



HAL
open science

Increasing computational reproducibility or how-to share **your** research code

Camille Maumet

► **To cite this version:**

Camille Maumet. Increasing computational reproducibility or how-to share **your** research code. MRI Together 2024 - Global workshop on open, reproducible, and inclusive MR, Dec 2024, Online, Australia. hal-04814565v2

HAL Id: hal-04814565

<https://inria.hal.science/hal-04814565v2>

Submitted on 2 Dec 2024

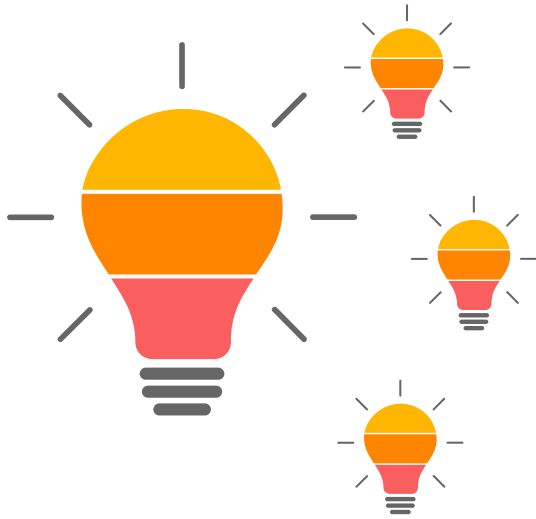
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

MRI Together
December 5, 2024



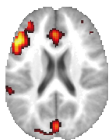
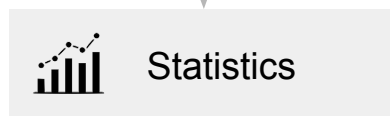
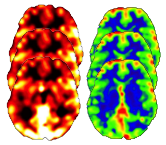
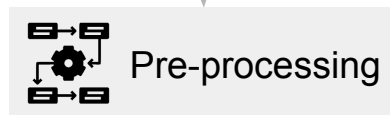
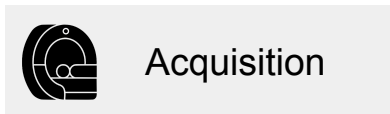
Increasing computational reproducibility or how-to share *your* research code

Camille Maumet

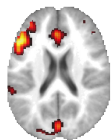
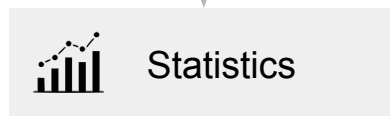
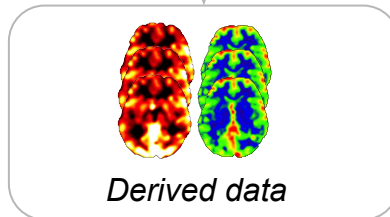
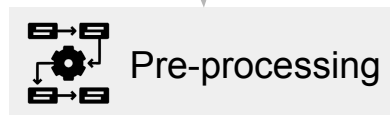
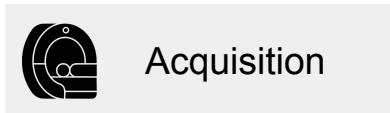
Inria, Univ Rennes, Inserm, CNRS, IRISA



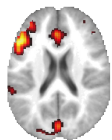
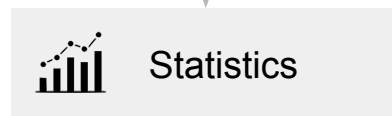
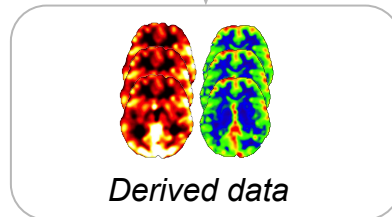
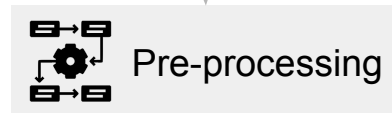
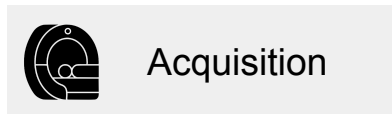
An **MRI** study



