Journée INS2I Santé et IA - 23 Janvier 2020

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

Camille Maumet

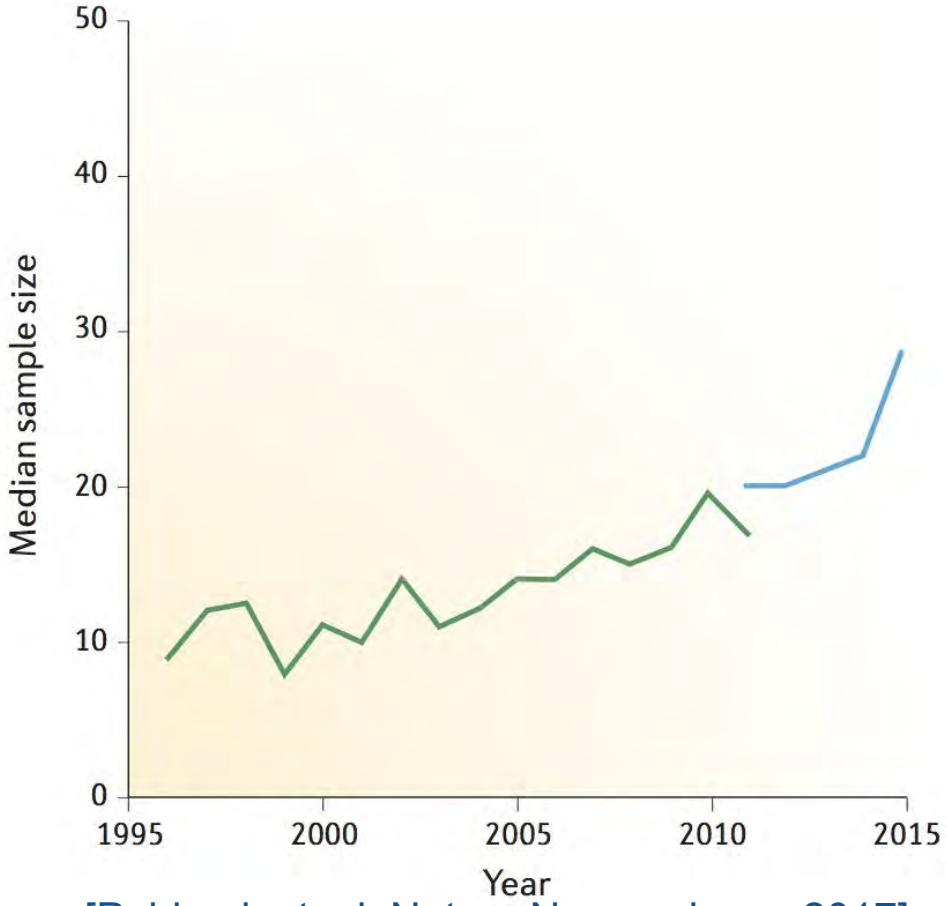
Univ Rennes, Inria, CNRS, Inserm
IRISA UMR 6074, Empenn ERL U-1228

@cmaumet

UNIVERSITÉ DE RENNES 1 — Inria inventeurs du monde numérique — Inserm Institut national de la santé et de la recherche médicale — cnrs — UMR IRISA

La neuroimagerie aujourd'hui

Tailles des études de neuroimagerie



[Poldrack et. al, Nature Neuroscience 2017]

2015 : médiane de 30 participants par étude

M PIXELS | CHRONIQUES DES (R)ÉVOLUTIONS NUMÉRIQUES

VIE EN LIGNE

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

Le Monde Pixels, "Une étude démontre les biais de la reconnaissance faciale", Perrine Signoret.

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.

Biais de sélection

Plus rapides à 40 ans qu'à 20 ans ?

The Guardian, "Why are middle-aged marathon runners faster than twenty somethings?". Oliver Balch

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

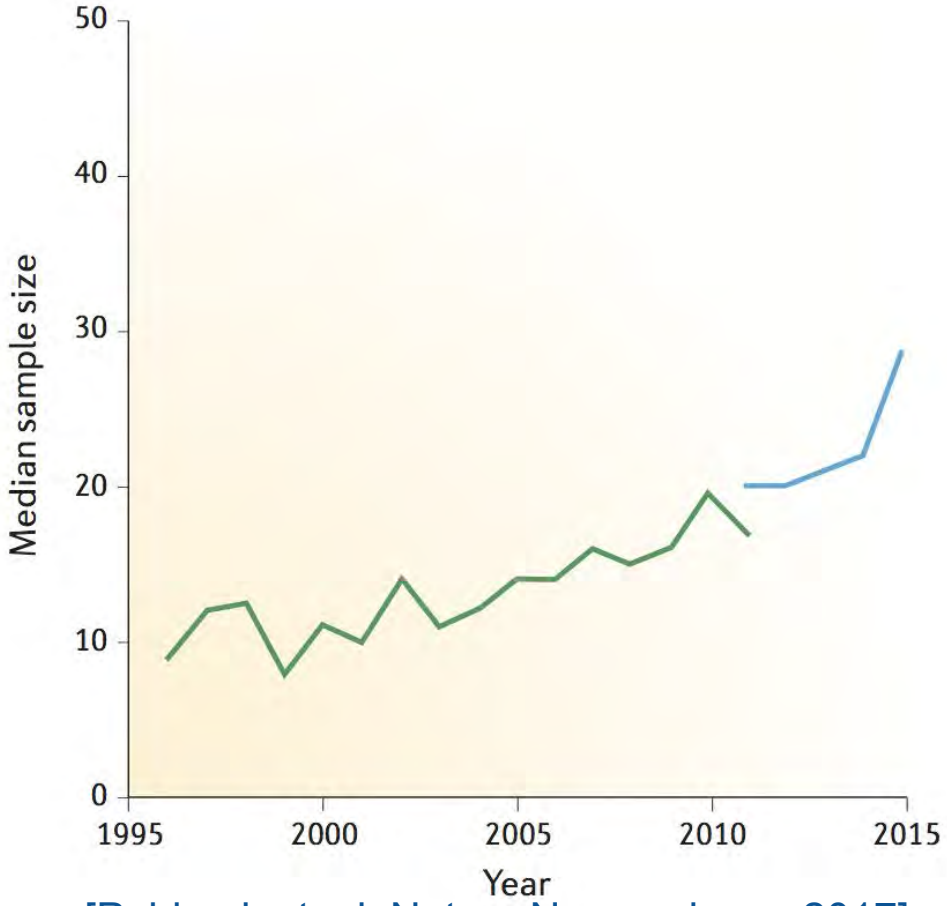
In 1996, a group of European researchers found that a certain gene, called *SLC6A4*, might influence a person's risk of depression.

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

The Atlantic Science, "A waste of 1000 research papers", Ed Yong.

Tailles des études de neuroimagerie



[Poldrack et. al, Nature Neuroscience 2017]

2015 : médiane de 30 participants par étude

Représentativité
Biais de sélection
Effet de vibration

Quand les données s’ouvrent…

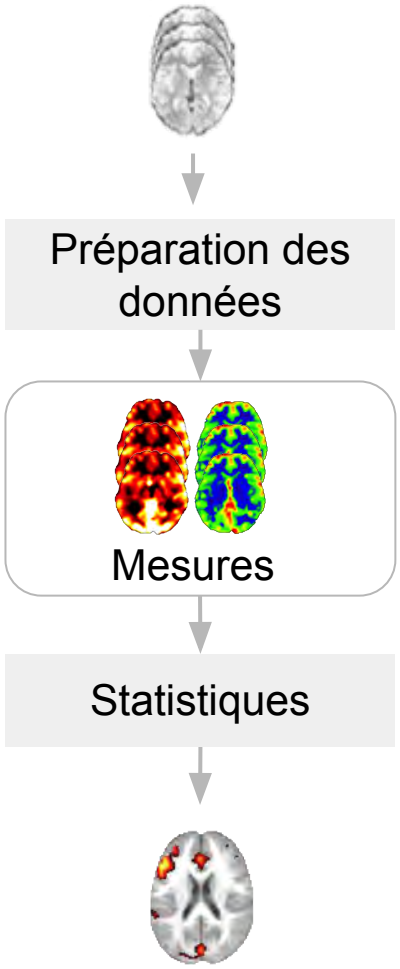
Les données ouvertes

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

Wikipedia

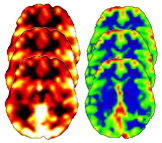
Crédits : Cerveaux, Neil Conway, Flickr (CC BY 2.0)

Ouvrir les données d'une étude



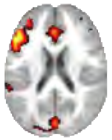
Ouvrir les données d'une étude

Préparation des données



Mesures

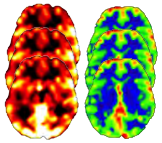
Statistiques



Communauté scientifique

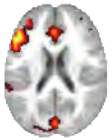
Ouvrir les données d'une étude

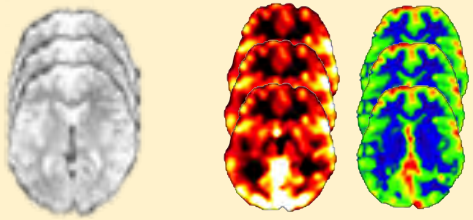
Préparation des données



Mesures

Statistiques





Communauté scientifique

Données ouvertes

Étude unique

30 participants

OpenNEURO

studyforrest.org

NEUROVAULT

MYCONNECTOME

NITRC

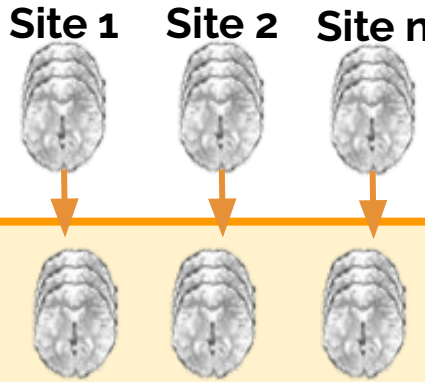
OSF

Crédits : Cerveaux, Neil Conway, Flickr (CC BY 2.0)

