



**HAL**  
open science

# Data sharing: towards large open datasets in brain imaging

Camille Maumet

► **To cite this version:**

Camille Maumet. Data sharing: towards large open datasets in brain imaging. IABM 2024 - Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale, Mar 2024, Grenoble, France. pp.1-42, 2024. hal-04534429

**HAL Id: hal-04534429**

**<https://inria.hal.science/hal-04534429v1>**

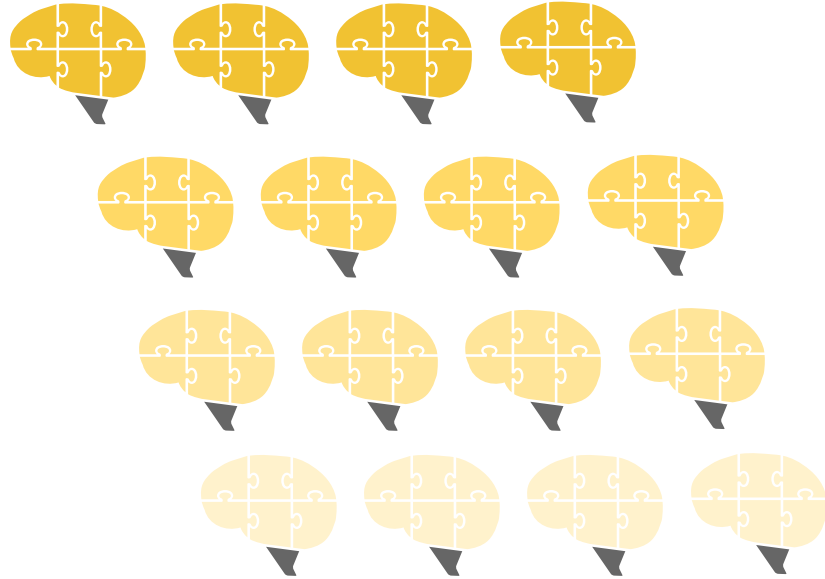
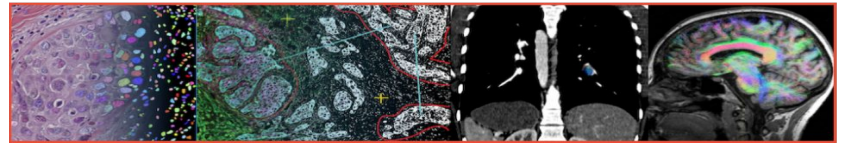
Submitted on 5 Apr 2024

**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



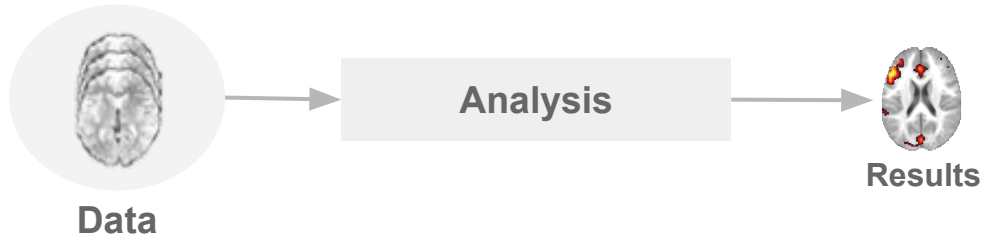
Partage de données : vers de  
larges bases de **données**  
**ouvertes** en neuroimagerie

—  
Data sharing: towards large **open**  
**datasets** in brain imaging

Camille Maumet

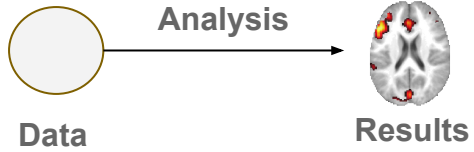
Inria, Univ Rennes, Inserm, CNRS, IRISA

# Reproducible research



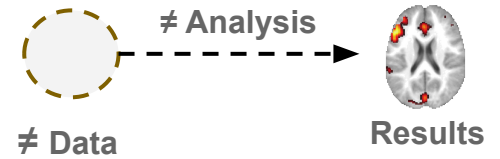
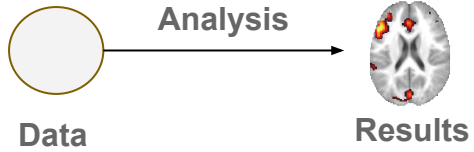
# Reproducible research

