

Journée INS2I Santé et IA - 23 Janvier 2020

Quand les données s'ouvrent : Opportunités et nouveaux défis pour mieux comprendre notre cerveau

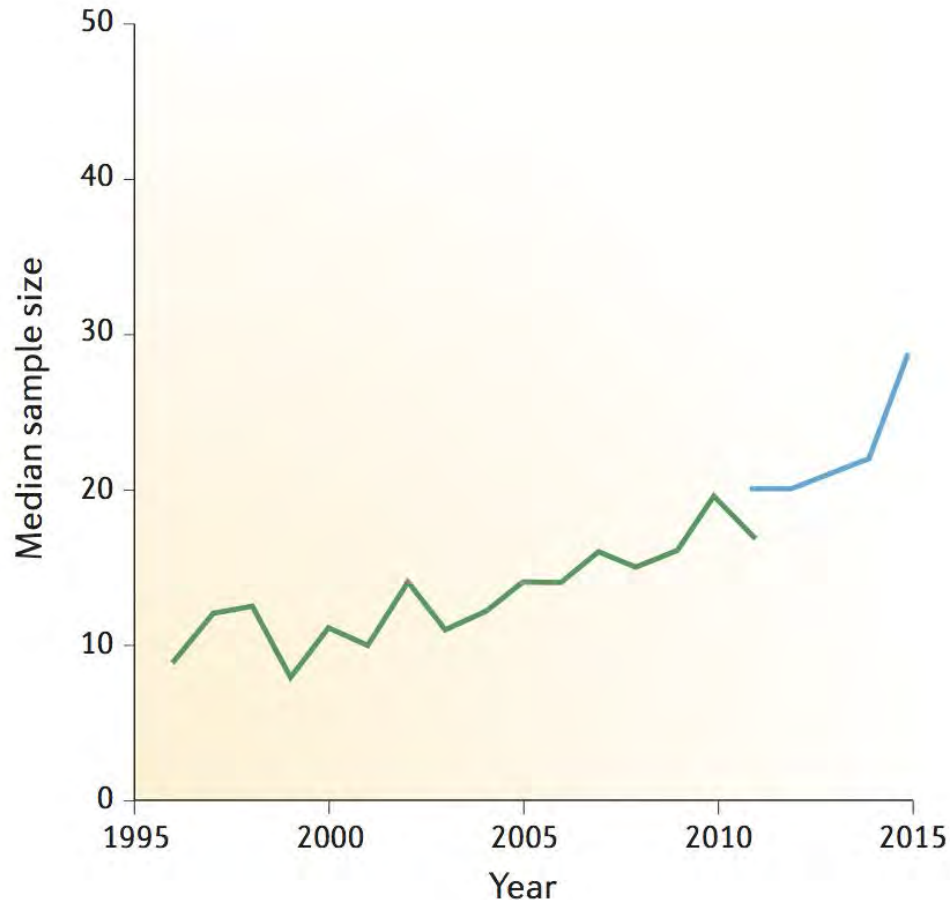
Camille Maumet

Univ Rennes, Inria, CNRS, Inserm
IRISA UMR 6074, Empenn ERL U-1228



La neuroimagerie aujourd'hui

Tailles des études de neuroimagerie



2015 : médiane de 30 participants par étude

[Poldrack et. al, Nature Neuroscience 2017]

Une étude démontre les biais de la reconnaissance faciale, plus efficace sur les hommes blancs

Lorsqu'il s'agit de reconnaître le genre d'un homme blanc, des logiciels affichent un taux de réussite de 99 %. La tâche se complique lorsque la peau d'une personne est plus foncée, ou s'il s'agit d'une femme.

LE MONDE | 12.02.2018 à 14h52 • Mis à jour le 13.02.2018 à 18h18 |

Par Perrine Signoret

Représentativité de l'espace d'entraînement

Why are middle-aged marathon runners faster than twentysomethings?

According to new data from the running app Strava, runners in their 40s are streets ahead of younger rivals



▲ 'Middle-aged runners outperform runners in the 20s ...' Photograph: FatCamera/Getty

According to data released by the running app Strava, middle-aged runners consistently average faster marathon times than their younger rivals, apparently defying the usual rules of athletic performance. Men in the 40-49 age bracket clock an average time of four hours and 17 minutes for a marathon, according to the recent figures. Women in the same age range typically come in at just under the five-hour mark.

Plus rapides à 40 ans
qu'à 20 ans ?

Biais de sélection

SCIENCE

A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



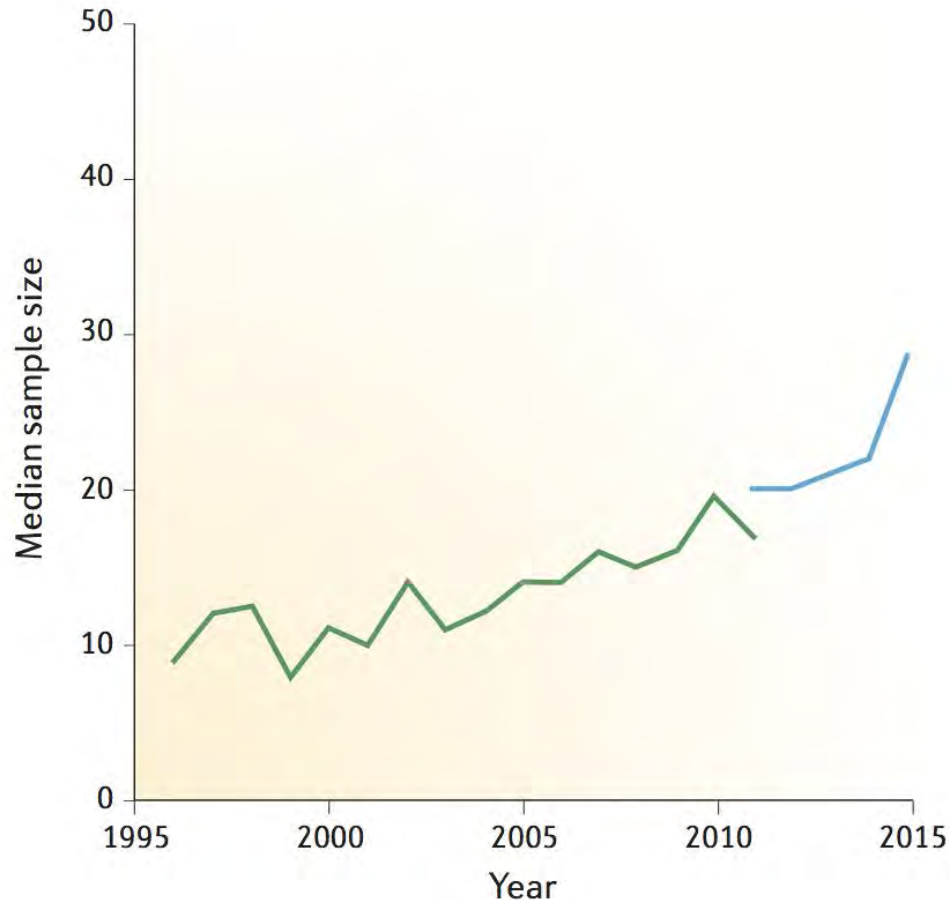
SEAN NEL / SHUTTERSTOCK

In 1996, [a group of European researchers](#) found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

It was a blockbuster discovery at the time. The team found that a [less active version](#) of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

Effet de vibration des résultats

Tailles des études de neuroimagerie



2015 : médiane de 30 participants par étude

Représentativité
Biais de sélection
Effet de vibration

[Poldrack et. al, Nature Neuroscience 2017]

**Quand les données
s'ouvrent...**

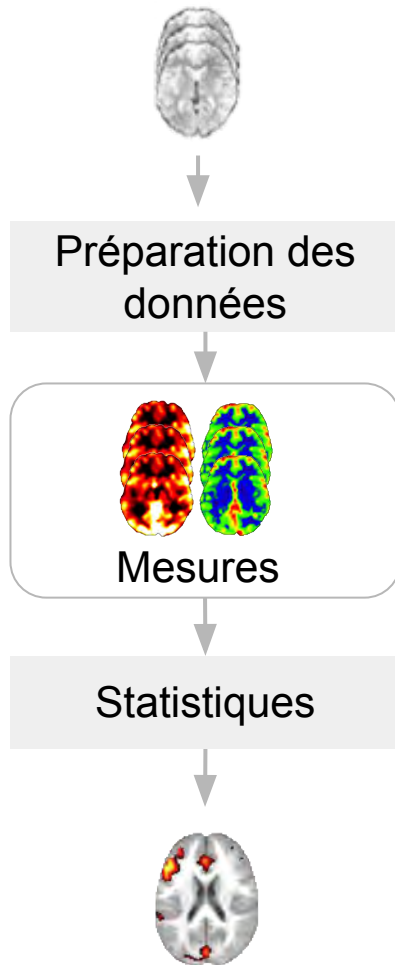
Les données ouvertes



“Les **données ouvertes** ou ***open data*** sont des données numériques dont l'accès et l'usage sont laissés libres aux usagers.”

[Wikipedia](#)

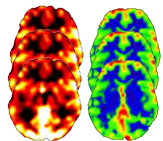
Ouvrir les données d'une étude



Ouvrir les données d'une étude



Préparation des données



Mesures



Statistiques

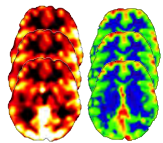


Communauté scientifique

Ouvrir les données d'une étude



Préparation des données



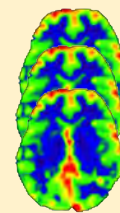
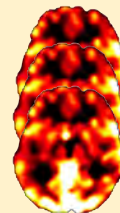
Mesures



Statistiques



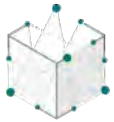
Communauté scientifique



Données ouvertes



Étude unique
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T Q R V
S M Y C O N N E C T O M E Q
G S P K K W A R R G K E H R

 NITRC

 OSF

Créer un consortium

Site 1



Site 2



Site n



Créer un consortium

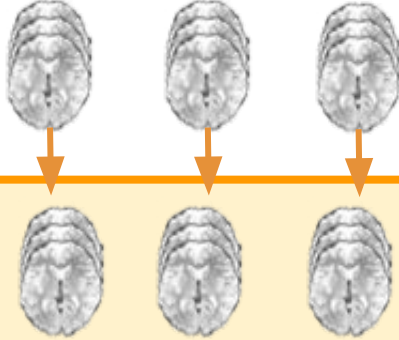
Site 1 Site 2 Site n



Communauté
scientifique

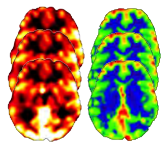
Créer un consortium

Site 1 Site 2 Site n



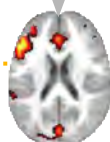
Communauté scientifique

Préparation des données



Mesures dérivées

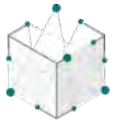
Statistiques



Données ouvertes



Étude unique
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T Q R V
S M Y C O N N E C T O M E Q
G S P K K W A R R G K E H R