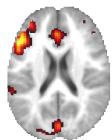
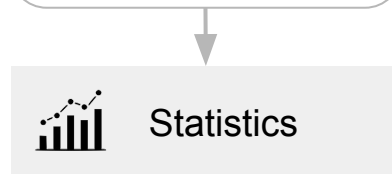
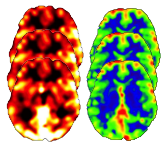
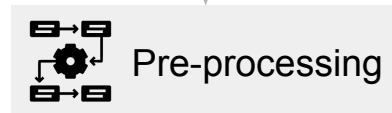
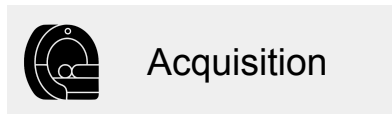
An **MRI** study



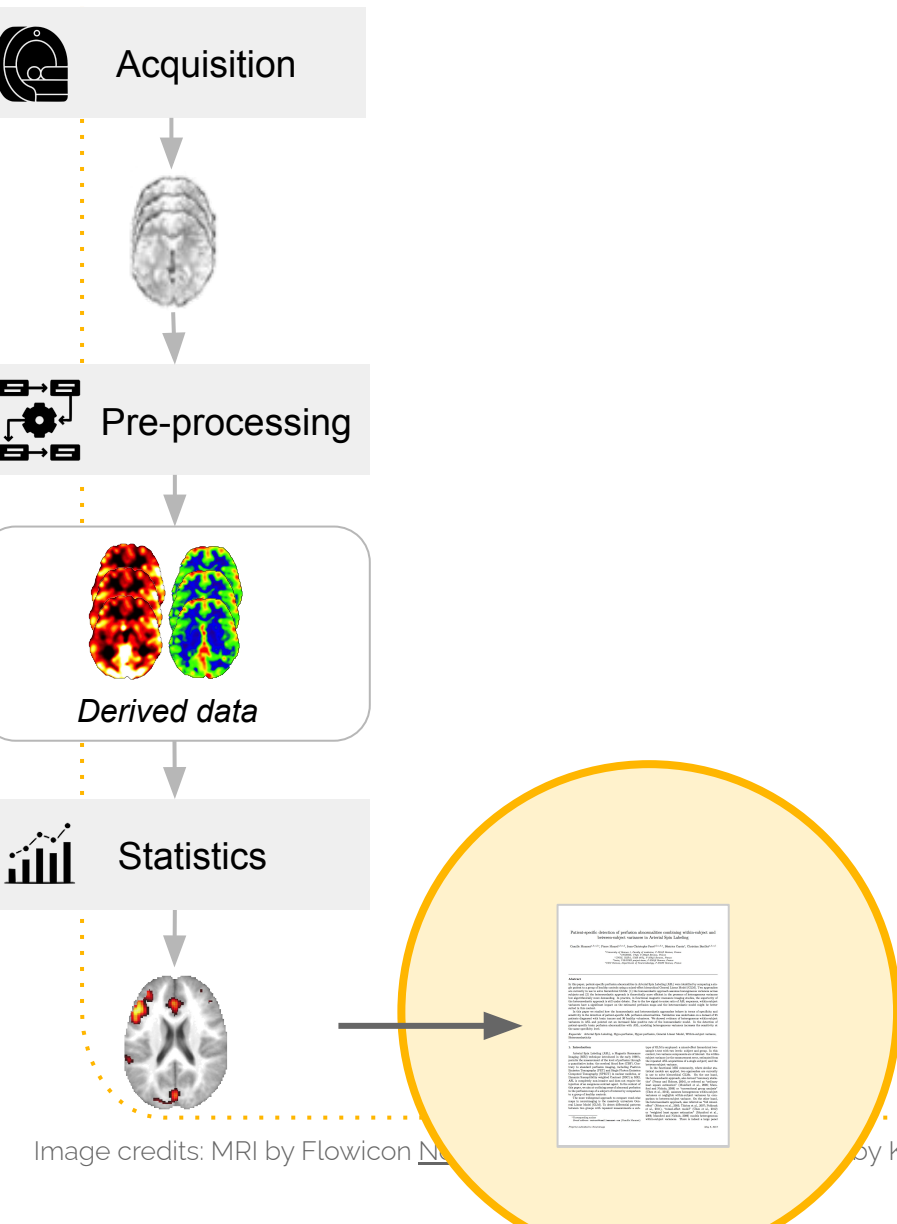
An **MRI** study



An **MRI** study



An **MRI** study

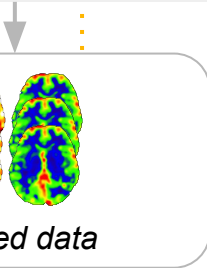


An **MRI** study

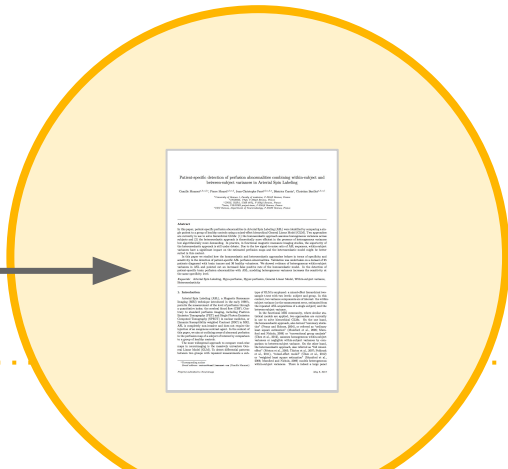
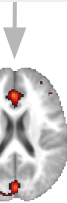
Acquisition



Pre-processing



Statistics



An **MRI** study



An **MRI** study



An **MRI** study



Which version of the SPM software was used?

Which registration algorithm?