Original study



# Reproducible research

Original study



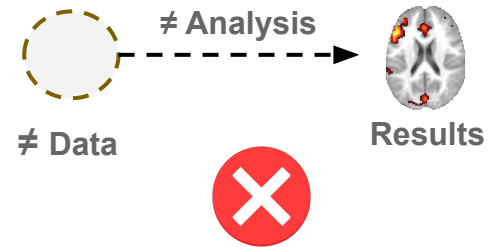
# Reproducible research

## Original study

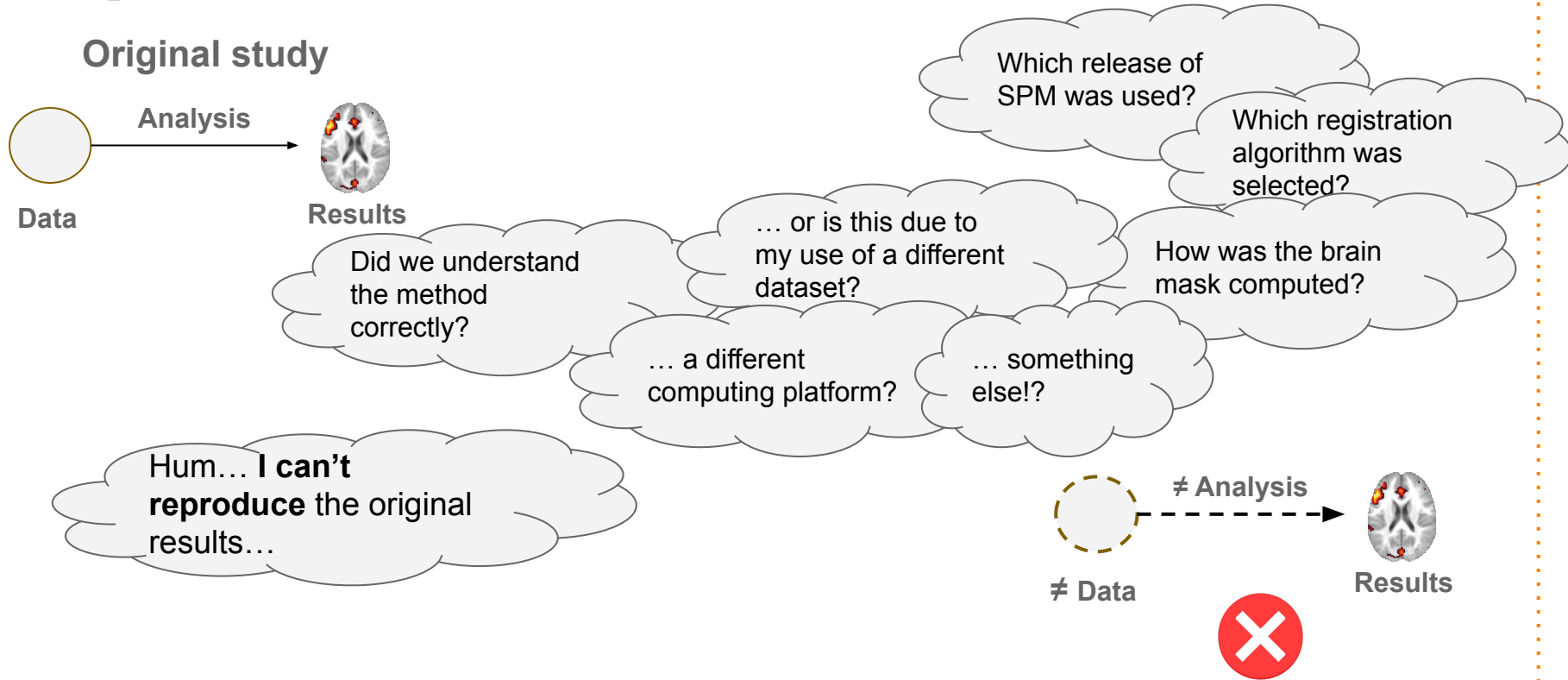