Créer un consortium

Site 1 Site 2 Site n

Créer un consortium



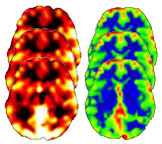
Communauté scientifique

Créer un consortium

Site 1 Site 2 Site n

Communauté scientifique

Préparation des données



Mesures dérivées

Statistiques

Données ouvertes

Étude unique
30 participants

Consortium
1000 participants

OpenNEURO

studyforrest.org

NEUROVAULT

MYCONNECTOME

ABIDE
Autism Brain Imaging Data Exchange

1000 Functional Connectomes Project

NITRC

ADHD 200

nki

OSF

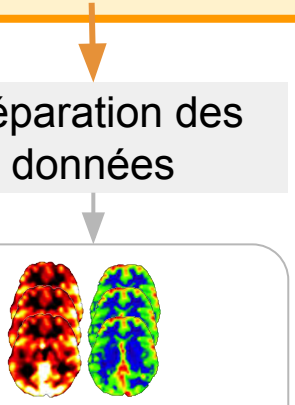
CORR
CONSORTIUM FOR RELIABILITY AND REPRODUCIBILITY

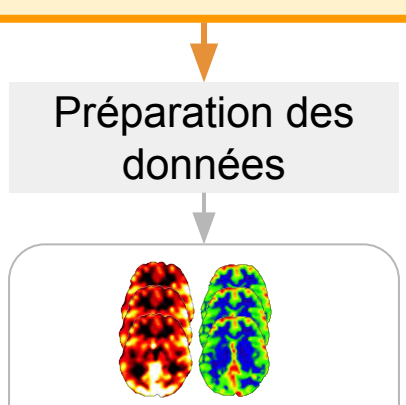
Créer une ressource commune

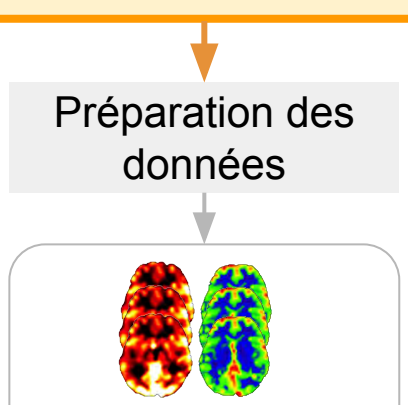
Communauté scientifique

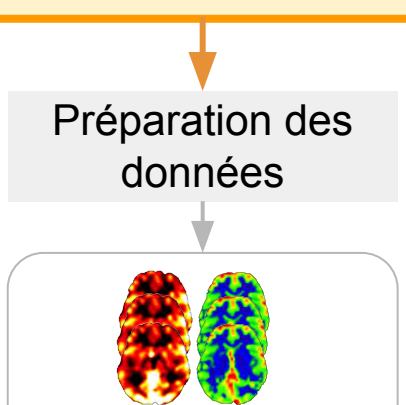
Créer une ressource commune

Communauté scientifique









Données ouvertes

Étude unique
30 participants

Consortium
1000 participants

Cohorte
100 000 participants

OpenNEURO

studyforrest.org

NEUROVAULT

MYCONNECTOME

ABIDE
Autism Brain Imaging Data Exchange

1000 Functional Connectomes Project

NITRC

ADHD 200

CORR
CONSORTIUM FOR RELIABILITY AND REPRODUCIBILITY

nki

HUMAN Connectome PROJECT

OSF

biobank

Crédits : Cerveaux, Neil Conway, Flickr (CC BY 2.0)

Données ouvertes

Étude unique
30 participants

Consortium
1000 participants

Cohorte
100 000 participants

OpenNEURO

studyforrest.org

NEUROVAULT

MYCONNECTOME

ABIDE
Autism Brain Imaging Data Exchange

1000 Functional Connectomes Project

ADHD 200

NITRC

CORR
CONSORTIUM FOR RELIABILITY AND REPRODUCIBILITY

nki

HUMAN Connectome PROJECT

OSF

biobank

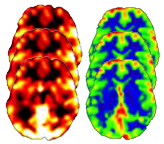
+ Images
+ Homogène
- Exemplaires

Crédits : Cerveaux, Neil Conway, Flickr (CC BY 2.0)

De nouveaux défis

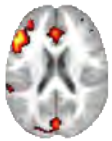
Défi 1 : Publier plus que nos articles !

Préparation des données

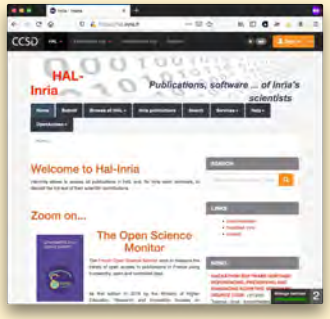


Mesures dérivées

Statistiques

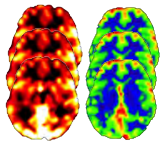


HAL, Pubmed



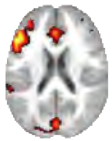
Défi 1 : Publier plus que nos articles !

Préparation des données



Mesures dérivées

Statistiques



OpenfMRI

OpenNEURO

XNAT

LORIS

SHANOIR

HAL,
Pubmed

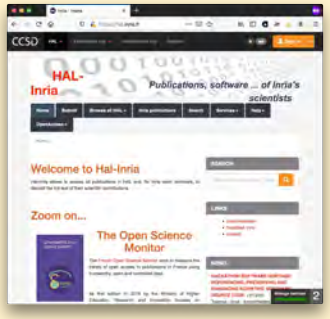
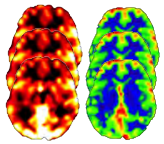


Image credits: Parcels 1 2 & 4 (CC0), Parcel 3 (CC0), Parcel 5 (CC0).

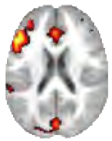
Défi 1 : Publier plus que nos articles !

Préparation des données



Mesures dérivées

Statistiques



OpenfMRI

OpenNEURO

XNAT

LORIS

SHANOIR

HAL,
Pubmed

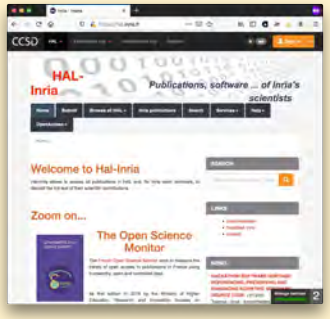


Image credits: Parcels 1 2 & 4 (CC0), Parcel 3 (CC0), Parcel 5 (CC0).