How was the brain mask computed?

An **MRI** study



Which version of the SPM software was used?

Which registration algorithm?

How was the brain mask computed?

Humm... **I can't reproduce** the published results...

Did I understand correctly the method that was used?

... is this due to my (different) data?

... or is this due to my (different) computing resources?

... or even something else?

An **MRI** study



Humm... **I can't reproduce** the published results...



Which version of the SPM software was used?

Which registration algorithm?

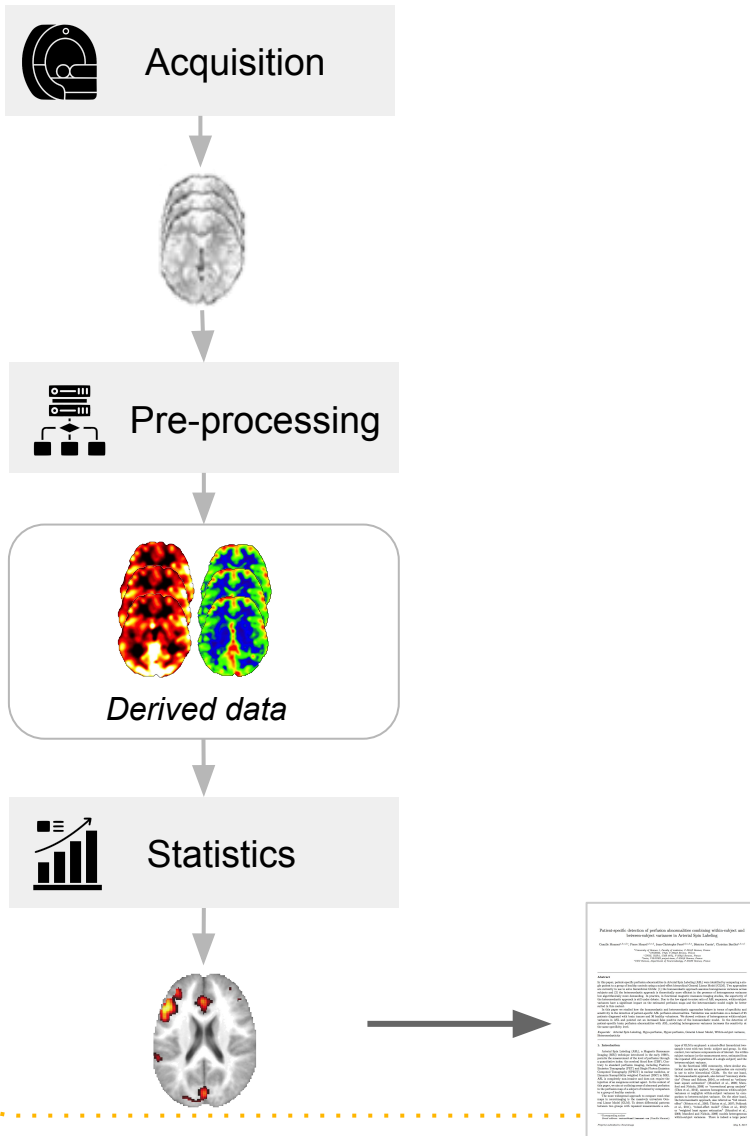
How was the brain mask computed?