# Reproducible research

Original study

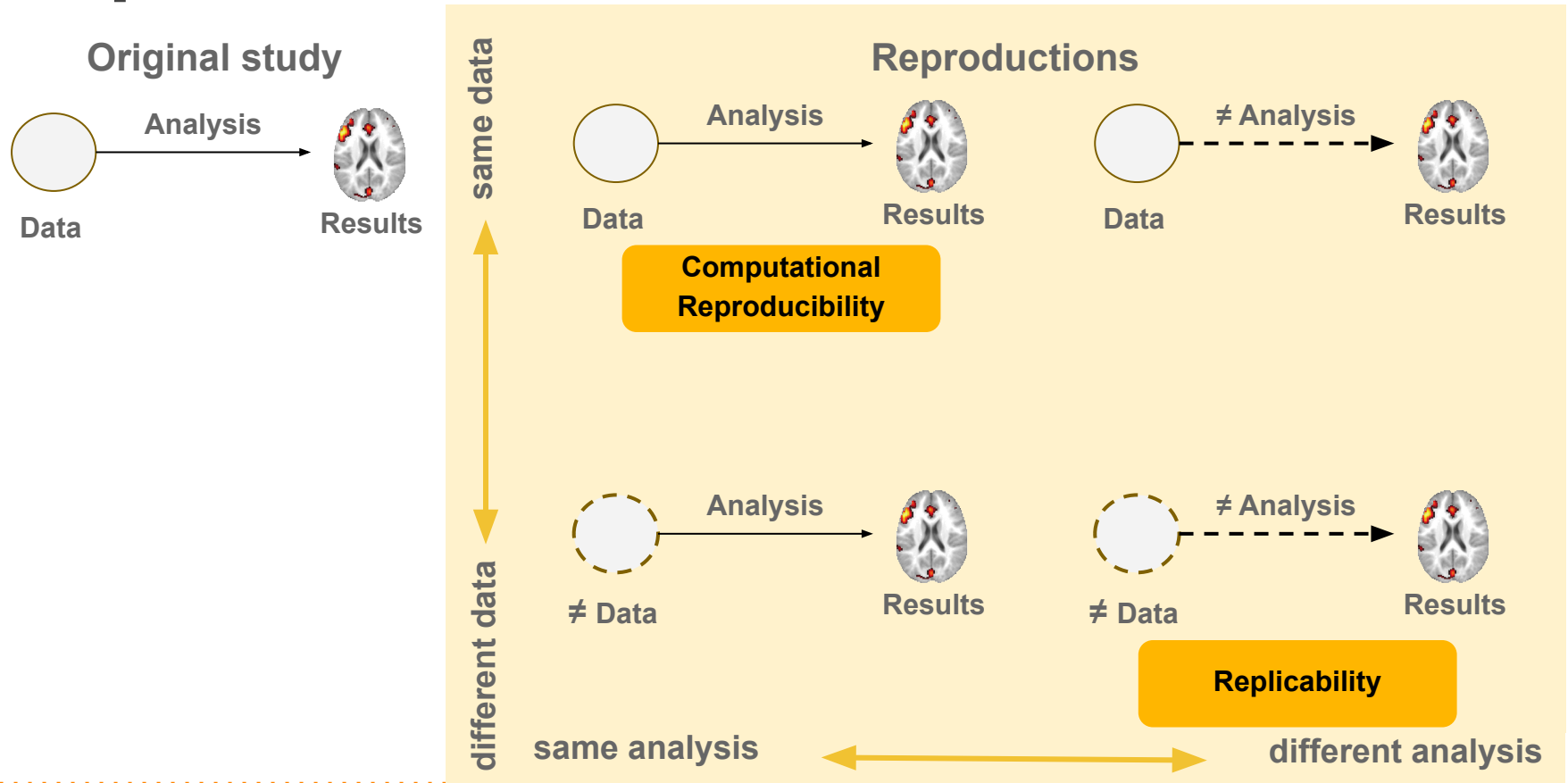


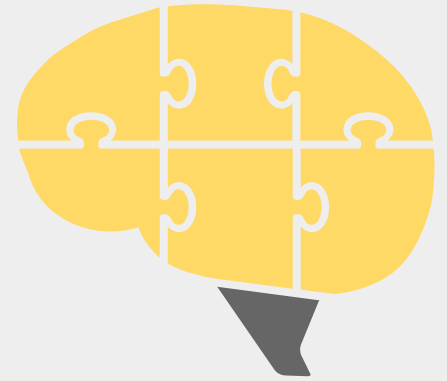
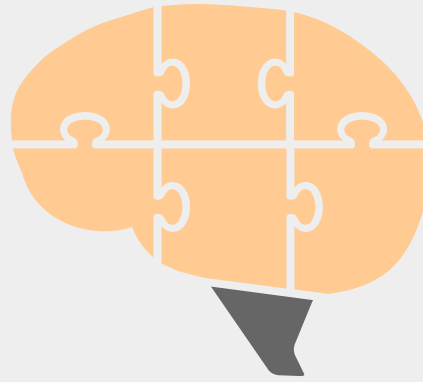
# Reproducible research