Standards pour le partage de données

- Données et **méta-données**
- Un langage commun

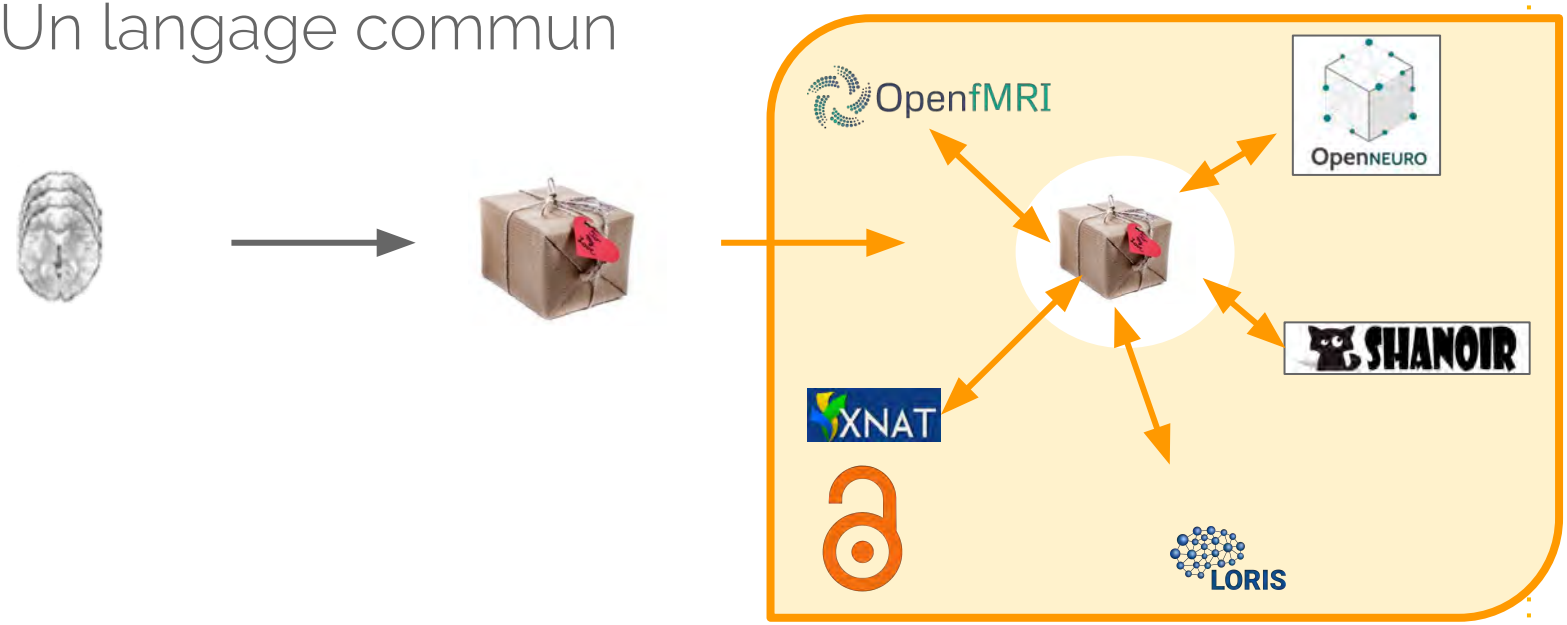
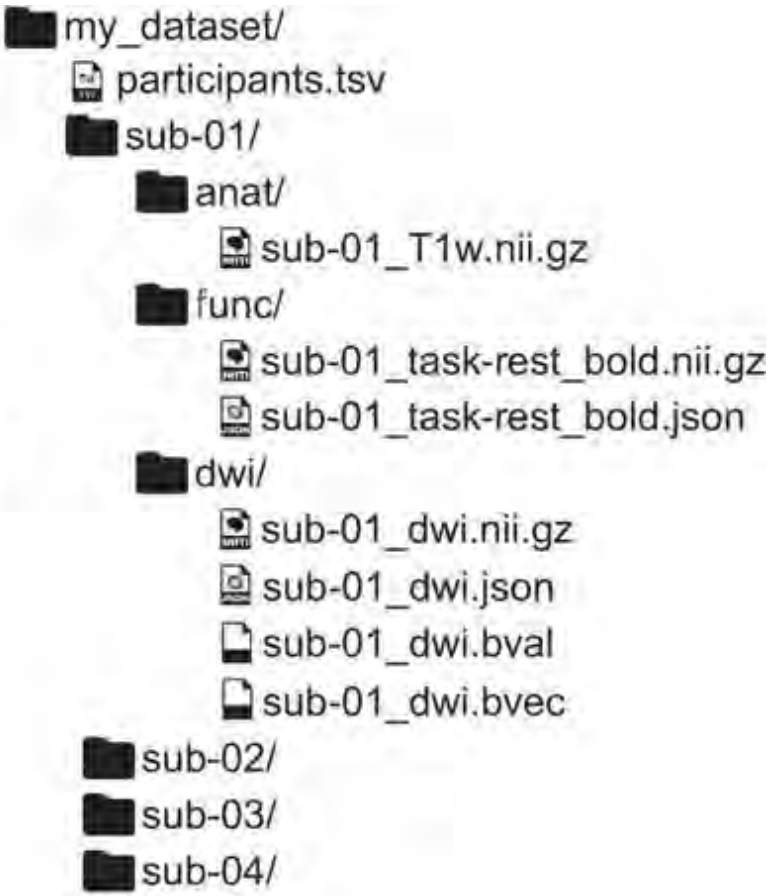


Image credits: Parcels 1 2 & 4 (CC0),

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

En collaboration avec Stanford Uni, Krys Gorgolewski.

Défi 1 : Publier plus que nos articles !

OpenfMRI

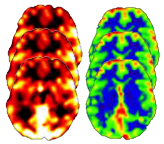
OpenNEURO

XNAT

LORIS

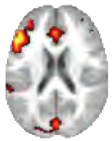
SHANOIR

Préparation des données

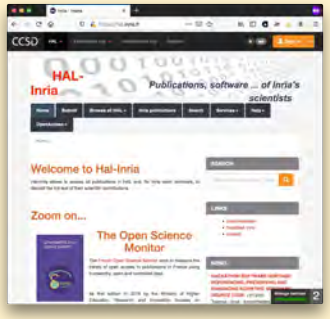


Mesures dérivées

Statistiques



HAL,
Pubmed



Défi 1 : Publier plus que nos articles !

OpenfMRI

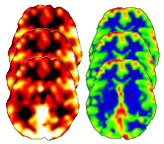
OpenNEURO

XNAT

LORIS

SHANOIR

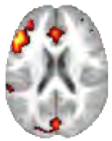
Préparation des données



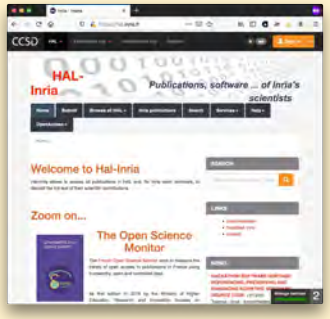
Mesures dérivées

NITRC

Statistiques

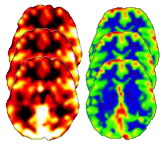


HAL,
Pubmed



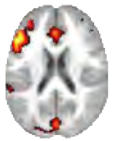
Défi 1 : Publier plus que nos articles !

Préparation des données



Mesures dérivées

Statistiques



OpenfMRI

OpenNEURO

XNAT

LORIS

SHANOIR

Software Heritage

zenodo

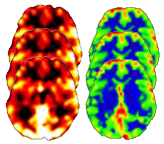
NITRC

HAL,
Pubmed

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

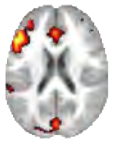
Défi 1 : Publier plus que nos articles !