... is this due to my (different) data?

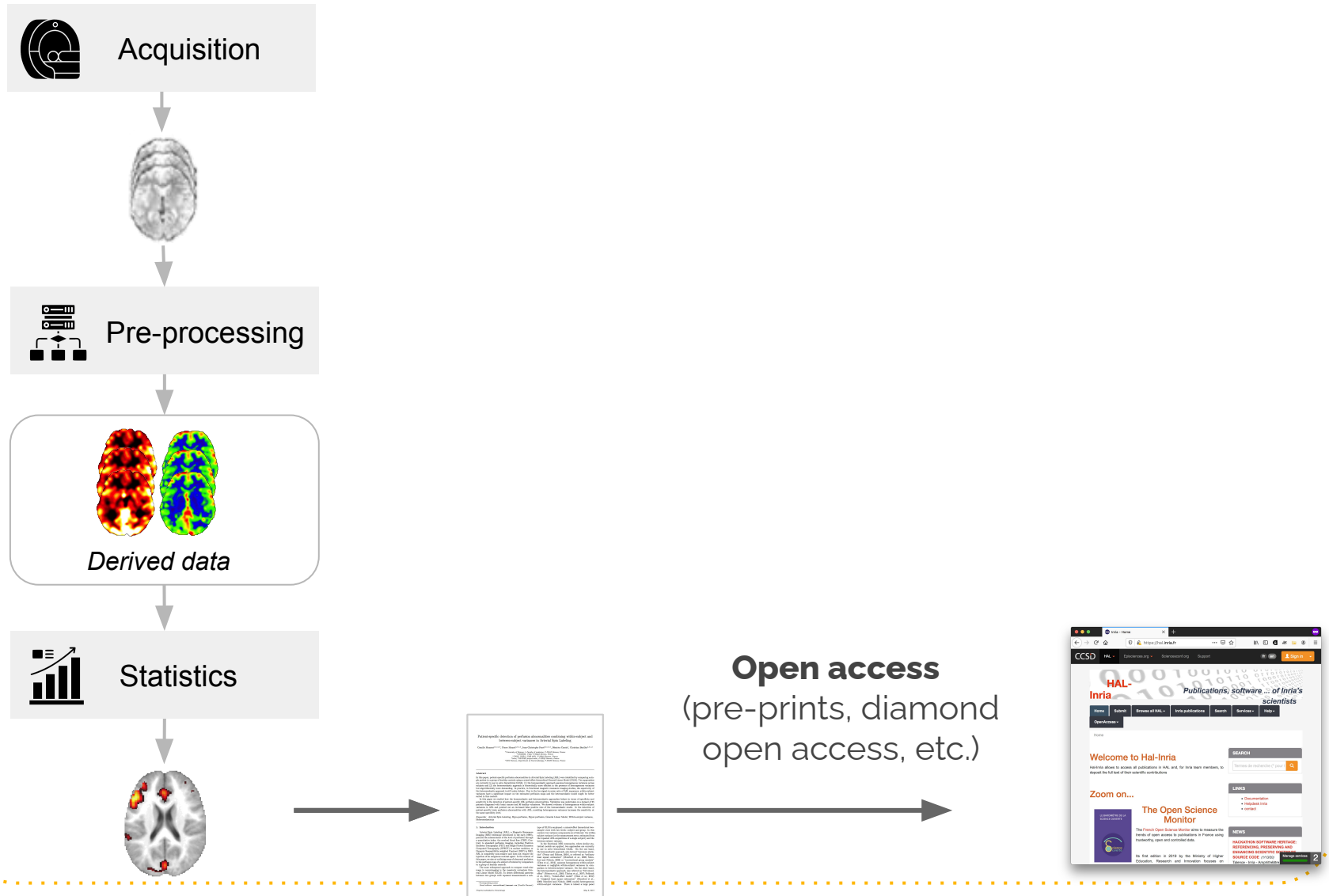
... is this due to (different) computing resources?

... or even something else?