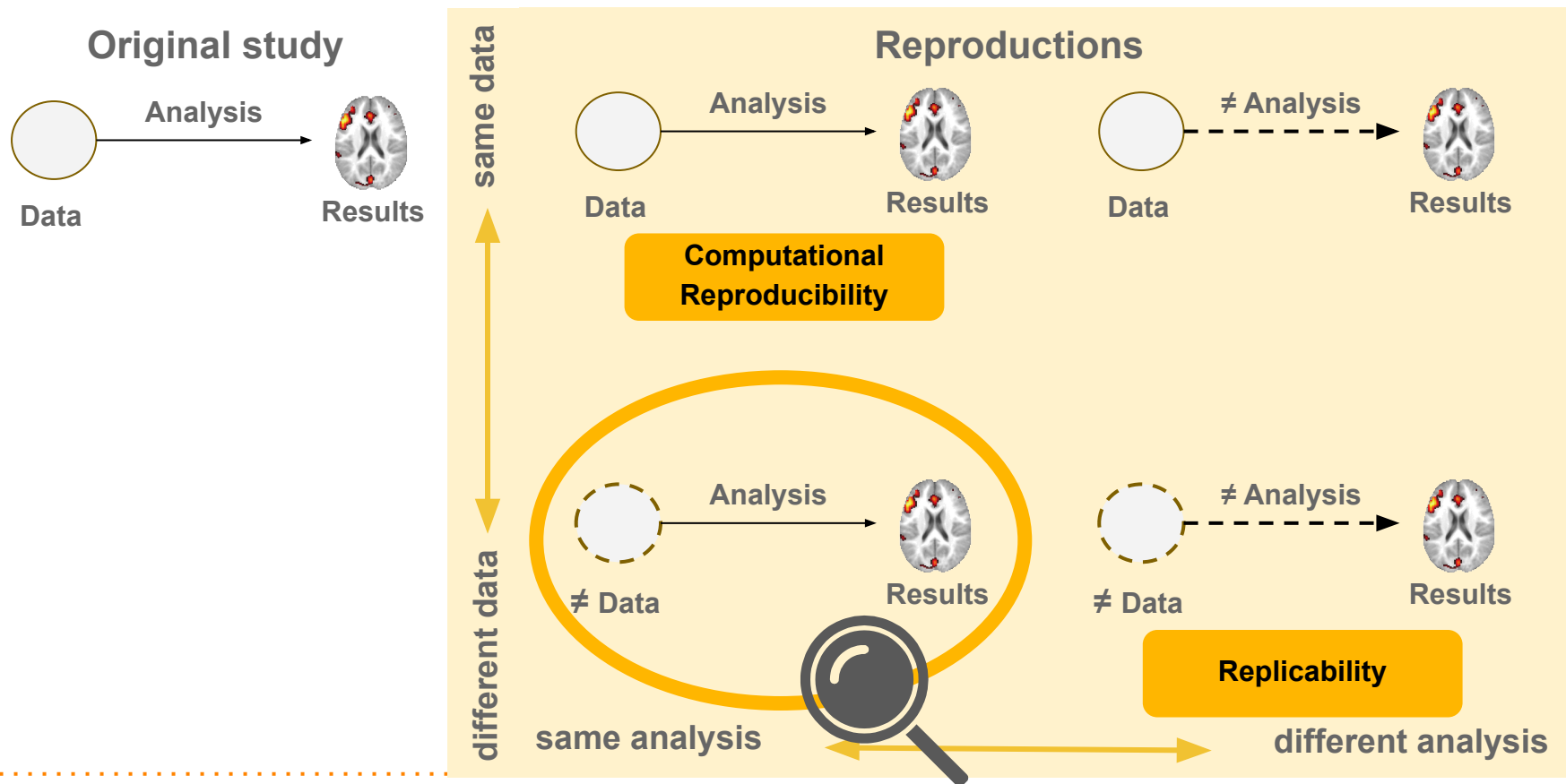
# Reproducible research





On our way to fix  
**reproducibility...**

# Reproducible research

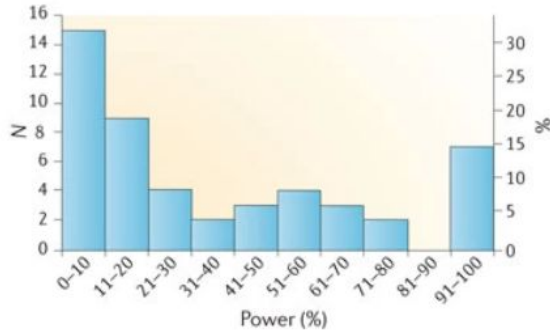


# Fixing reproducibility for different data

## Small samples

## Low statistical power

Power of neuroscience studies



(Button, Nat. Rev. Neurosci. 2013)

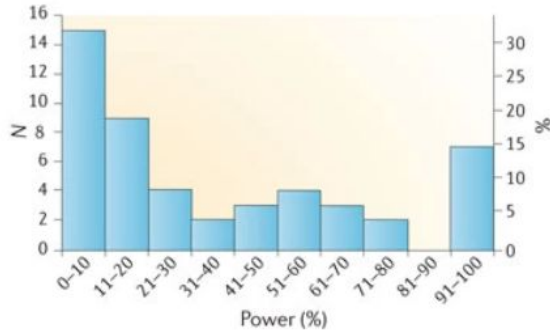
High rate of **false positives** in  
the literature

# Fixing reproducibility for different data

## Small samples

## Low statistical power

Power of neuroscience studies



(Button, Nat. Rev. Neurosci. 2013)

High rate of **false positives** in the literature

## Lack of representativity and diversity of the samples

X-ray: Lung opacity detection

Model trained on male images, tested on female images



(Larrazabal, PNAS 2020)

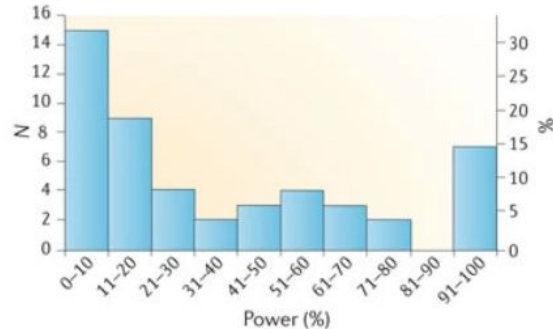
Lack of **generalisability** and **fairness**