Préparation des données



Mesures dérivées

Statistiques



OpenfMRI

OpenNEURO

XNAT

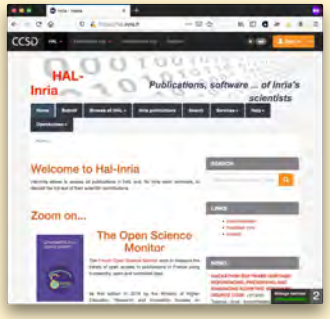
LORIS

SHANOIR

Software Heritage

zenodo

NITRC

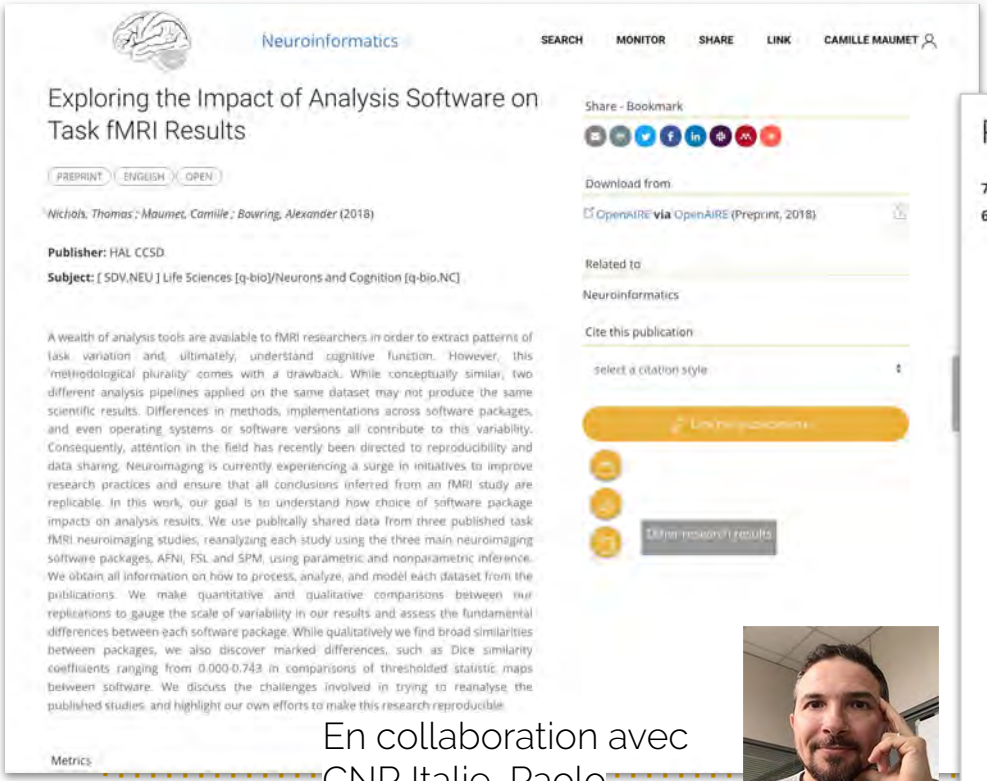


HAL,
Pubmed

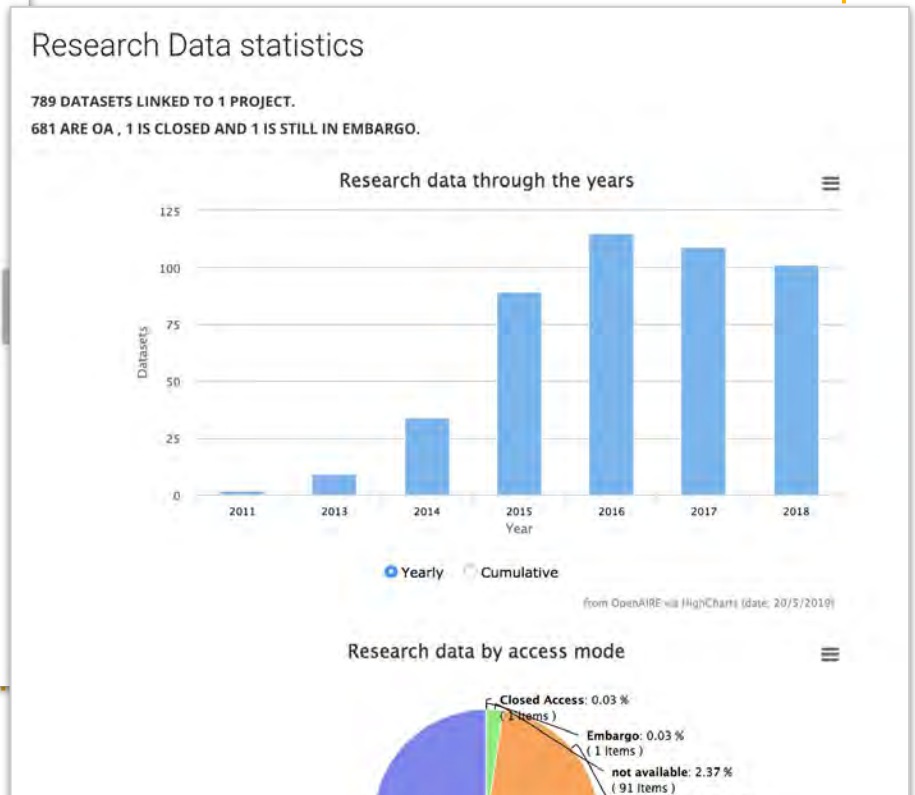
CONNECT

OpenAIRE-Connect

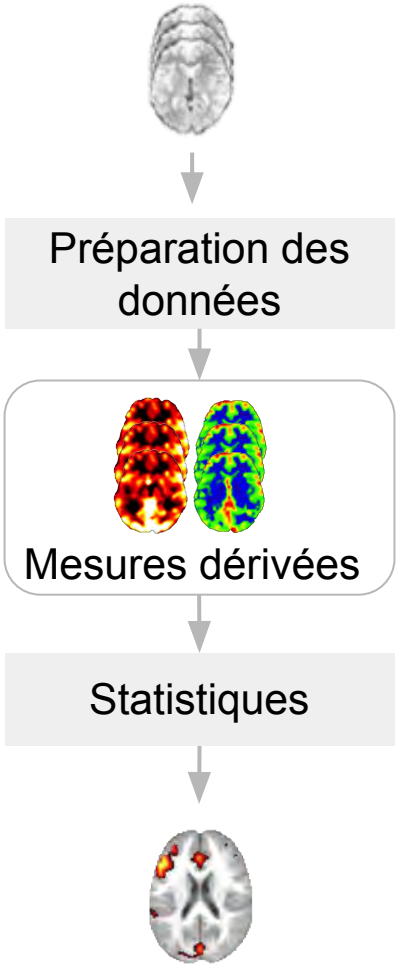
24,024	5,083	150	2,375	8	6
publications	research data	software	other research products	projects	content providers



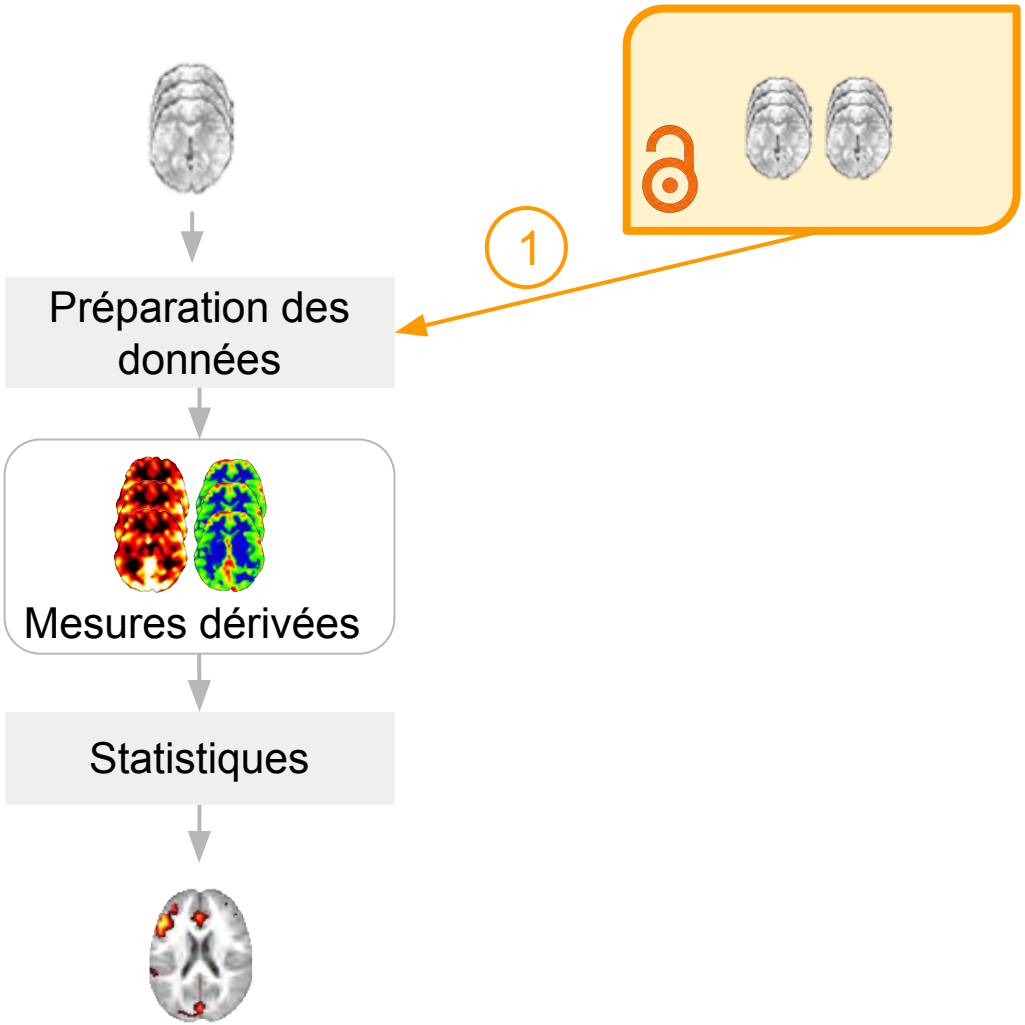
En collaboration avec CNR Italie, Paolo Manghi.



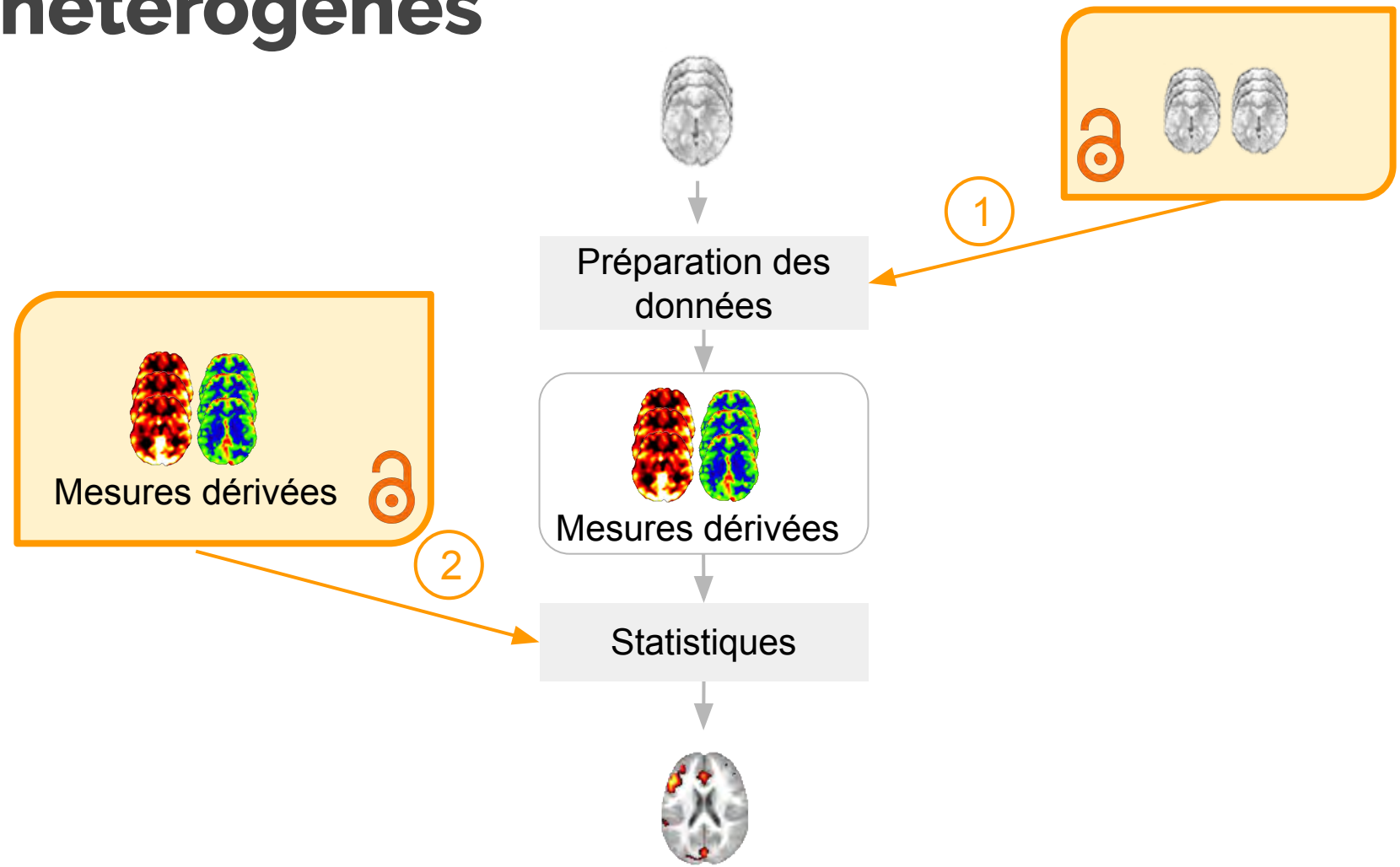
Défi 2 : Travailler avec des données hétérogènes