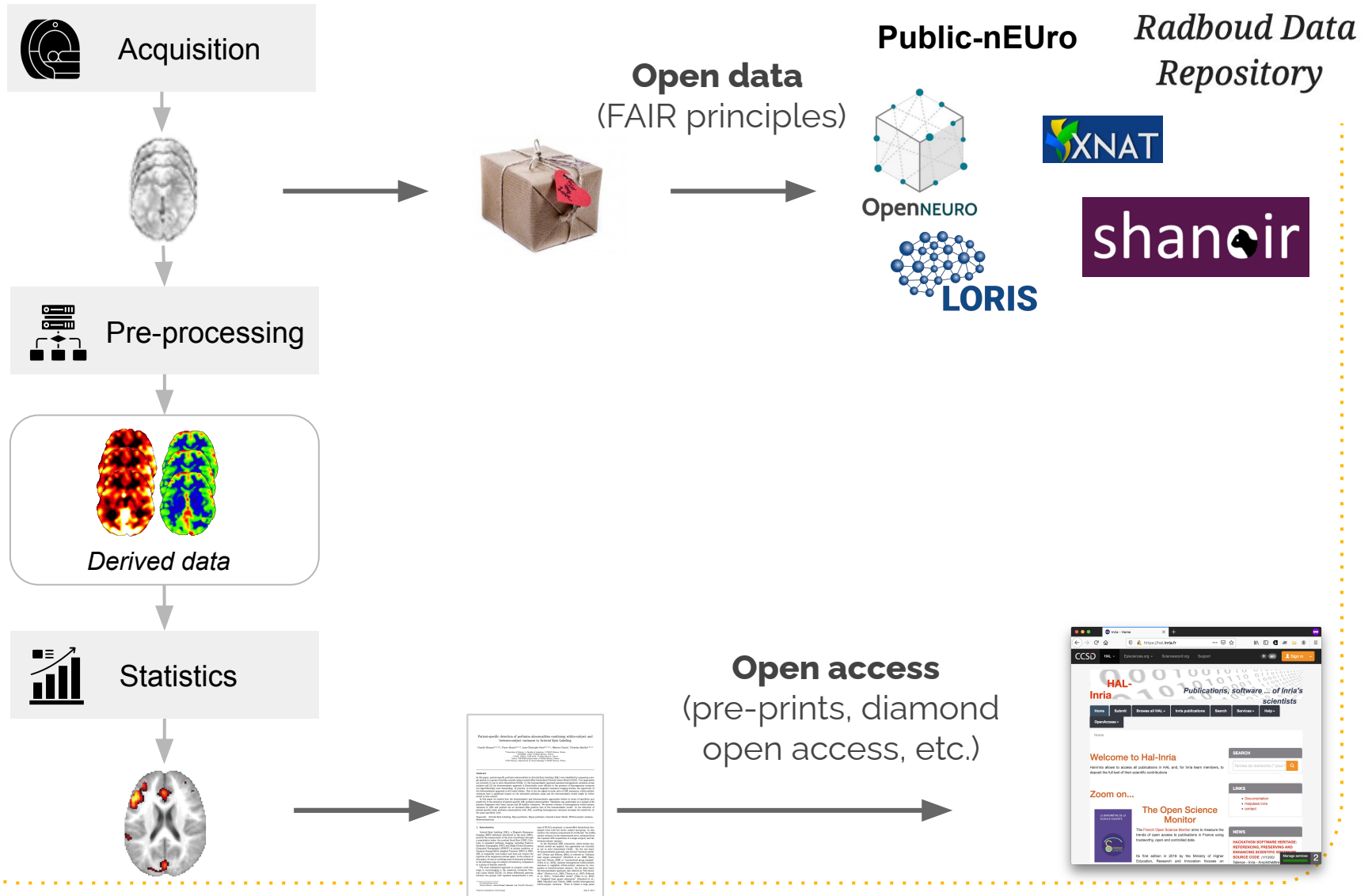
Towards **reproducible research**



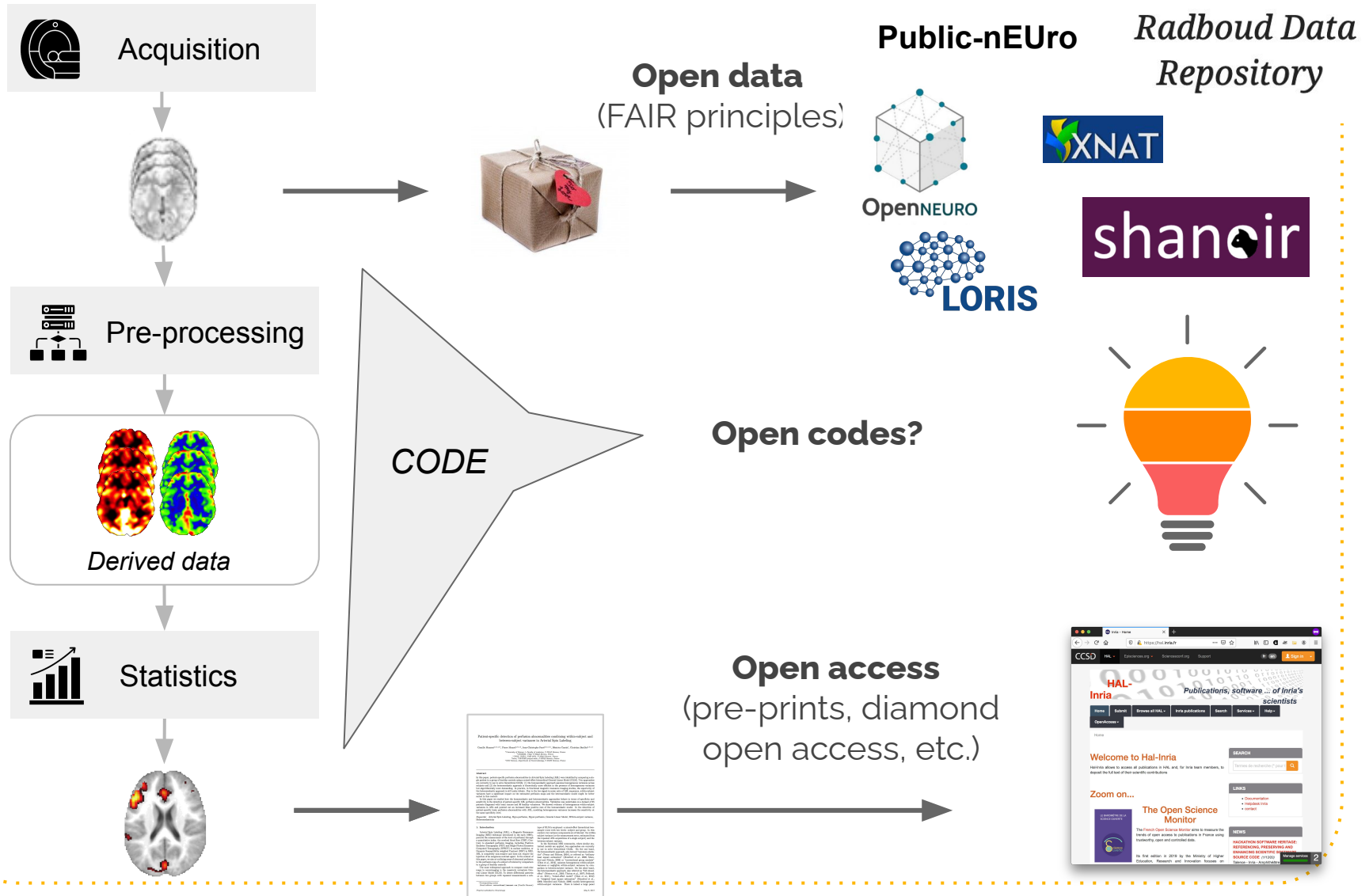
Towards **reproducible research**



Towards reproducible research



Towards reproducible research



Towards **reproducible research**

Computational reproducibility

Ability to reproduce **exactly the same results** using the **same data** and **same code**.

Replicability

Ability to reach the **same scientific findings** using **similar data** and a **similar method**.

Towards **reproducible research**

Computational reproducibility

Ability to reproduce **exactly the same results** using the **same data** and **same code**.