# Fixing reproducibility for different data

## Small samples

## Low statistical power

Power of neuroscience studies



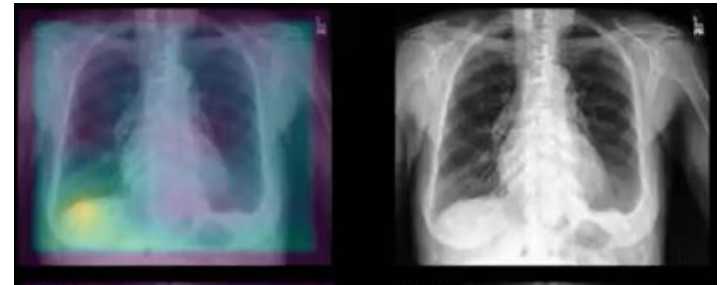
(Button, Nat. Rev. Neurosci. 2013)

High rate of **false positives** in the literature

## Lack of representativity and diversity of the samples

X-ray: Lung opacity detection

Model trained on male images, tested on female images



(Larrazabal, PNAS 2020)

Lack of **generalisability** and **fairness**

**Solution: Data Sharing**

# Neuroimaging data sharing

## Single study

30 sub.



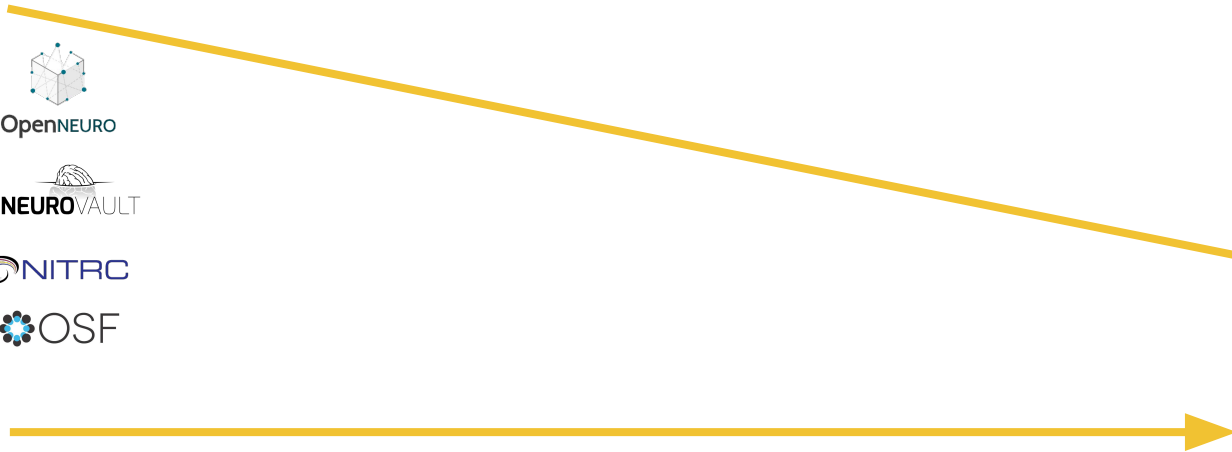
OpenNEURO



NEUROVAULT

NITRC

OSF



# Neuroimaging data sharing

## Single study

30 sub.



OpenNEURO



NEUROVAULT

NITRC



OSF

## Consortium

1000 sub.



ABIDE

Autism Brain Imaging  
Data Exchange



ADHD  
2000



CORR  
CONSORTIUM FOR  
RELIABILITY AND  
REPRODUCIBILITY



1000 Functional  
Connectomes Project



# Neuroimaging data sharing

## Single study

30 sub.



OpenNEURO



NEUROVAULT



## Consortium

1000 sub.



ABIDE  
Autism Brain Imaging  
Data Exchange



CORR  
CONSORTIUM FOR  
RELIABILITY AND  
REPRODUCIBILITY



1000 Functional  
Connectomes Project



HUMAN  
Connectome  
PROJECT



Adolescent Brain Cognitive Development®  
Teen Brains. Today's Science. Brighter Future.