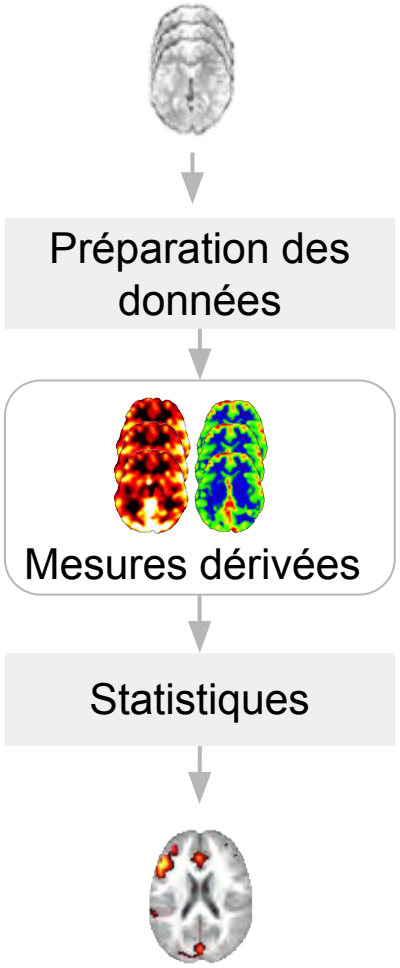
Défi 2 : Travailler avec des données hétérogènes



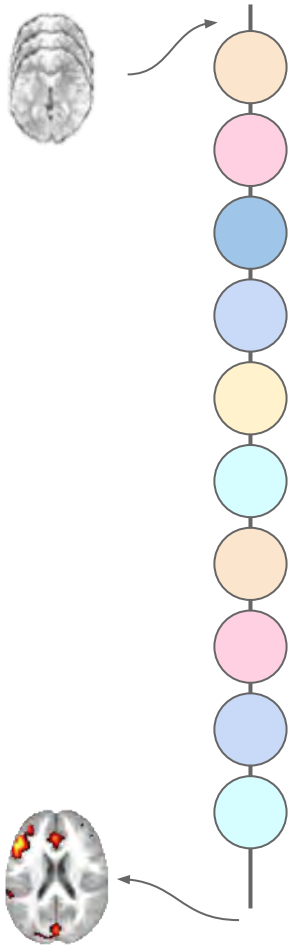
Défi 2 : Travailler avec des données hétérogènes



Variabilité analytique



Variabilité analytique

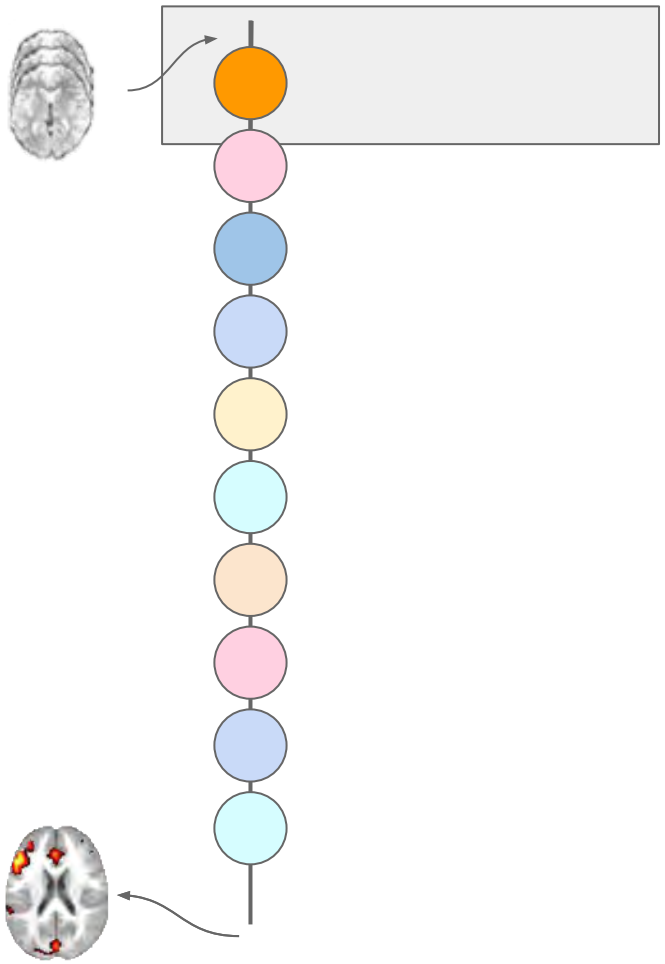


Correction des effets liés au mouvement

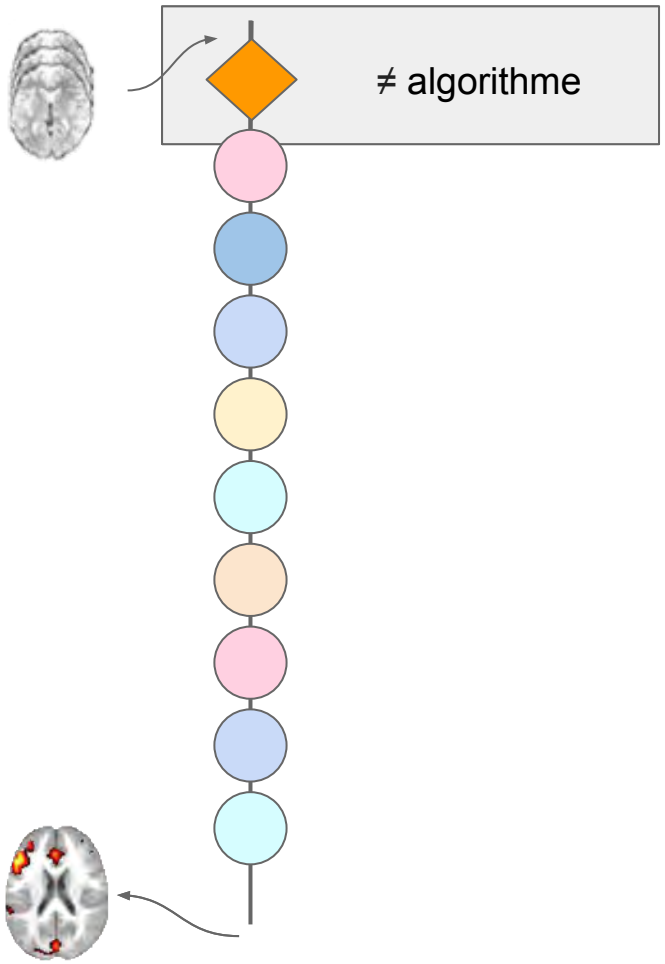
Débruitage

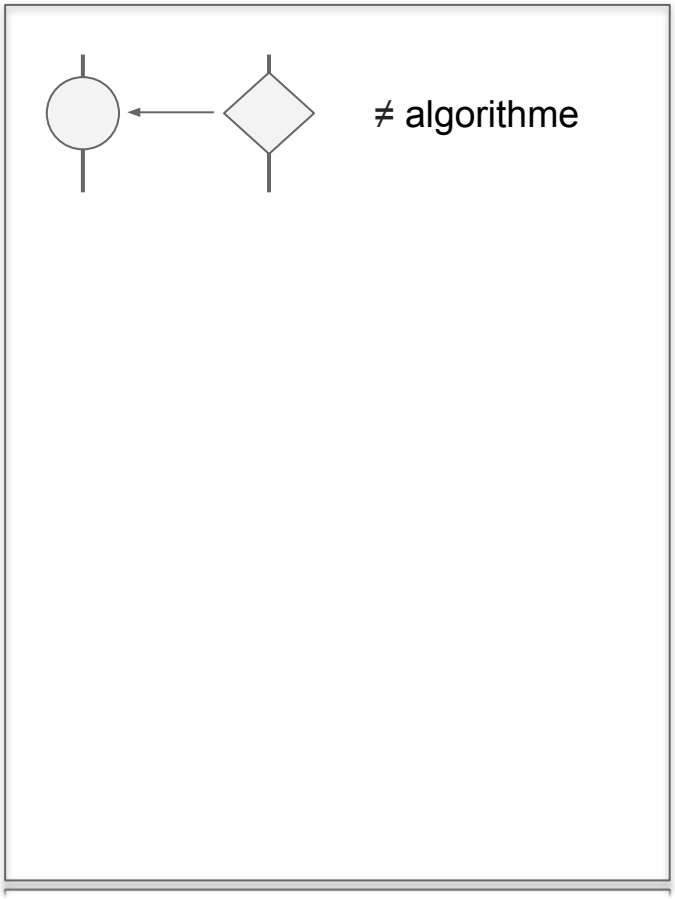
etc.