Consortium
1000 participants



1000 Functional
Connectomes Project

NITRC



OSF

CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY

Créer une ressource commune

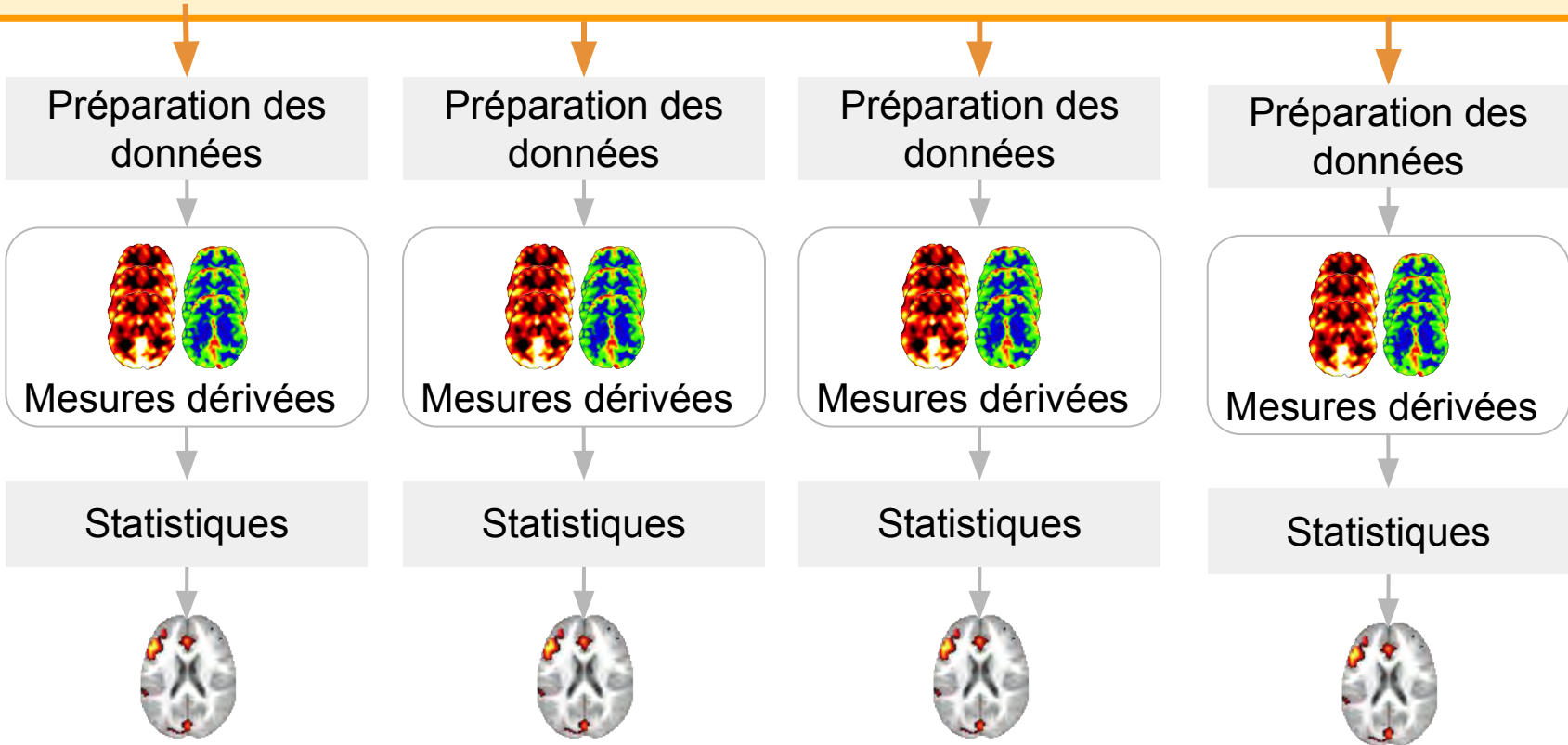


 Communauté
scientifique

Créer une ressource commune



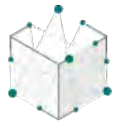
 Communauté scientifique



Données ouvertes



Étude unique
30 participants



OpenNEURO

studyforrest.org



NEUROVAULT

L I E L N L F T Q K T O R V
S M Y C O N N E C T O M E Q
G S P K K W A R R G K E H R

Consortium
1000 participants



ABIDE
Autism Brain Imaging
Data Exchange

1000 Functional
Connectomes Project



CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY



Cohorte
100 000 participants



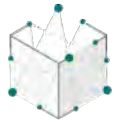
HUMAN
Connectome
PROJECT

biobank[™]
Promoting the health of future generations

Données ouvertes



Étude unique
30 participants



OpenNEURO

studyforrest.org

Consortium
1000 participants



ABIDE
Autism Brain Imaging
Data Exchange

1000 Functional
Connectomes Project



CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY



Cohorte
100 000 participants



biobank[™]
Promoting the health of future generations

- + Images
- + Homogène
- Exemplaires

De nouveaux défis

Défi 1 : Publier plus que nos articles !

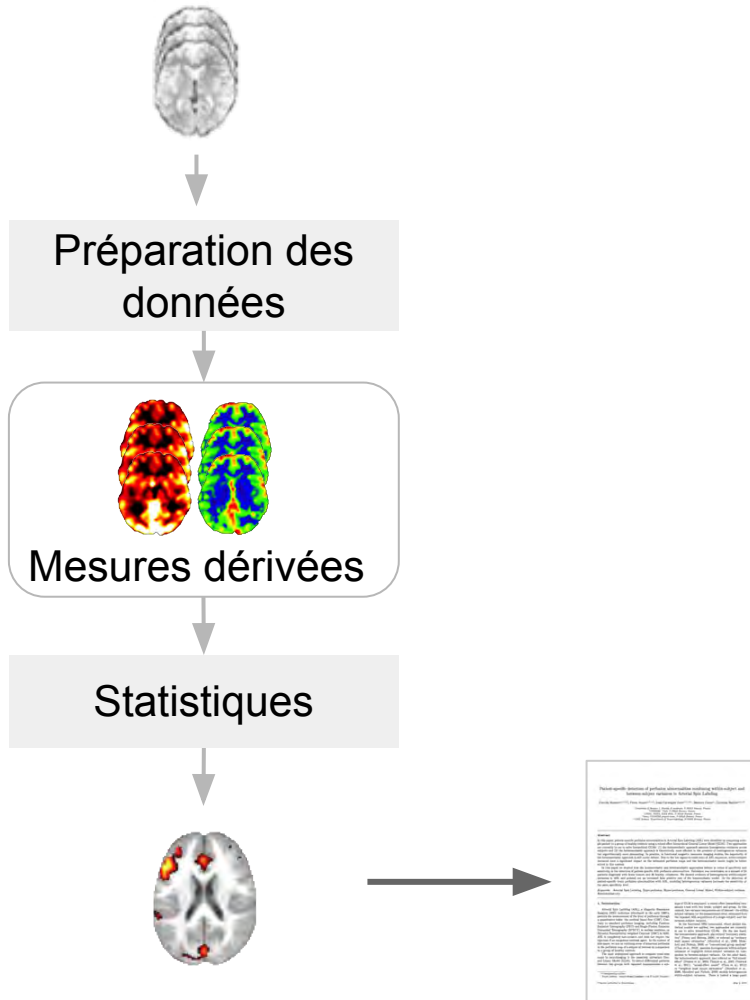


Image credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

HAL,
Pubmed

The screenshot shows the HAL website interface. At the top, it says "HAL - Inria Publications, software ... of Inria's scientists". Below that, there's a search bar and a "Welcome to Hal-Inria" message. A "Zoom on..." section highlights "The Open Science Monitor". At the bottom of the screenshot, there's a date: "In this version: 10/2019".

An orange padlock icon is positioned at the bottom left of the yellow box, symbolizing open access.

Défi 1 : Publier plus que nos articles !

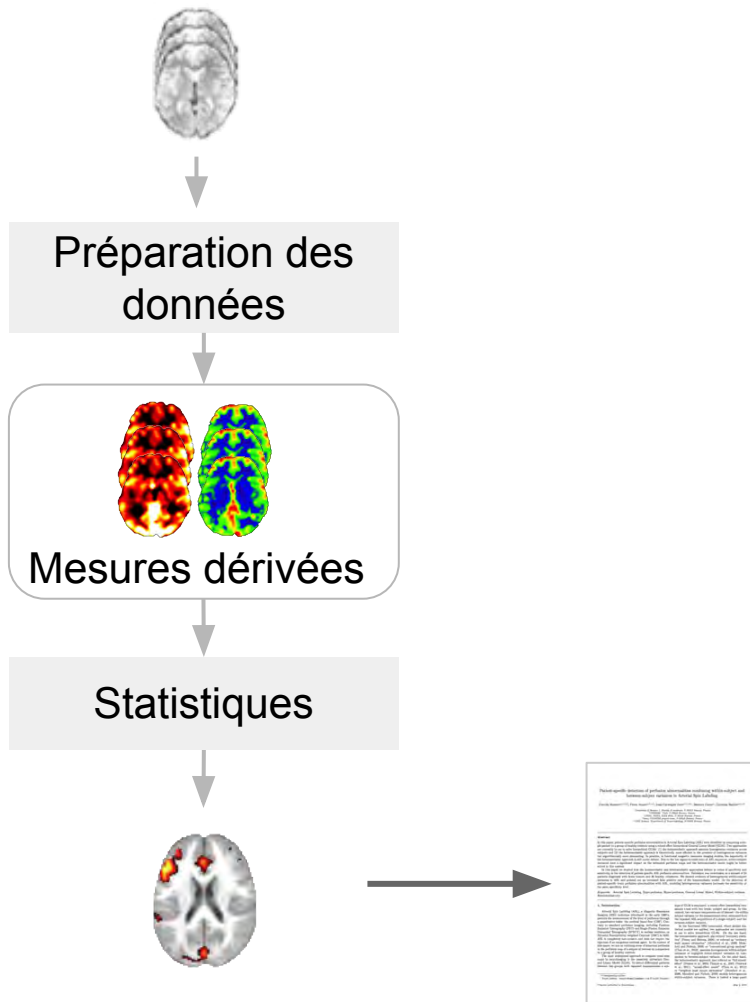
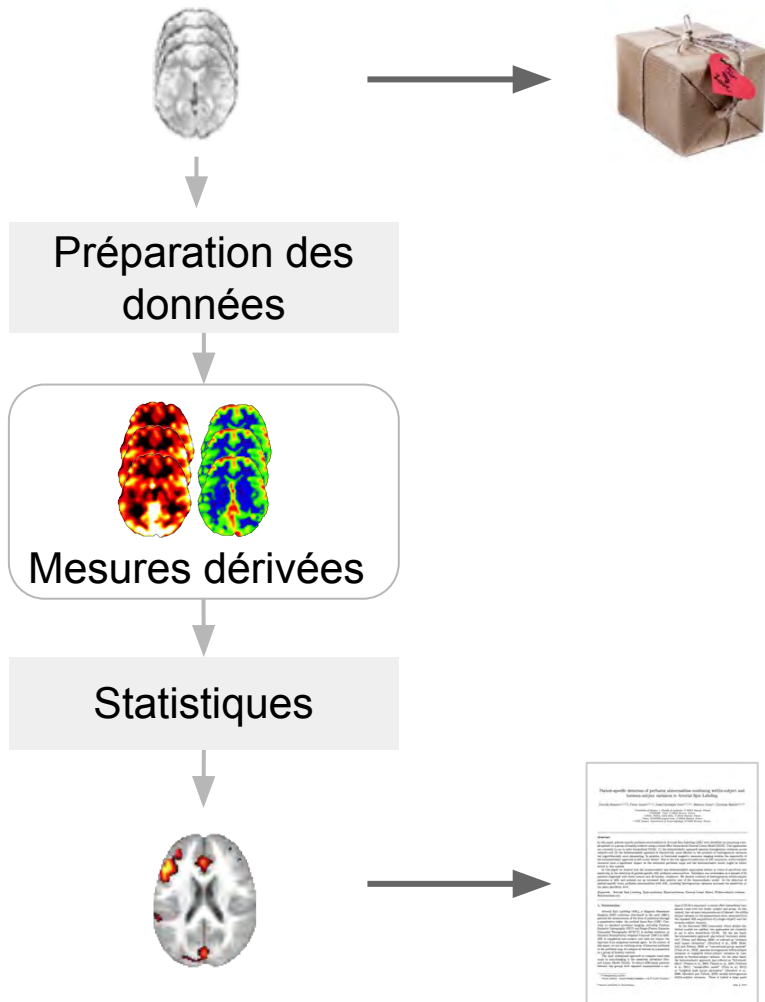