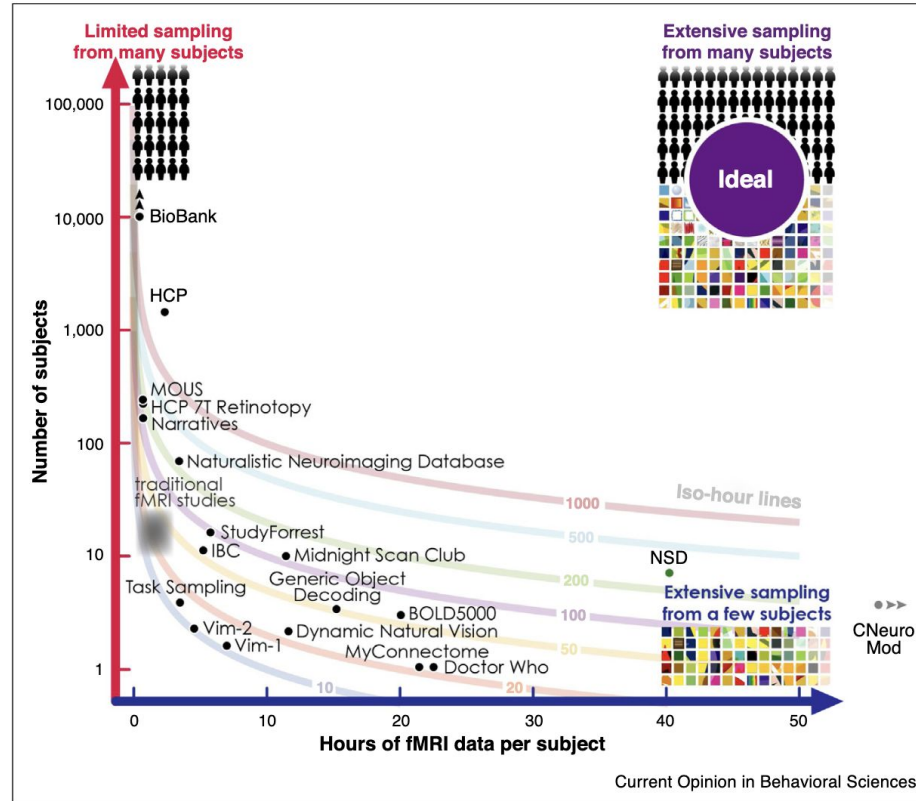
## Cohort

100 000 sub.

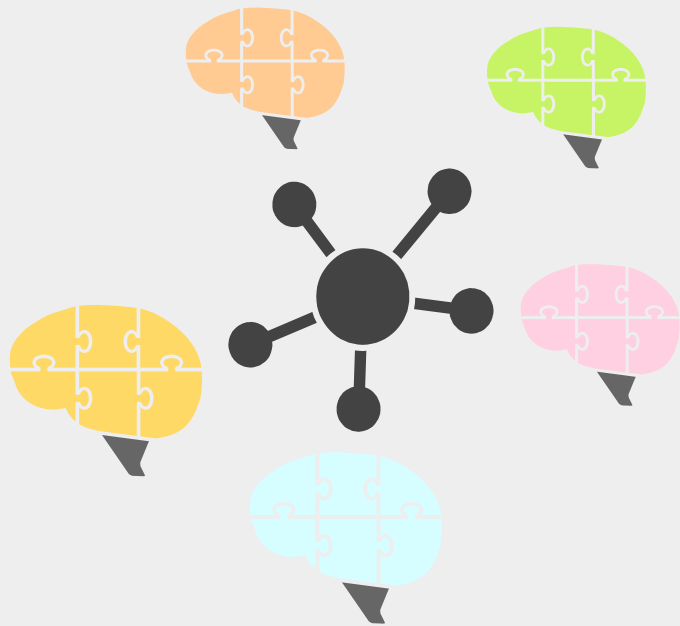


++ participants  
++ homogeneous  
-- instances

# Neuroimaging data sharing



(Nasalaris, Curr. Opin. Behav. Sci. 2021)



# Opening up neuroimaging data

# Sharing FAIR datasets



# Sharing FAIR datasets



OpenNEURO



LORIS

*Donders Repository*



**Findable** we can search for  
open datasets in common  
repositories

# Sharing FAIR datasets



OpenNEURO



LORIS

*Donders Repository*



**Findable** we can search for open datasets in common repositories



**Accessible** Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

# Data sharing and privacy

Are neuroimages anonymous?