Variabilité analytique

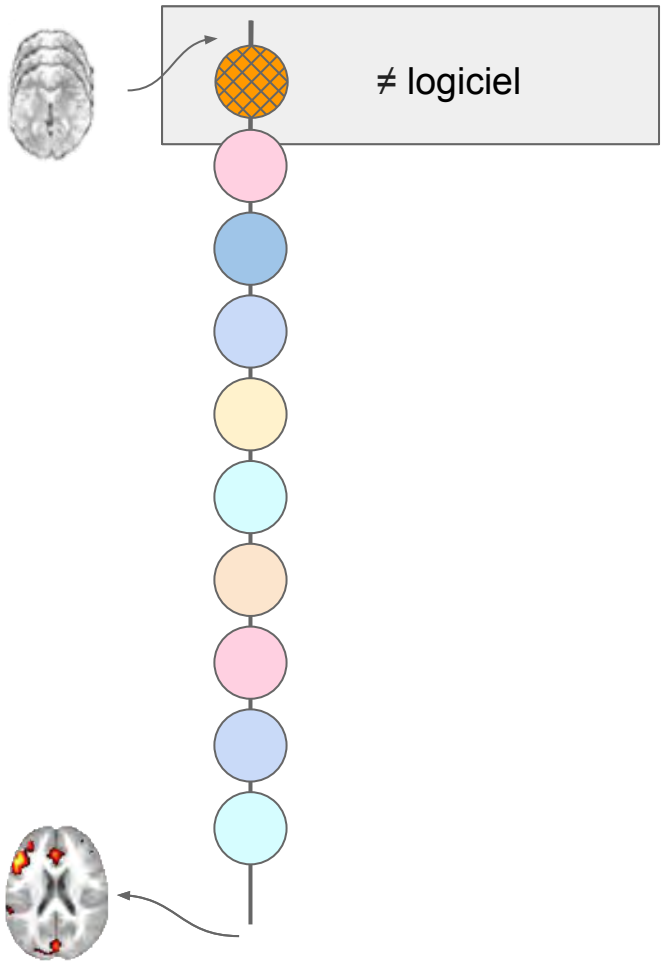


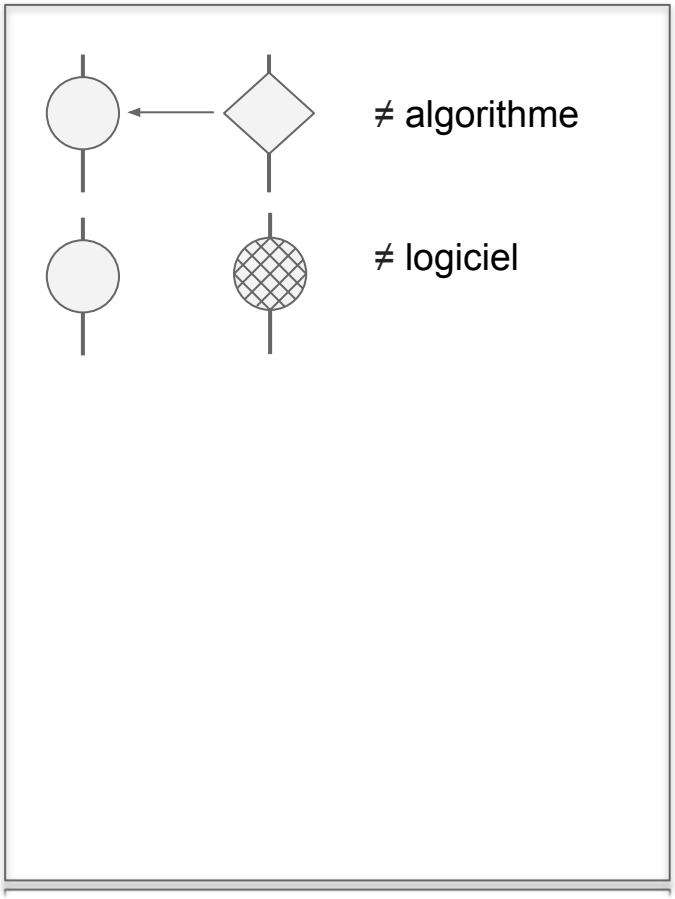
Variabilité analytique



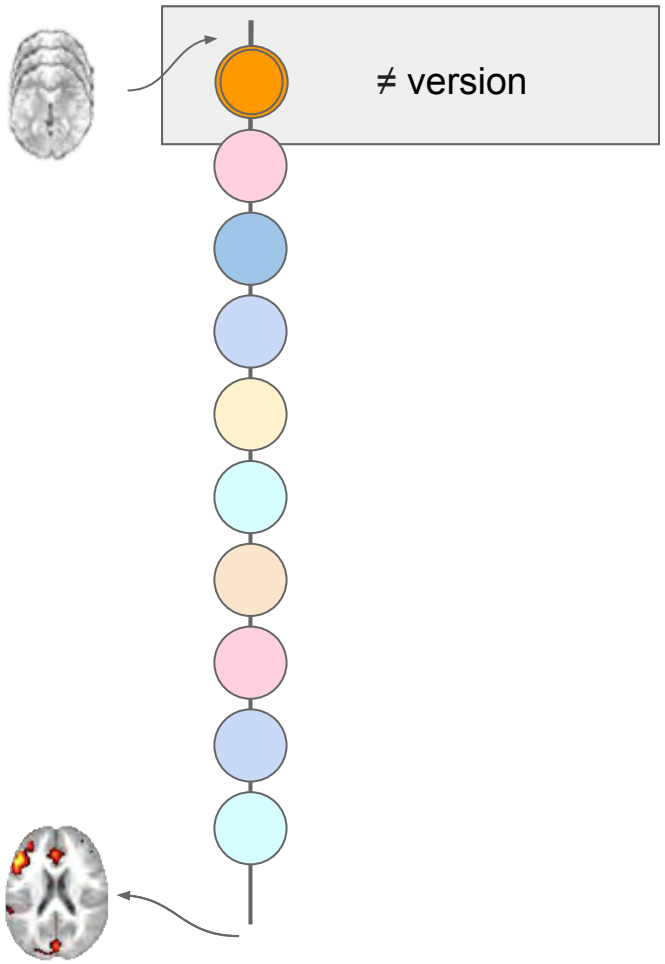


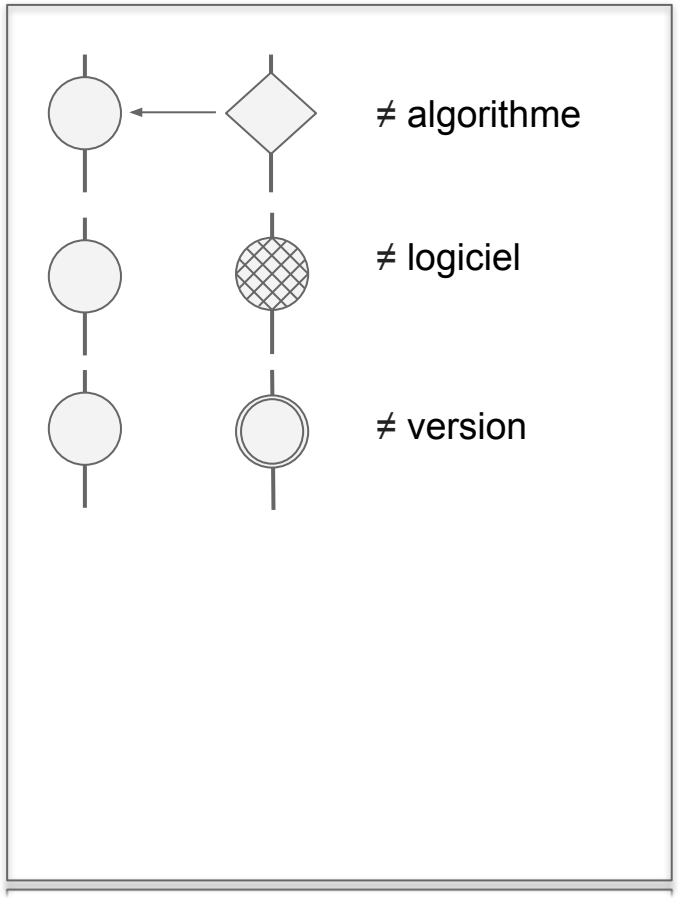
Variabilité analytique



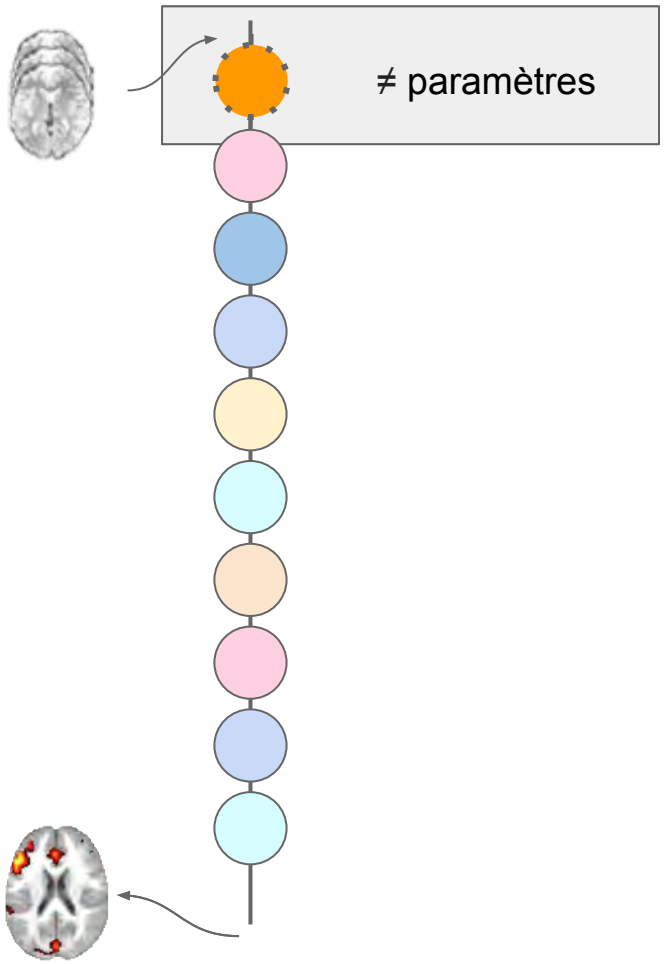


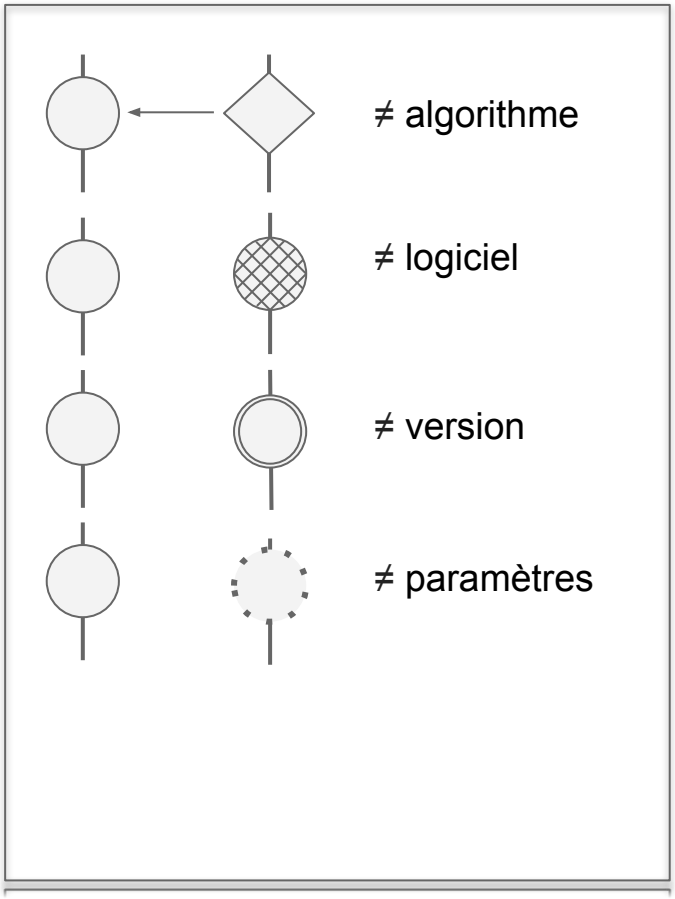
Variabilité analytique