Image credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

This panel highlights various open science platforms and repositories:

- OpenfMRI
- OpenNEURO
- XNAT
- LORIS
- SHANNOIR
- HAL, Pubmed
- A screenshot of the HAL-Inria website.
- An orange padlock icon at the bottom.

Défi 1 : Publier plus que nos articles !



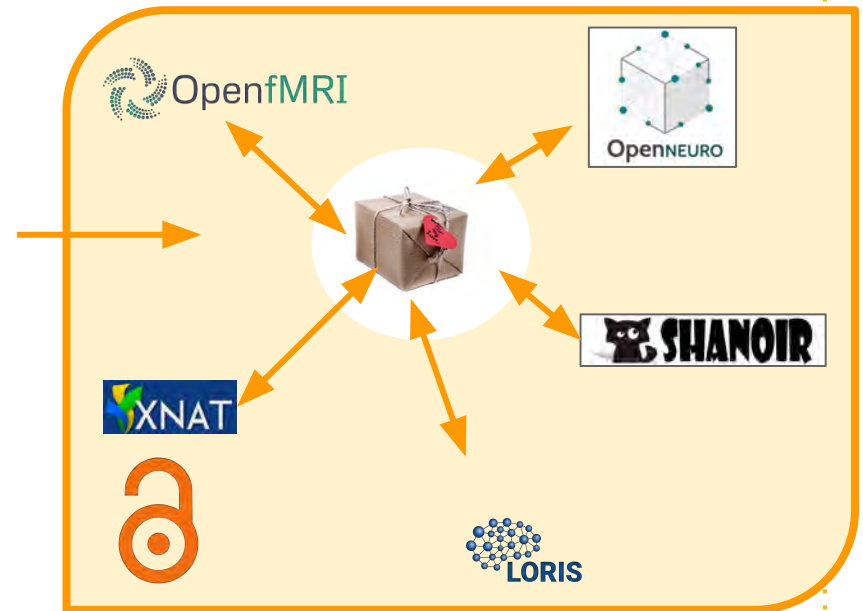
This section features logos for several open science and data management platforms: OpenfMRI, XNAT, LORIS, OpenNEURO, and SHANOIR. An orange arrow points from the cardboard box in the flowchart to this group of logos.

HAL,
Pubmed



Standards pour le partage de données

- Données et **mé**ta-données
- Un langage commun



Brain Imaging Data Structure



[Gorgolewski et. al, Scientific Data 2016]

- Utilisé par plus de 60 laboratoires dans le monde
- Données anonymisées de > 20 000 participants
- IRM, MEG, EEG

Traffic to bids.neuroimaging.io

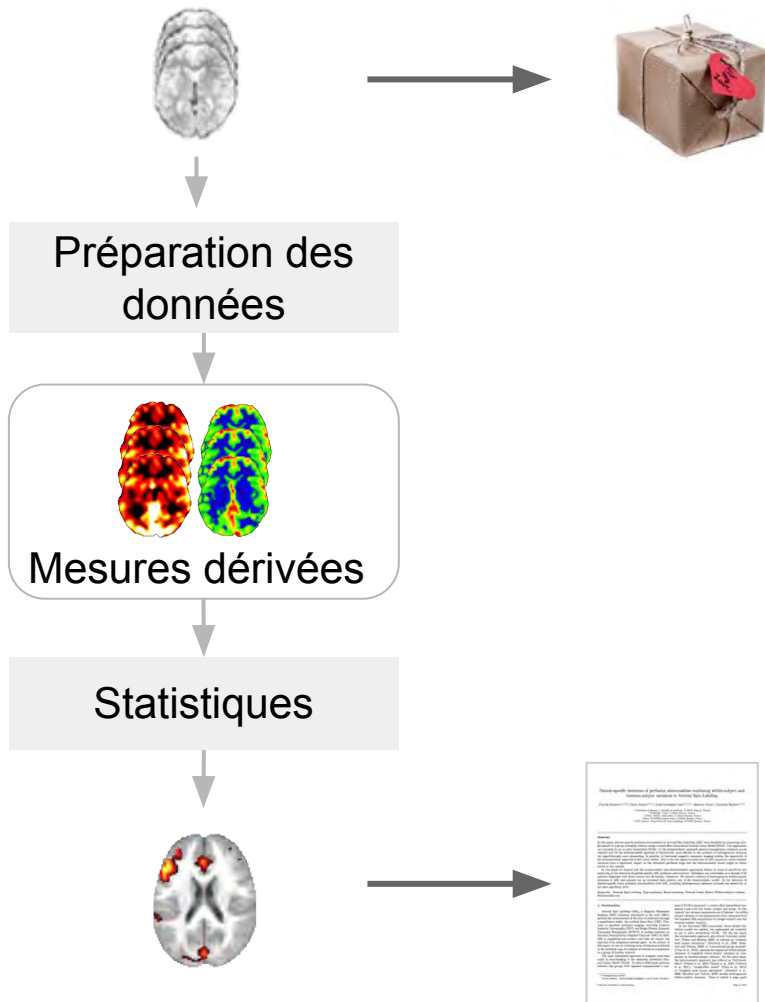


~1.8K visitors per month

En collaboration avec
Stanford Uni, Krys
Gorgolewski.

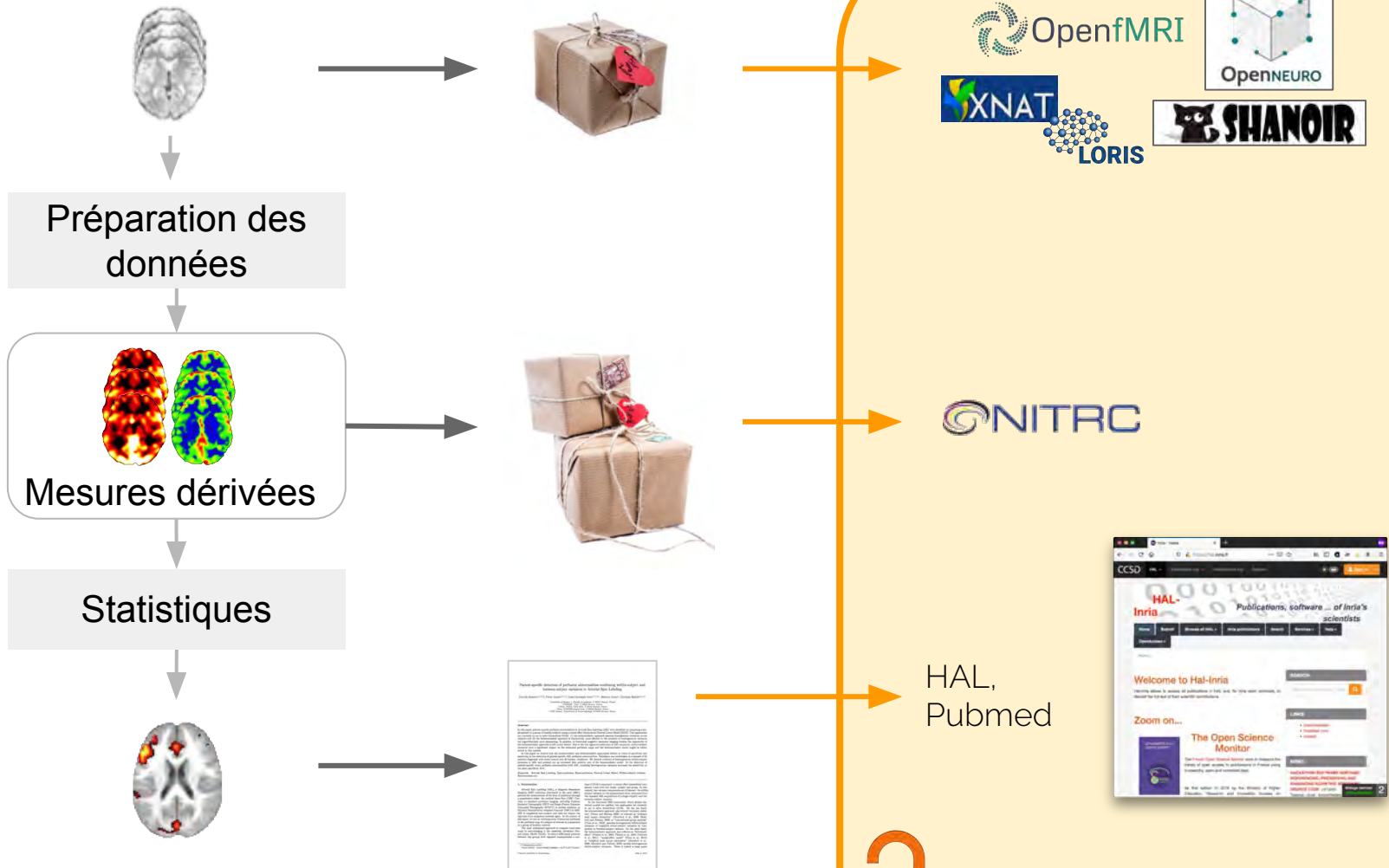


Défi 1 : Publier plus que nos articles !

