaging
- **TestKraken** – ***Continuous integration*** and ***stability*** assessment
  - https://github.com/ReproNim/testkraken
  - test workflows in a matrix of parametrized environments.

How can I use it for my research?

- Complex problem, many tools, little available time
- So what can we do about it
  - Train
  - Decide what is important and feasible depending on each project
  - Maybe VIP can help ?...

Boutiques and FAIR data analysis

- Describe, publish, integrate and execute applications **across platforms**
  - facilitate application porting
  - import and exchange of applications
  - Linux containers to facilitate application installation and sharing

- https://github.com/boutiques

Findable

1. Globally persistent records
2. Described with rich metadata
3. Searchable

We leverage **Zenodo [2]** to create DOIs for Boutiques descriptors which can be accessed via the Zenodo API.

Accessible

1. Easily retrievable
2. Universal access
3. Persistent metadata beyond data lifetime

The retrievable tool descriptions contain **immutable** human- and machine-readable instructions for testing and launching each tool.

Interoperable

1. Formalized and shared metadata standard
2. Metadata standards adopted are FAIR
3. Linking between objects where appropriate

CARMIN [3] and **Boutiques [4]** standards are used to describe and launch tools, either locally or through a RESTful API.

Re-Usable

1. Multiple accurate and relevant attributes
2. Clearly licensed
3. Meets minimum domain standards

Docker [5] and **Singularity [6]** virtualization enable re-runability across platforms and enclosed testing. Simulation and querying allow runtime evaluation.

FAIR tools. Credits: Gregory Kiar and Tristan Glatard

Project proposal on reproducibility

- Help **evaluate the numerical reproducibility of results** through
  - Continuous integration (active tests) of scientific applications
  - Provenance (passive tests), provided that we dispose of enough data

- Foster collaboration and entice the use of multiple analysis pipelines by multiple researchers by providing them with a **complete solution easily accessible**

THANK YOU FOR YOUR ATTENTION!

QUESTIONS ?

sorina.pop@creatis.insa-lyon.fr