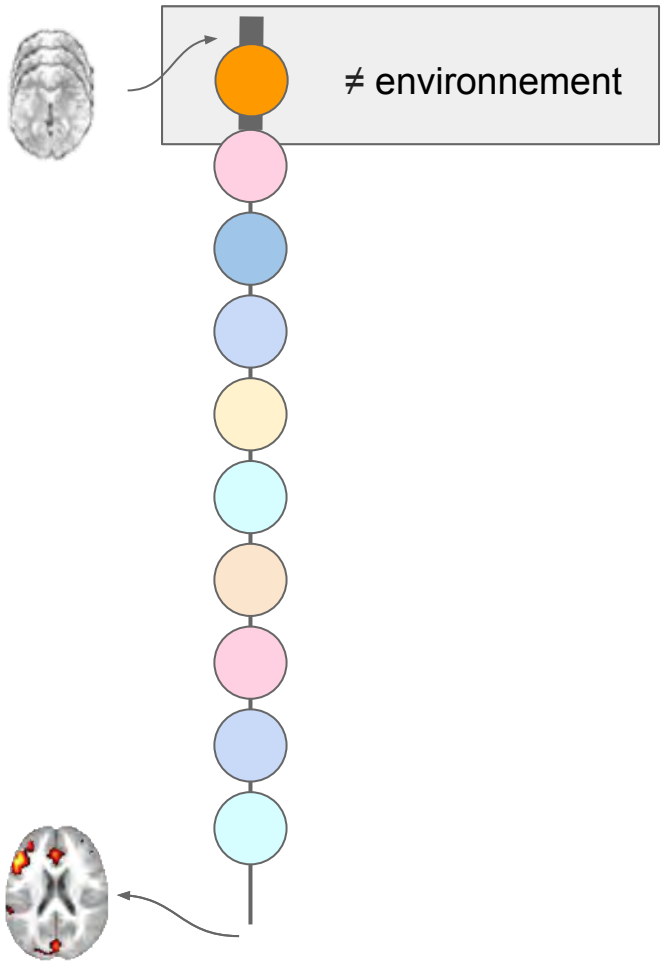


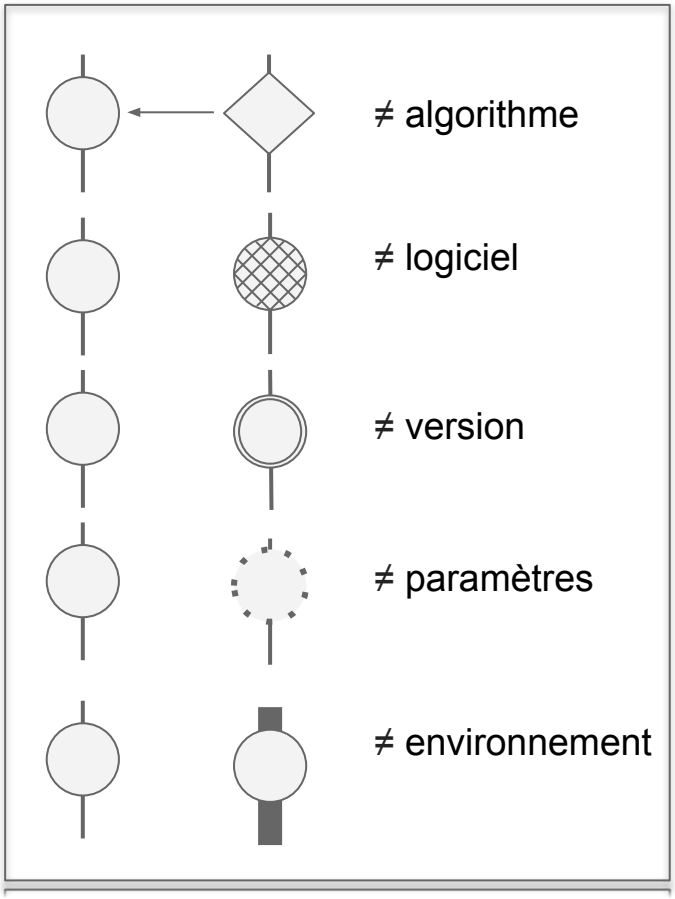
Variabilité analytique



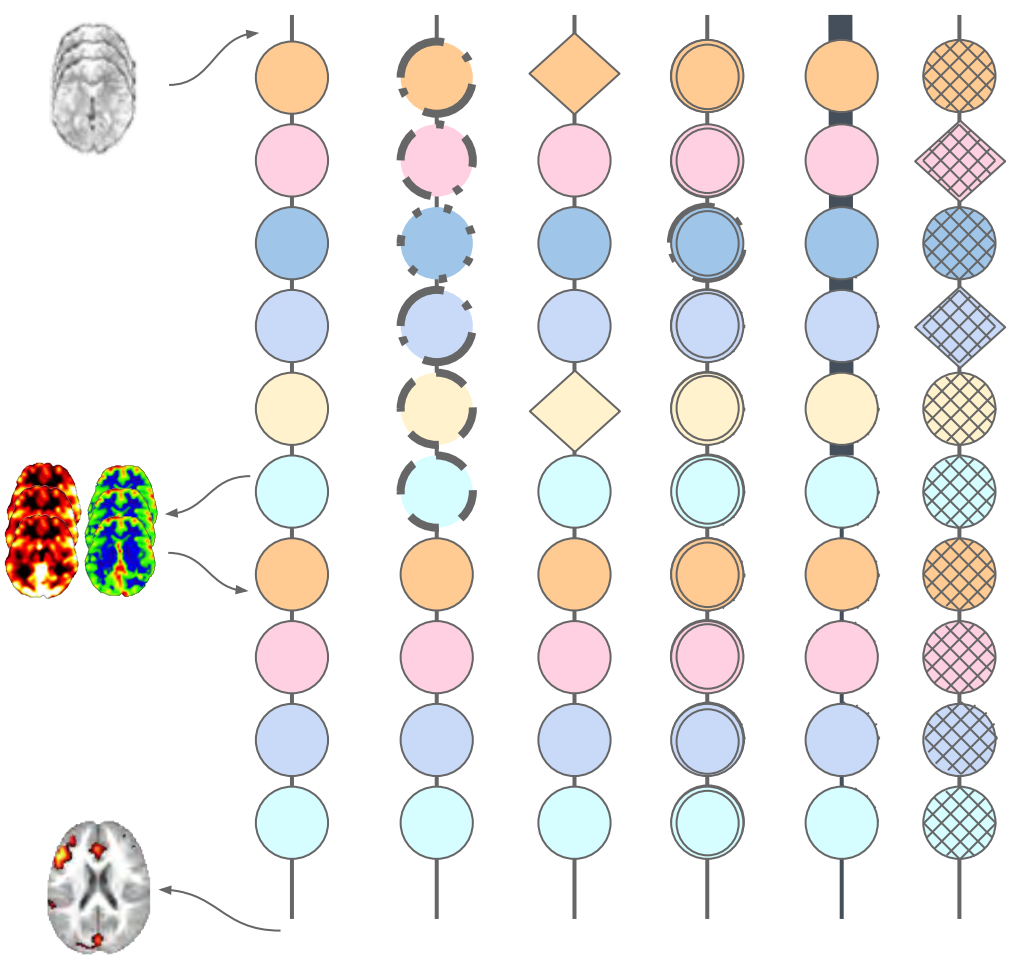


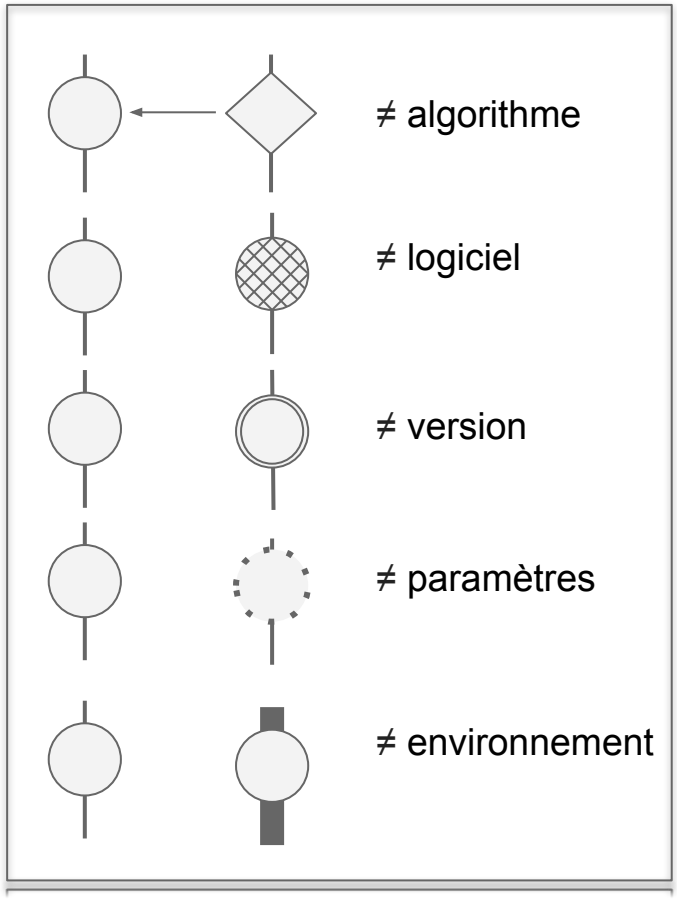
Variabilité analytique





Variabilité analytique





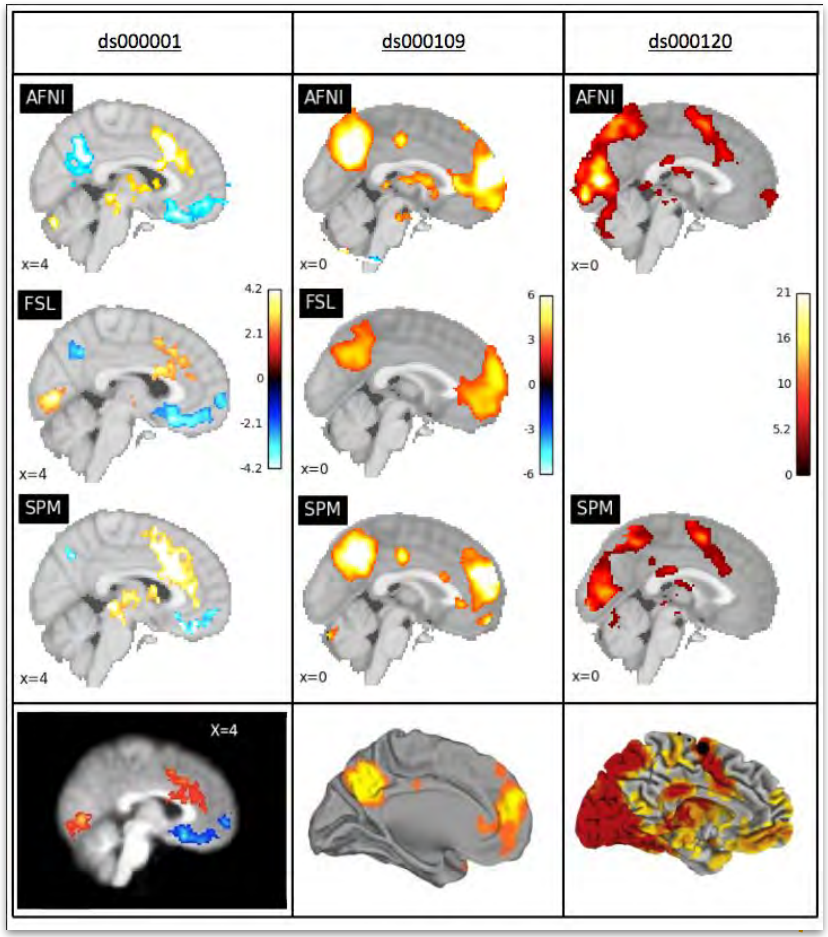
Variabilité analytique en IRM

Liée aux algorithmes
[Carp, Front. Neuroscience 2012]

Liée à la version