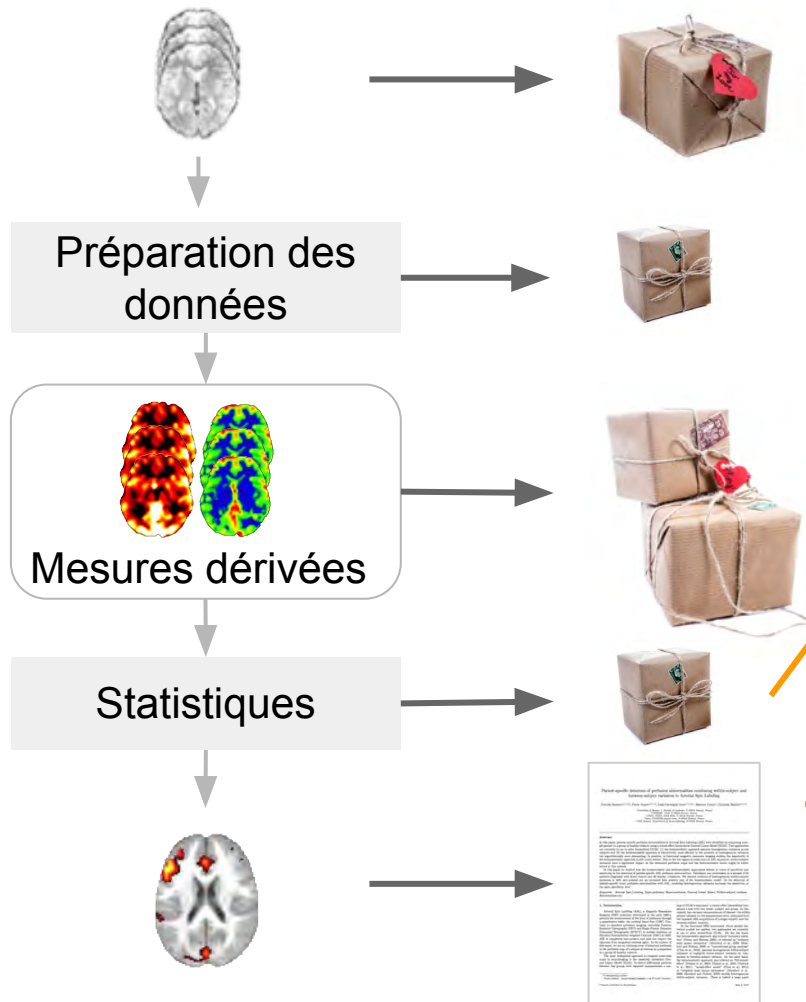


This panel highlights various open science and data sharing initiatives. At the top, it features logos for OpenfMRI, OpenNEURO, XNAT, LORIS, and SHANOIR. Below these, there is a screenshot of the HAL website, which is a platform for publishing research outputs. The text 'HAL, Pubmed' is written next to the screenshot. At the bottom of the panel is a large orange padlock icon, symbolizing open access.

Défi 1 : Publier plus que nos articles !

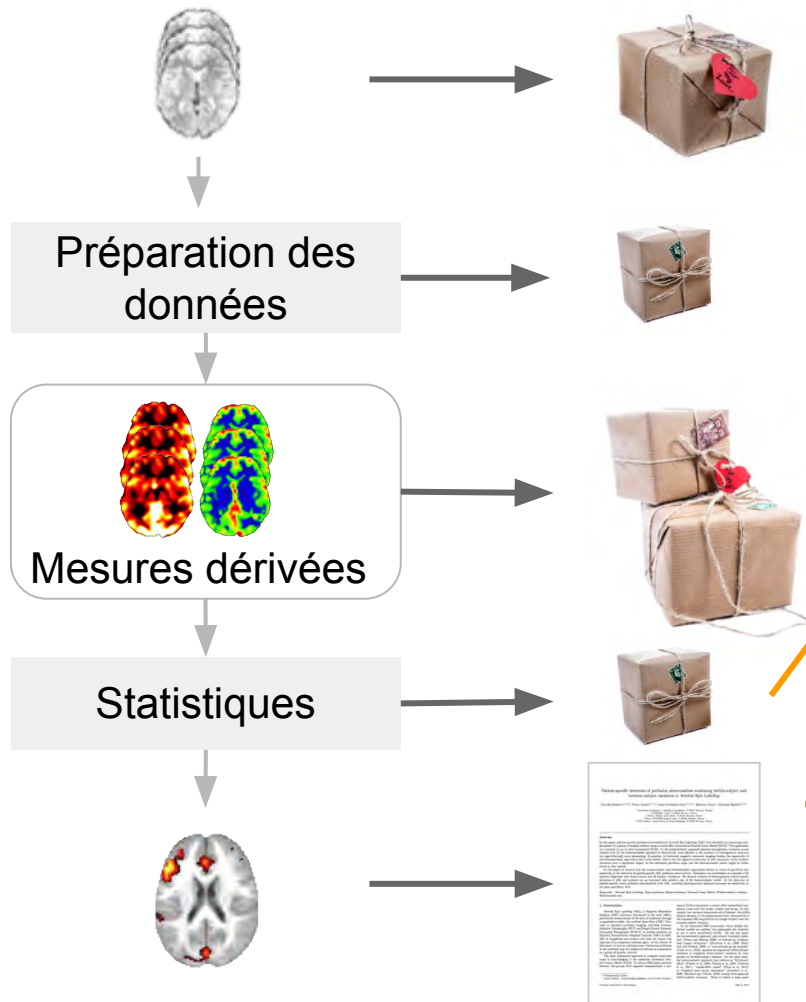


Défi 1 : Publier plus que nos articles !



A large orange-bordered box contains a collage of logos and platforms for open science and data sharing. The logos include OpenfMRI, OpenNEURO, XNAT, LORIS, SHANNOIR, Software Heritage, zenodo, and NITRC. Below these, the text 'HAL, Pubmed' is displayed next to a screenshot of the HAL website. At the bottom left of this box is the Open Access lock icon.

Défi 1 : Publier plus que nos articles !



OpenAIRE-Connect

24,024

publications

5,083

research data

150

software

2,375

other research products

8

projects

6

content providers



Neuroinformatics

SEARCH MONITOR SHARE LINK CAMILLE MAUMET

Exploring the Impact of Analysis Software on Task fMRI Results

PREPRINT ENGLISH OPEN

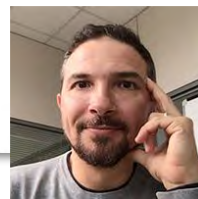
Nichols, Thomas; Maumet, Camille; Bawring, Alexander (2018)

Publisher: HAL CCSD

Subject: [SDV.NEU] Life Sciences [q-bio]/Neurons and Cognition [q-bio.NC]

A wealth of analysis tools are available to fMRI researchers in order to extract patterns of task variation and, ultimately, understand cognitive function. However, this 'methodological plurality' comes with a drawback. While conceptually similar, two different analysis pipelines applied on the same dataset may not produce the same scientific results. Differences in methods, implementations across software packages, and even operating systems or software versions all contribute to this variability. Consequently, attention in the field has recently been directed to reproducibility and data sharing. Neuroimaging is currently experiencing a surge in initiatives to improve research practices and ensure that all conclusions inferred from an fMRI study are replicable. In this work, our goal is to understand how choice of software package impacts on analysis results. We use publically shared data from three published task fMRI neuroimaging studies, reanalyzing each study using the three main neuroimaging software packages, AFNI, FSL and SPM, using parametric and nonparametric inference. We obtain all information on how to process, analyze, and model each dataset from the publications. We make quantitative and qualitative comparisons between our replications to gauge the scale of variability in our results and assess the fundamental differences between each software package. While qualitatively we find broad similarities between packages, we also discover marked differences, such as Dice similarity coefficients ranging from 0.000-0.743 in comparisons of thresholded statistic maps between software. We discuss the challenges involved in trying to reanalyse the published studies, and highlight our own efforts to make this research reproducible.

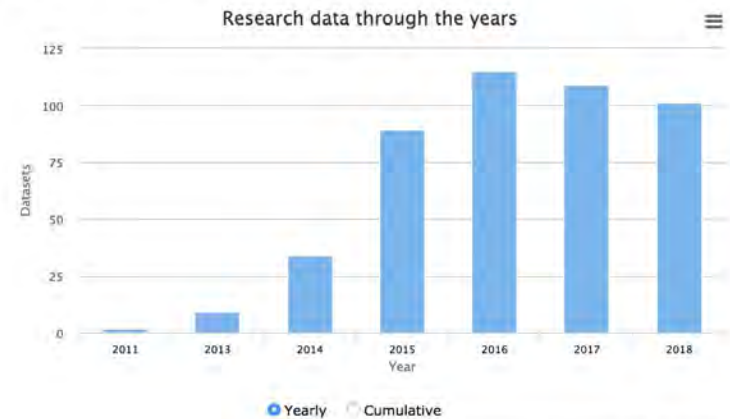
Metrics



En collaboration avec
CNR Italie, Paolo
Manghi.

Research Data statistics

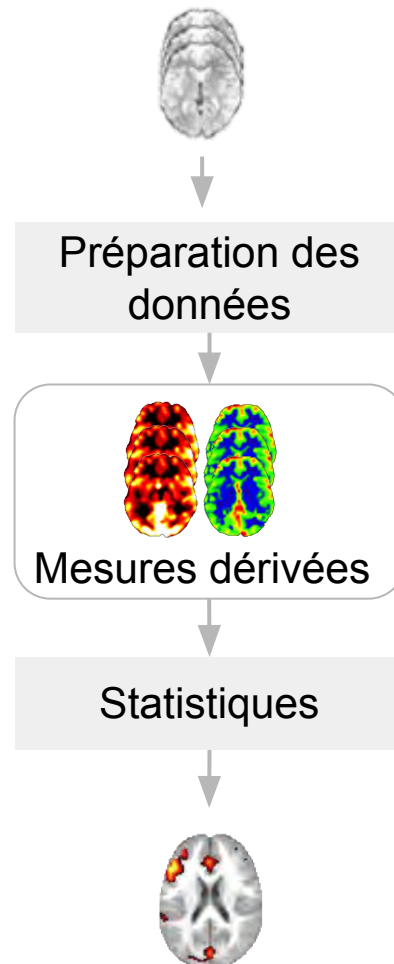
789 DATASETS LINKED TO 1 PROJECT.
681 ARE OA, 1 IS CLOSED AND 1 IS STILL IN EMBARGO.