Replicability

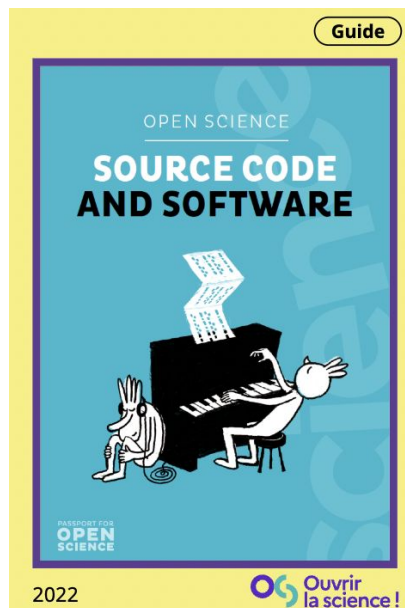
Ability to reach the **same scientific findings** using **similar data** and a **similar method**.

Open code



Open source
Transparency

- With an open **source licence** and **publicly accessible**.



<https://www.ouvrirlascience.fr/source-code-and-software/>

Open code

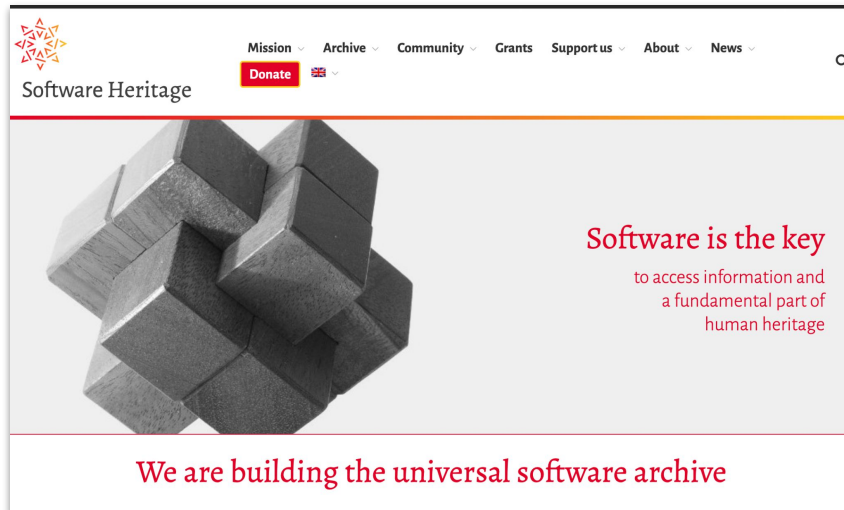


Archived
Sustainability

- **Archived** for **long-term preservation** (e.g. using Software Heritage, Zenodo)



Software Heritage



<https://www.softwareheritage.org/>

zenodo

<https://zenodo.org/>

Open code



Using coding best practices

Reutilisability

- Following **software development best practices**
 - **Versioning**: using a forge (Github, gitlab, etc.).
 - **Listing dependencies**.
 - **README**, file tree template, etc.

```
.
|-- CITATION
|-- README
|-- LICENSE
|-- requirements.txt
|-- data
|   -- birds_count_table.csv
|-- doc
|   -- notebook.md
|   -- manuscript.md
|   -- changelog.txt
|-- results
|   -- summarized_results.csv
|-- src
|   -- sightings_analysis.py
|   -- runall.py
```

[Wilson et. al, PlosCompBio 2017]

Project/Paper Title

Table of contents

- [How to cite?](#)
- [Contents overview](#)
- [Reproducing figures and tables](#)
 - [Table 1](#)
 - [Fig. 1](#)
 - [Fig. 2](#)
- [Reproducing full analysis](#)

How to cite?

See [CITATION](#).

Contents overview

<Summarise what's in this repository>

Reproducing figures and tables

<Instructions on how to use summary/derived data in the `results` directory to create figures and tables>

<Specify precise steps, including any datasets that need to be downloaded and path variables that need to be set>

Table 1

Fig. 1

Fig. 2

Reproducing full analysis

<Instructions on how to (1) obtain raw data; (2) process it to create summary/derived data in the `results`>

<Specify precise steps, including any datasets that need to be downloaded and path variables that need to be set>

Example of [README template](#)

How-to **open** your **code**



Open source
Transparency

- with an open **source licence** and **publicly accessible**.

Archived
Sustainability

- in an **archive** for **long-term preservation** (e.g. using Software Heritage, Zenodo)

Using coding best practices
Reutilisability

- following **software coding best practices**
 - **Versioning** using a forge (Github, GitLab, etc.)
 - **Listing dependencies**
 - **README**, file tree template, etc.