[Groenenschild, PlosOne, 2012]

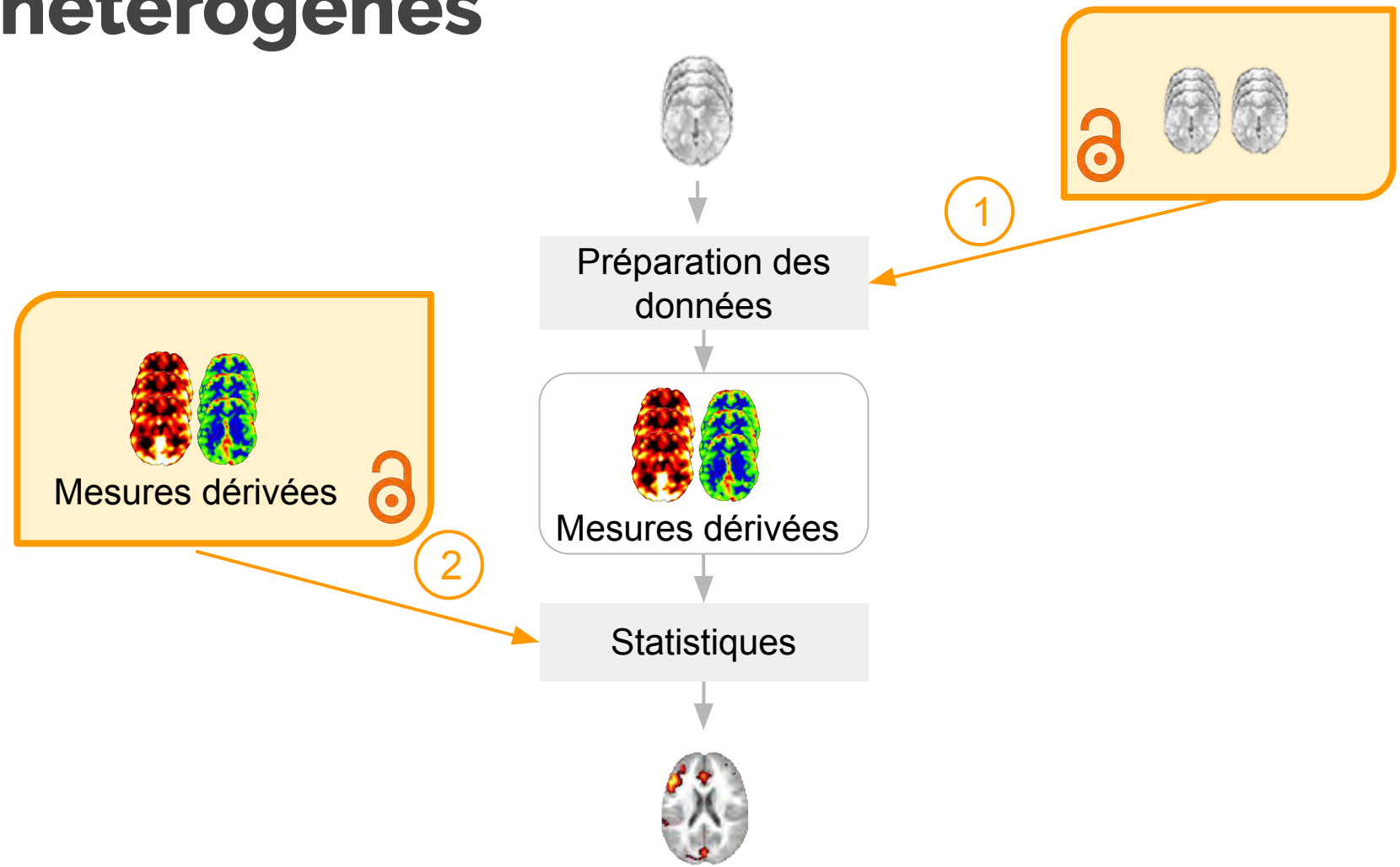
Liée au système d'exploitation
[Glatard, Front. Neuroinformatics 2015]

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



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

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 | cmaumet

https://team.inria.fr/empenn/ | empenn_lab