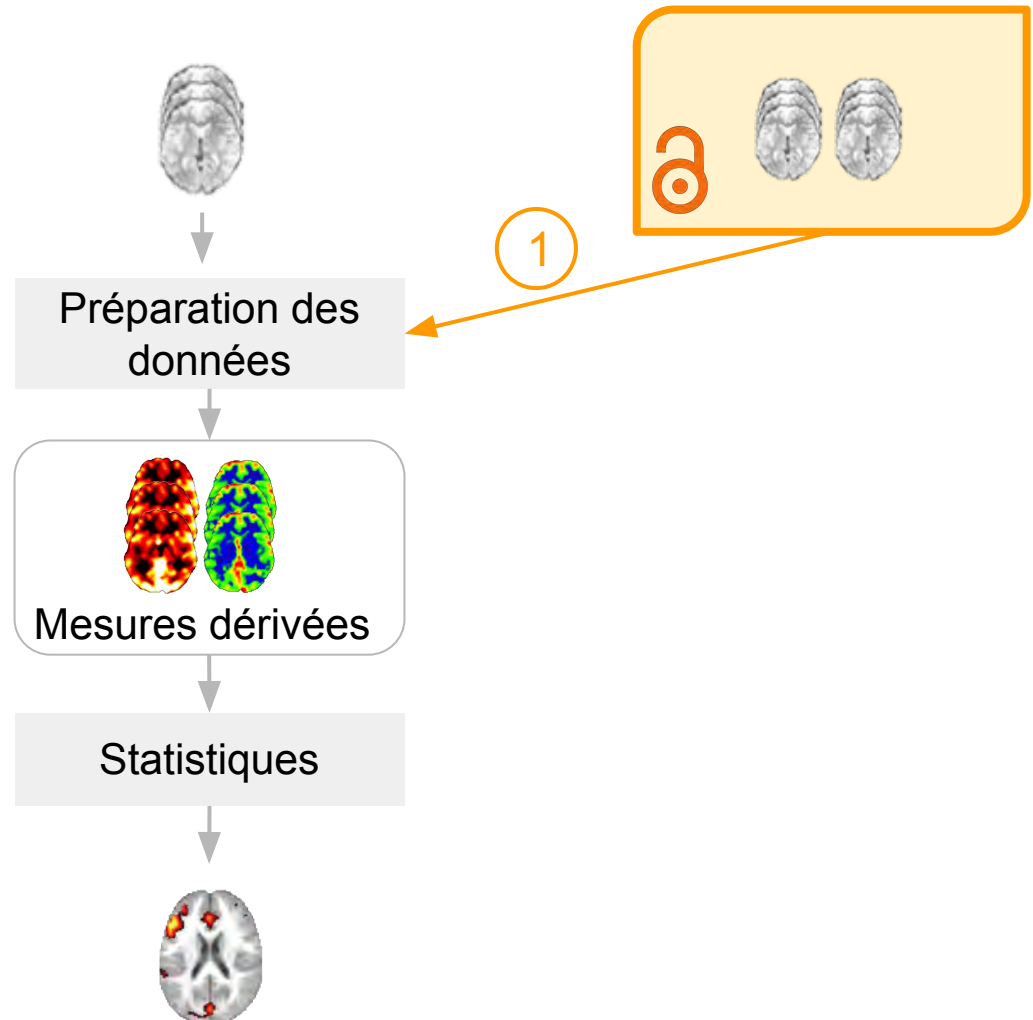
Research data by access mode



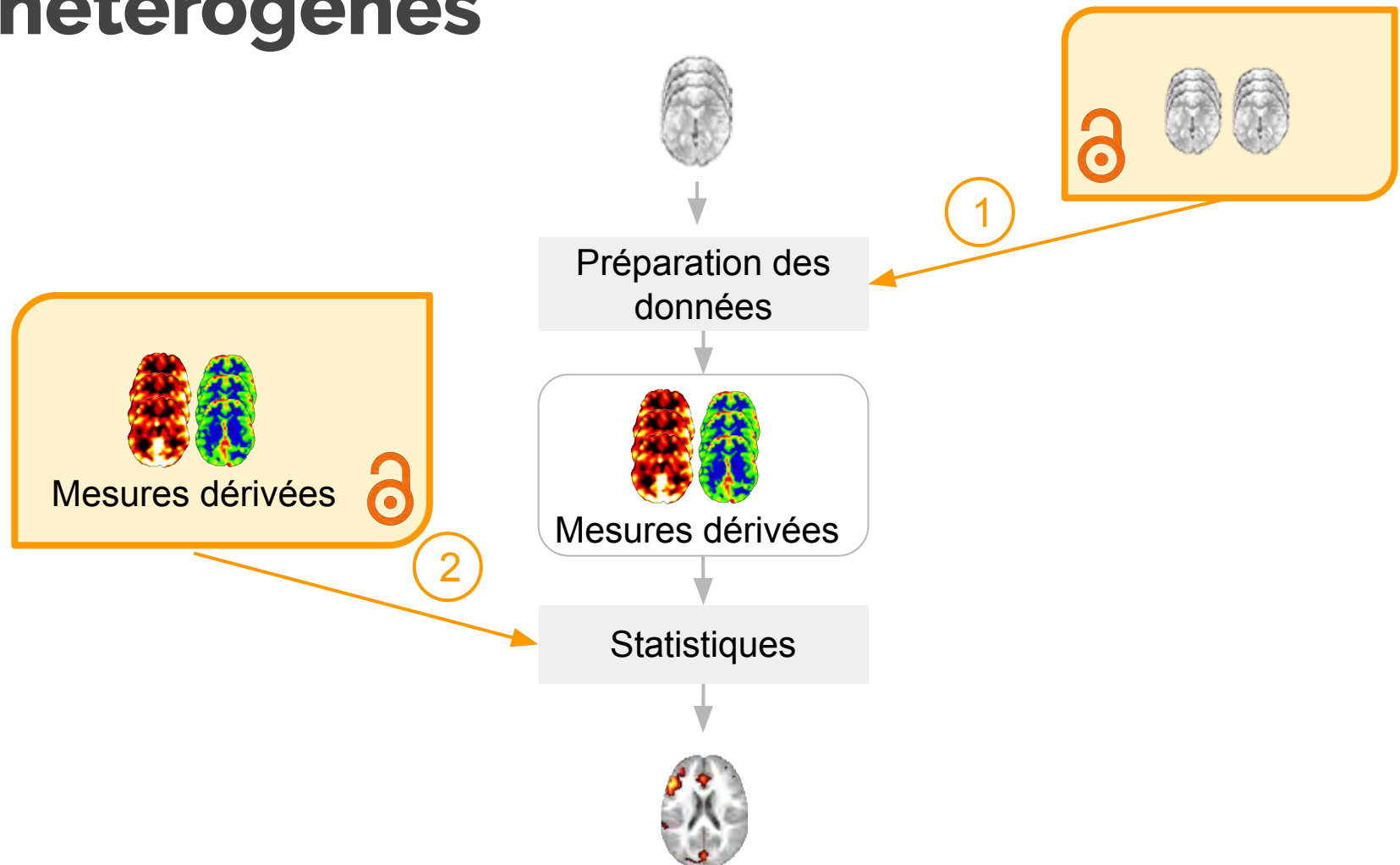
Défi 2 : Travailler avec des données hétérogènes



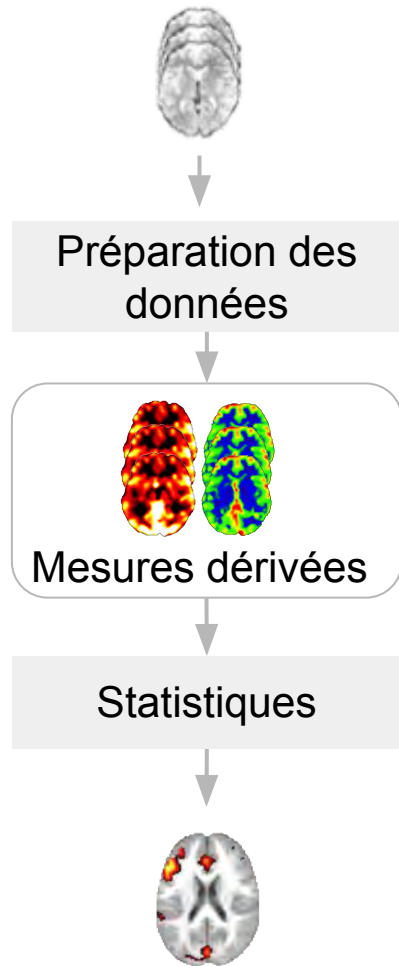
Défi 2 : Travailler avec des données hétérogènes



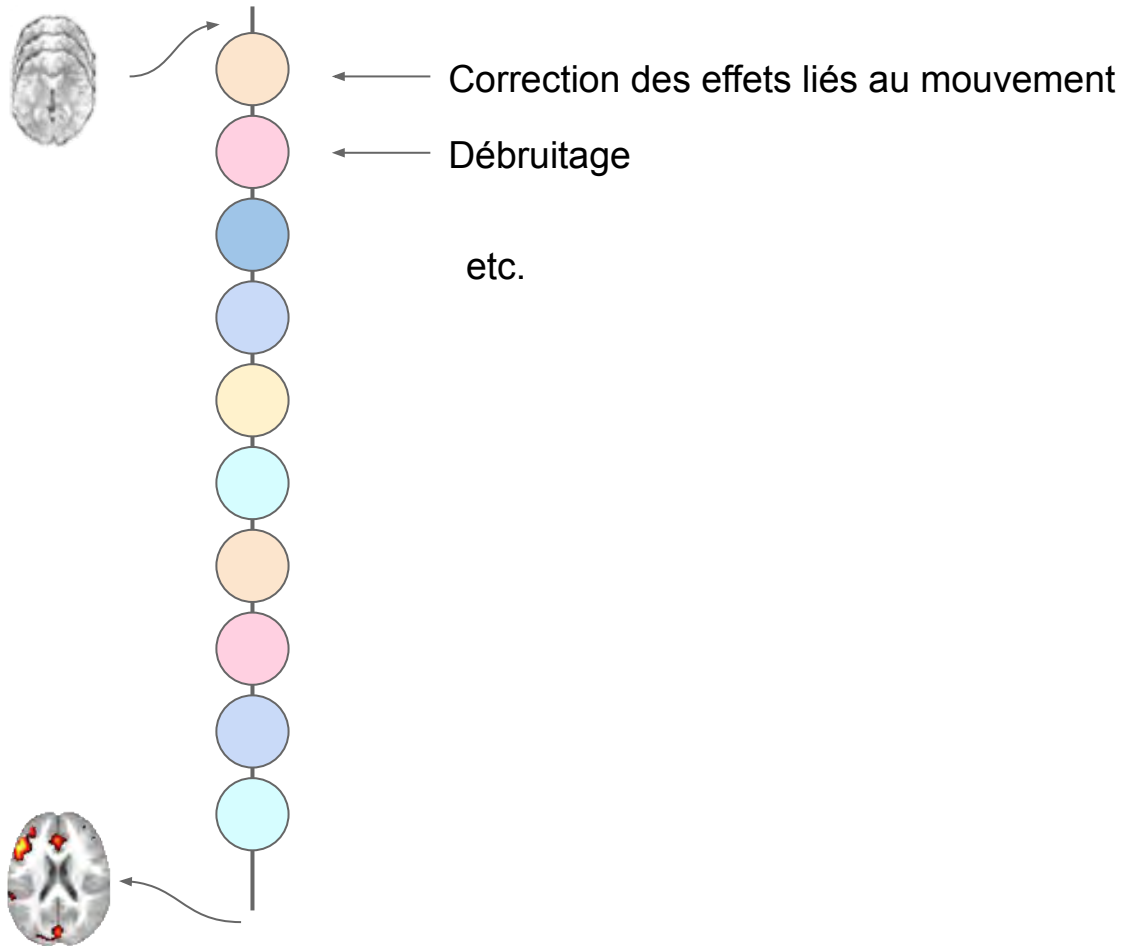
Défi 2 : Travailler avec des données hétérogènes