Are my results (computationally) reproducible now? **Well...**

Beyond source code,
environnements (OS, libraries, etc.)
can also have an impact...

→ *docker, Guix, Nix*

Computational reproducibility **in 10 years** is still an open question...

Perkel, Jeffrey M. "Challenge to scientists: does your ten-year-old code still run?." *Nature* 584.7822 (2020): 656-659.

Towards **reproducible research**

Computational reproducibility

Ability to reproduce **exactly the same results** using the **same data** and **same code**.

Replicability

Ability to reach the **same scientific findings** using **similar data** and a **similar method**.

Towards **reproducible research**

Computational reproducibility

Ability to reproduce **exactly the same results** using the **same data** and **same code**.

Replicability

Ability to reach the **same scientific findings** using **similar data** and a **similar method**.

Towards **reproducible research**

Computational reproducibility

Ability to reproduce **exactly the same results** using the **same data** and **same code**.

Replicability

Ability to reach the **same scientific findings** using **similar data** and a **similar method**.

Reusability

Ability to reuse some research **artefacts for new purposes**

Towards **reproducible research**

Computational reproducibility

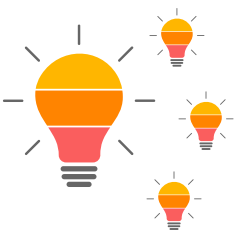
Ability to reproduce **exactly the same results** using the **same data** and **same code**.

Replicability

Ability to reach the **same scientific findings** using **similar data** and a **similar method**.

Reusability

Ability to reuse some research **artefacts for new purposes**

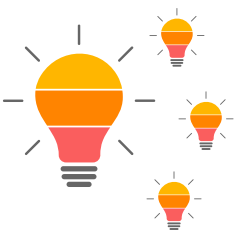


Open software



Open source

- Choose an **open source license**
- Make **code available for download**
- Use an **archive** for sustainability



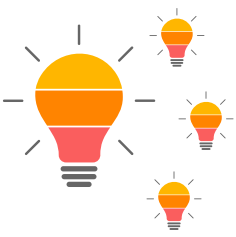
Open software

Who can **access**
and use the
software?



Open source

- Choose an **open source license**
- Make **code available for download**
- Use an **archive** for sustainability



Open software

Who can **access and use** the software?



Open source

- Choose an **open source license**
- Make **code available for download**
- Use an **archive** for sustainability

Who can **contribute**?



Open development

- Accept **external contributions** (e.g. Github PR)
- Be transparent in terms of **development choices** (e.g. issues, project board)

Open software



Who can **access and use** the software?



Open source

- Choose an **open source license**
- Make **code available for download**
- Use an **archive** for sustainability

Who can **contribute**?



Open development

- Accept **external contributions** (e.g. Github PR)
- Be transparent in terms of **development choices** (e.g. issues, project board)

Who can **decide** for **future directions**?



Open governance

- Transparency in how decision are made
- Possibility to join the **leadership of the project**

Towards a more **collaborative** science

Beyond technical solutions, open science brings a new culture

- Sharing resources towards the **scientific community**
- **Reusing existing resources** (code, data etc.) generated by others

toward a more collaborative research.



Towards a more **collaborative** science

Beyond technical solutions, open science brings a new culture

- Sharing resources towards the **scientific community**
- **Reusing existing resources** (code, data etc.) generated by others

toward a more collaborative research.



Open science awards for
open software / open data /
open PhD thesis

<https://www.ouvrirlascience.fr/category/openscience-prices/>



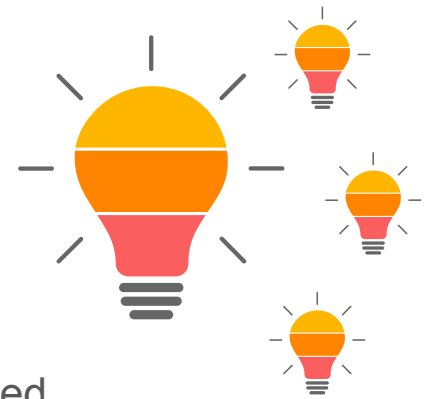
MRI Together
December 5, 2024



Increasing computational reproducibility or how-to share *your* research code

Thank you

Camille Maumet



Credit: Presentation template and icons by SlidesCarnival, adapted



@cmaumet.bsky.social

@cmaumet@fediscience.org