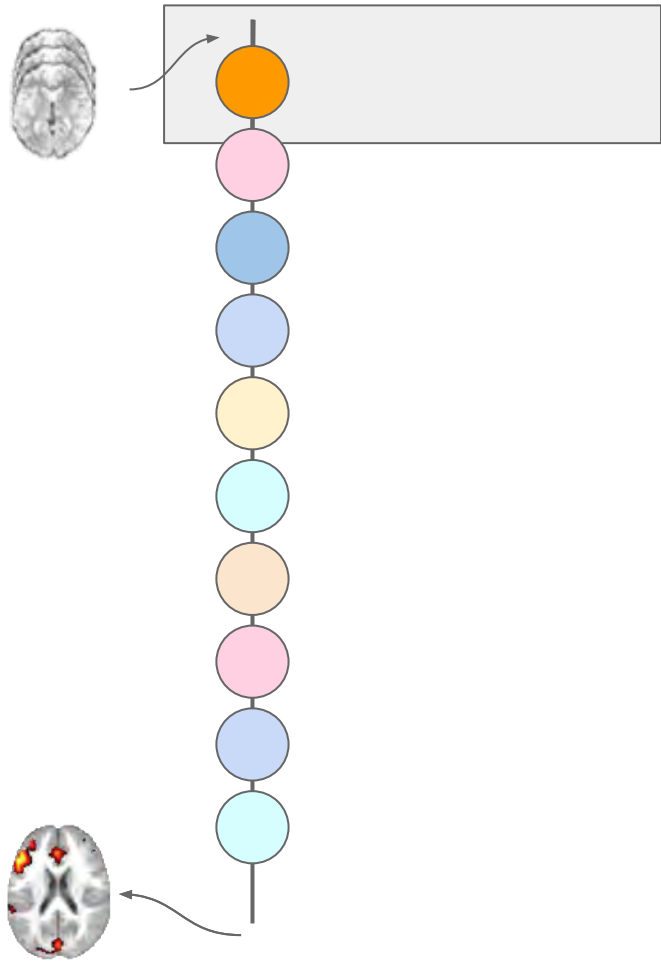
Variabilité analytique



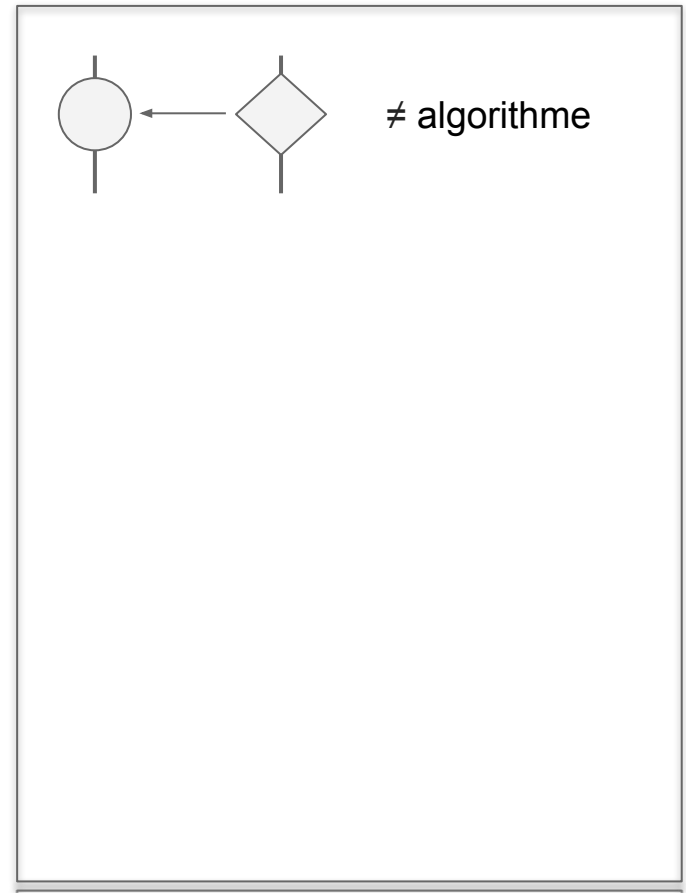
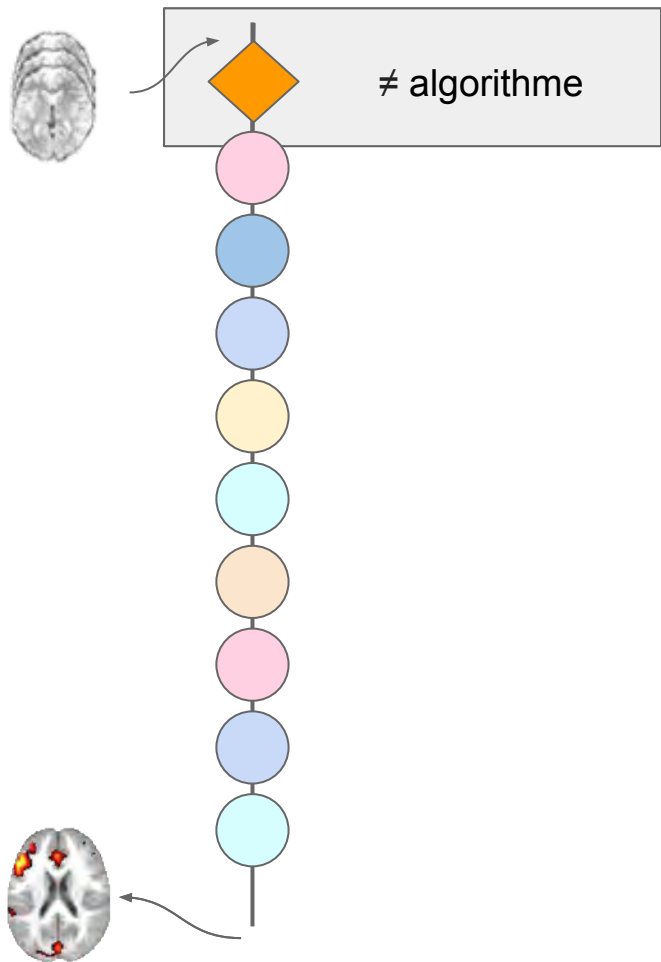
Variabilité analytique



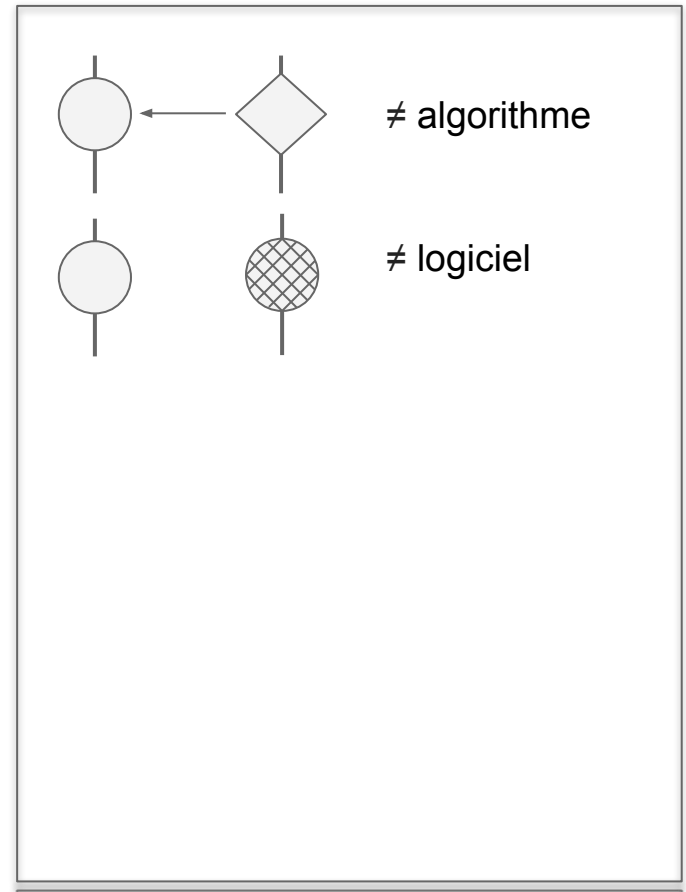
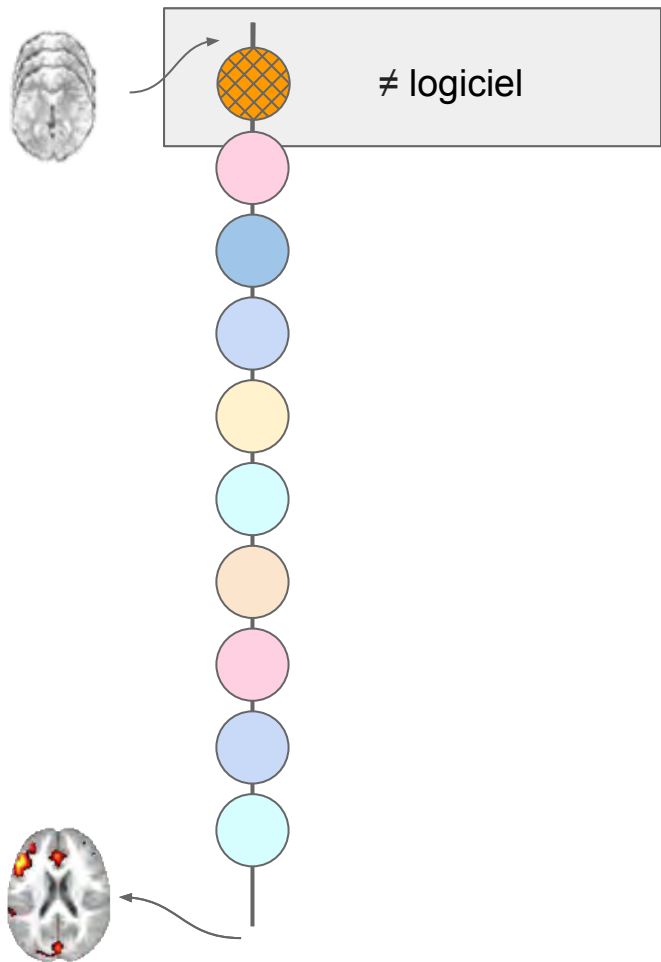
Variabilité analytique



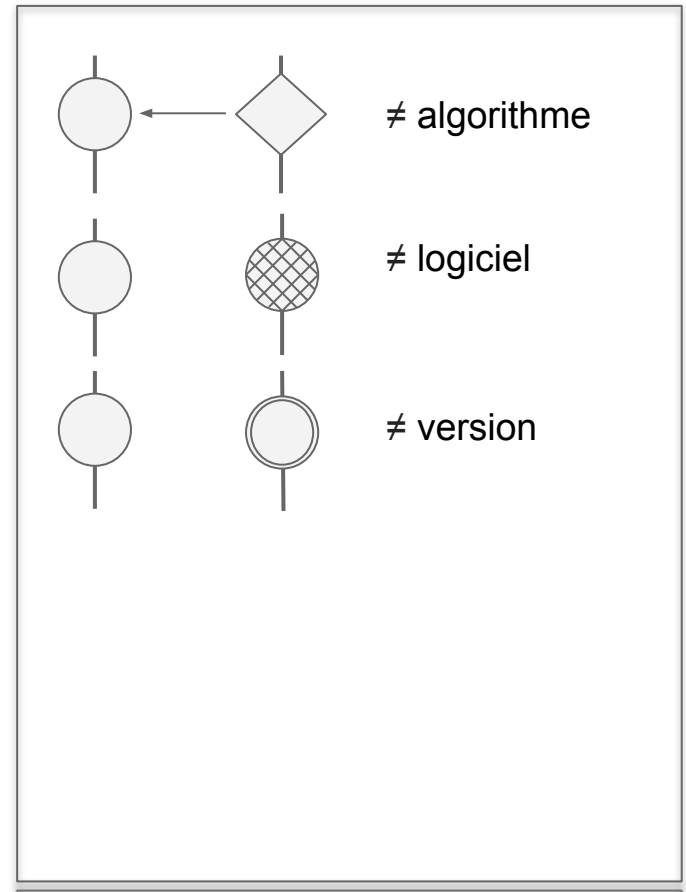
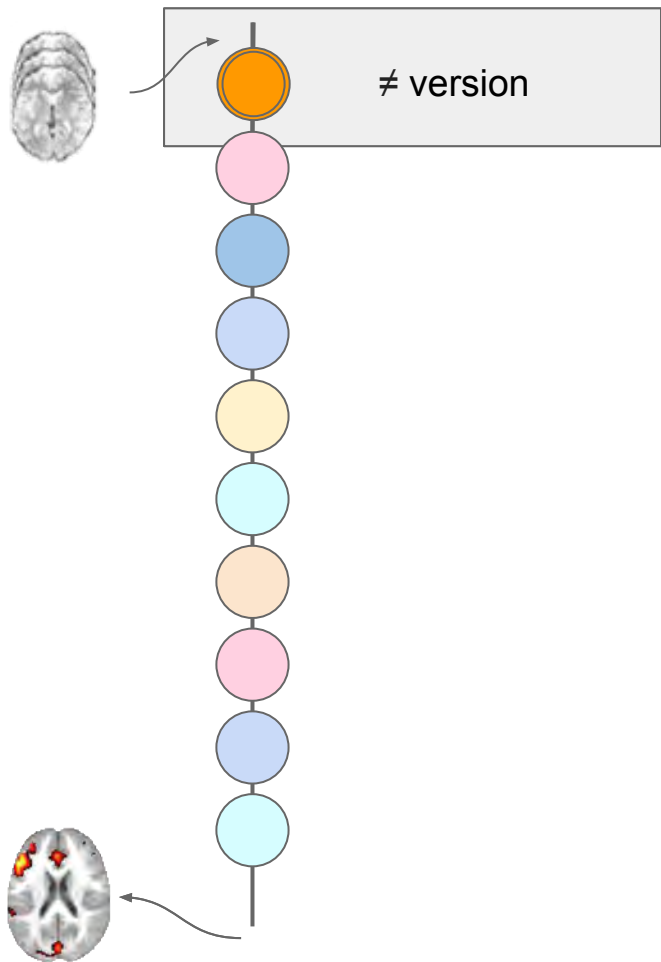
Variabilité analytique



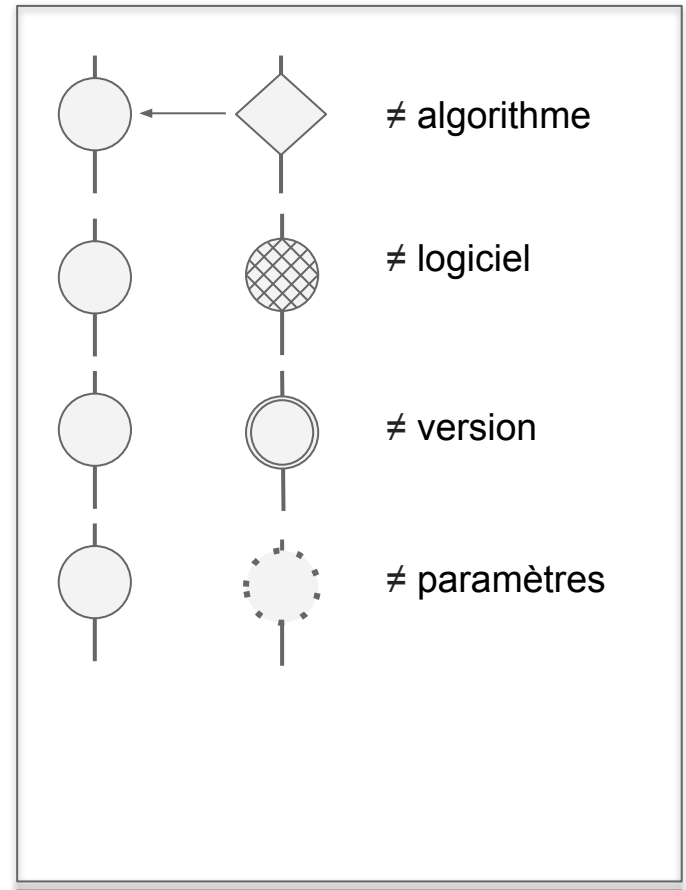
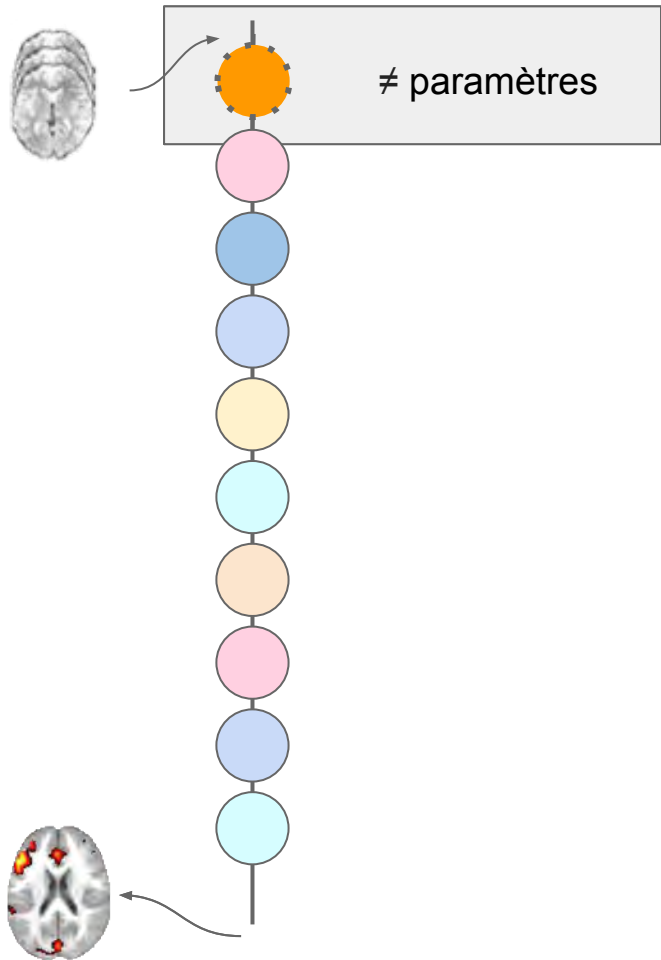
Variabilité analytique



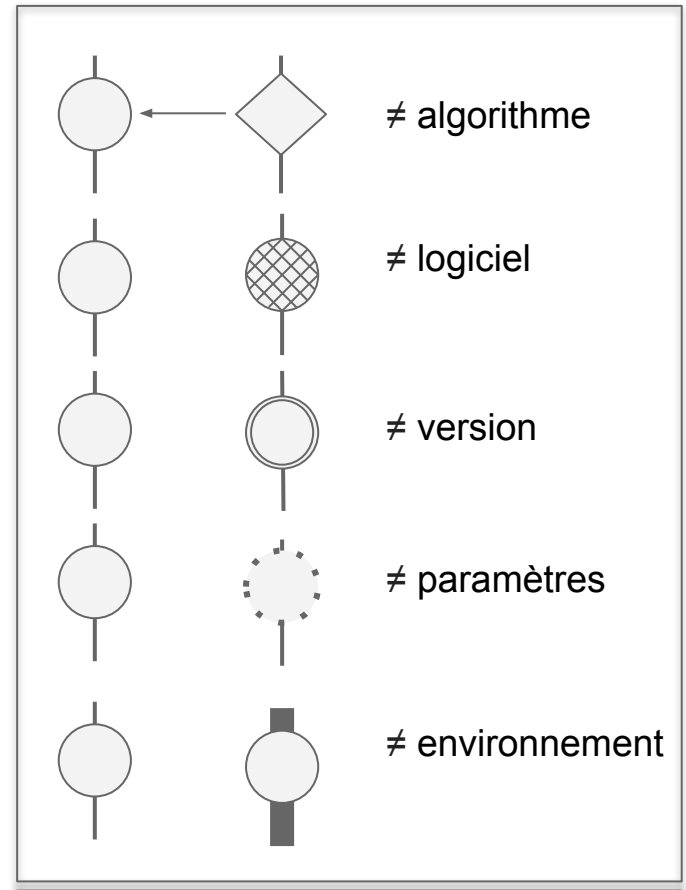
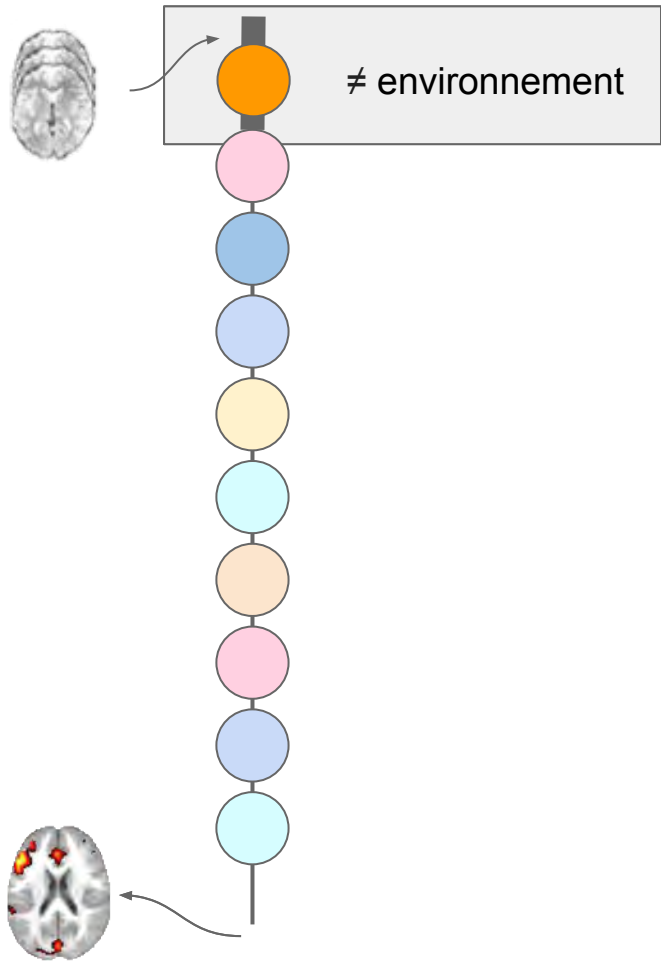
Variabilité analytique



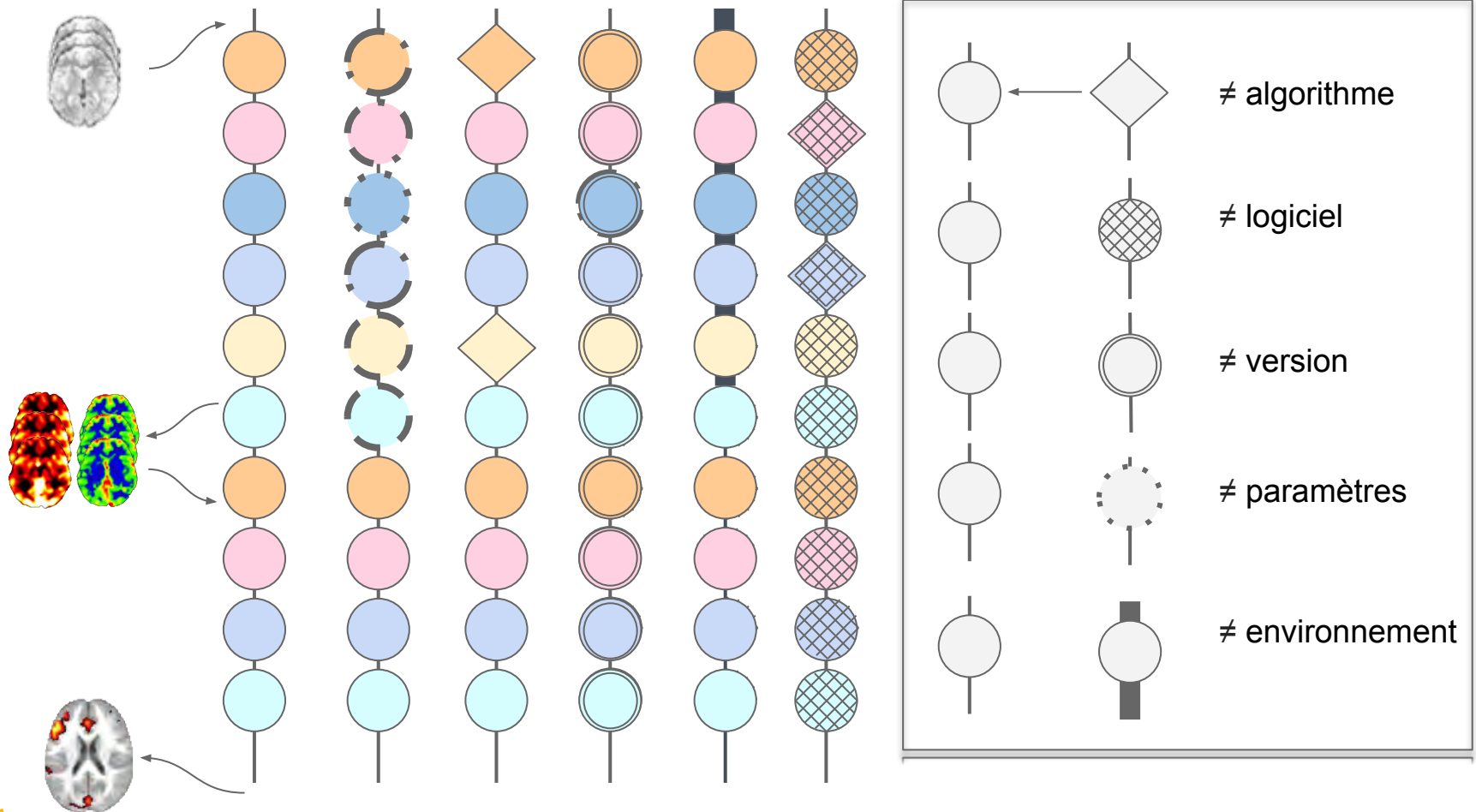
Variabilité analytique



Variabilité analytique



Variabilité analytique



Variabilité analytique en IRM

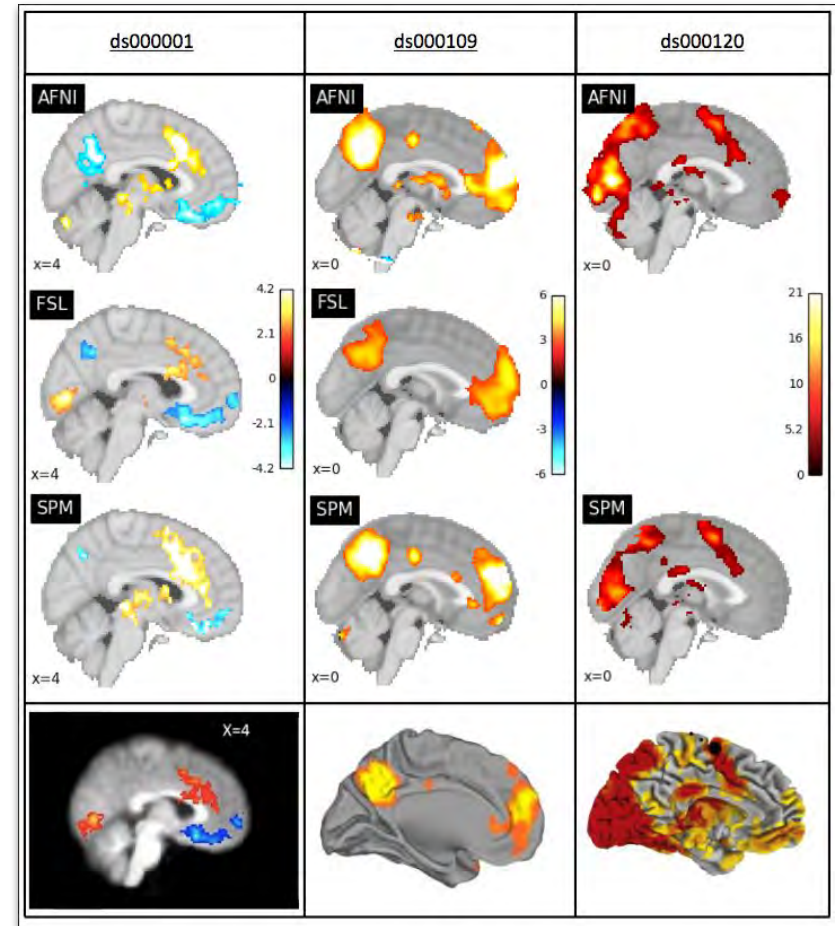
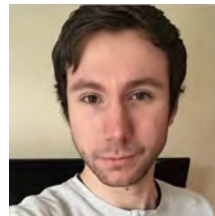
Liée aux algorithmes
[Carp, Front. Neuroscience 2012]

Liée à la version
[Groenenschild, PlosOne, 2012]

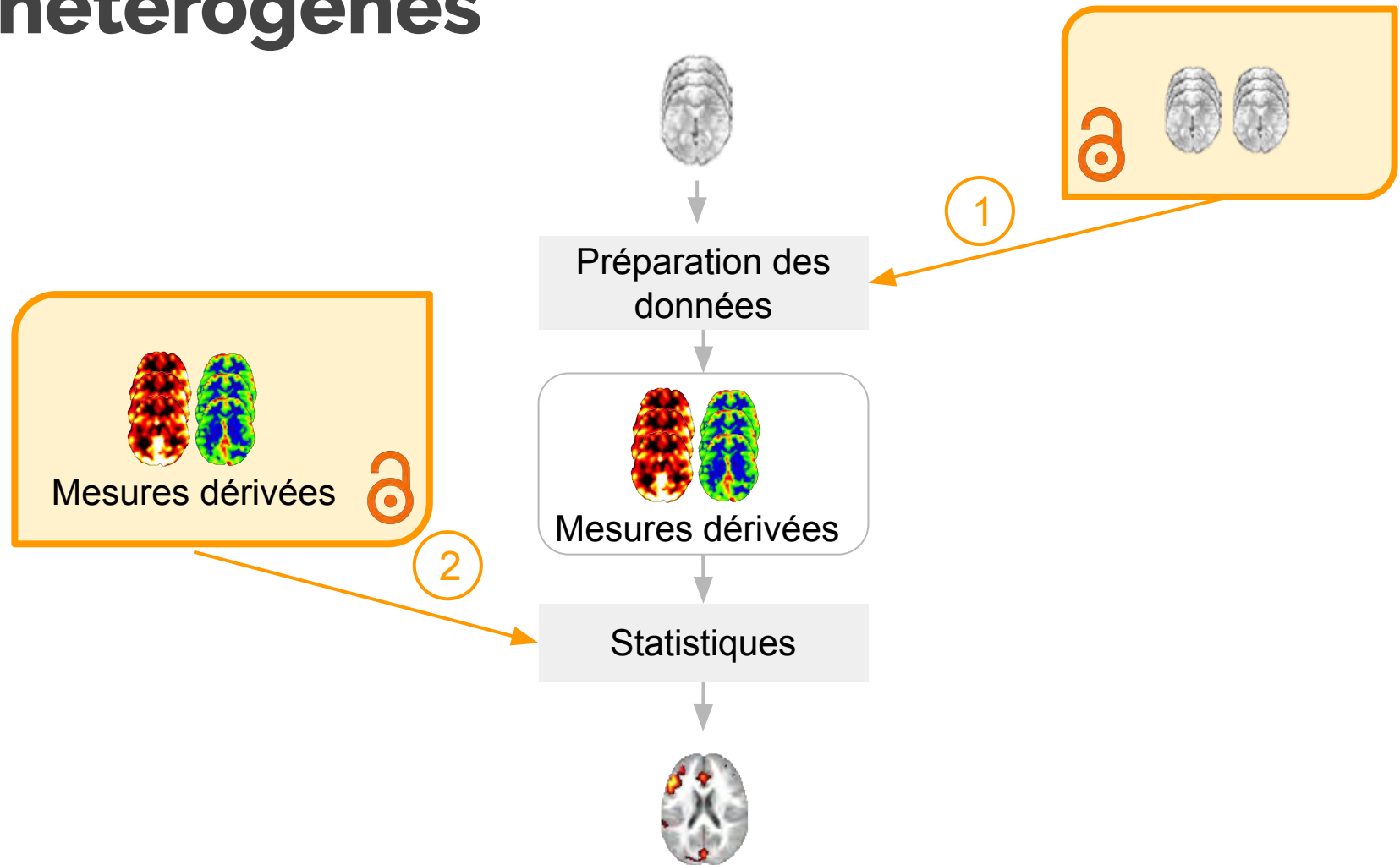
Liée au système d'exploitation
[Glatard, Front. Neuroinformatics 2015]

Liée au logiciel
[Bowring et. al, HBM 2019]

En collaboration avec
Uni. of Oxford, Thomas
Nichols & Alex Bowring.



Défi 2 : Travailler avec des données hétérogènes



Quand les données s'ouvrent :

Opportunités et nouveaux défis pour mieux comprendre notre cerveau



Merci



Camille Maumet

Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, Empenn ERL U-1228

<http://camillemaumet.com>

<https://team.inria.fr/empenn/>



[cmaumet](https://twitter.com/cmaumet)

[empenn_lab](https://twitter.com/empenn_lab)