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)



Stephan Heunis  
Hosp Rsrch Ctr Jülich  
(Germany)

(Open Brain Consent, HBM 2021)

# Data sharing and privacy

## Are neuroimages anonymous?



Name: Depont  
First : Nathalie  
Birth date : 10/11/1995  
...



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)



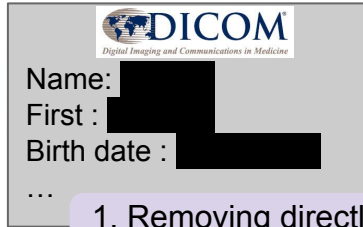
Stephan Heunis  
Hosp Rsrch Ctr Jülich  
(Germany)

(Open Brain Consent, HBM 2021)



# Data sharing and privacy

## Are neuroimages anonymous?



1. Removing directly identifiable information



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)



Stephan Heunis  
Hosp Rsrch Ctr Jülich  
(Germany)

(Open Brain Consent, HBM 2021)

# Data sharing and privacy

## Are neuroimages anonymous?



1. Removing directly identifiable information



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)

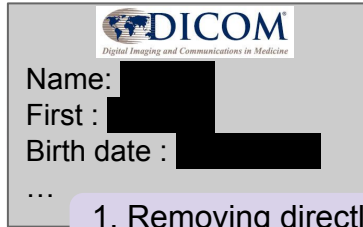


Stephan Heunis  
Hosp Rsrch Ctr Jülich  
(Germany)

(Open Brain Consent, HBM 2021)

# Data sharing and privacy

## Are neuroimages anonymous?



1. Removing directly identifiable information



2. Defacing



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)



Stephan Heunis  
Hosp Rsrch Ctr Jülich  
(Germany)

(Open Brain Consent, HBM 2021)

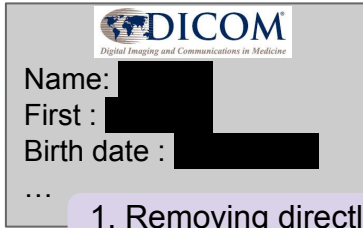
# Data sharing and privacy



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)

Stephan Heunis  
Rsrch Ctr Jülich  
(Germany)

## Are neuroimages anonymous?



1. Removing directly identifiable information



2. Defacing



Anonymous (USA)

(Open Brain Consent, HBM 2021)

# Data sharing and privacy



Cyril Pernet  
Copenhagen Uni Hosp  
(Denmark)

Stephan Heunis  
Rsrch Ctr Jülich  
(Germany)

## Are neuroimages anonymous?



1. Removing directly identifiable information



2. Defacing



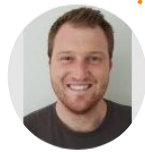
Anonymous (USA)

Not anonymous (EU)

**General Data Protection Regulation (GDPR)** → Individualisation

(Open Brain Consent, HBM 2021)

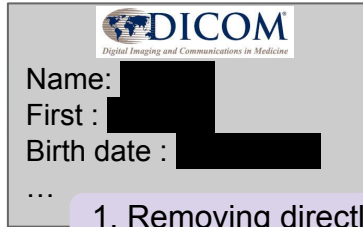
# Data sharing and privacy



Cyril Pernet  
Copenhagen Uni Hosp Rsrch Ctr Jülich  
(Denmark)

Stephan Heunis  
(Germany)

## Are neuroimages anonymous?



1. Removing directly identifiable information



2. Defacing



Anonymous (USA)

Not anonymous (EU)

**General Data Protection Regulation (GDPR)** → Individualisation

## Open Brain Consent GDPR edition

### Consent form

Signed by participants

“[...] **data might also be used for other future research projects** in the field of medical and cognitive neuroscience.”

### Data User Agreement

Signed by researchers

“[...] I will **not attempt to [...] retrieve the identity** of the study participants.”

(Open Brain Consent, HBM 2021)

French



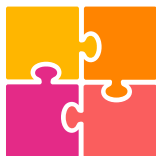
Anne Hespel  
CHU Rennes



Elise Bannier  
CHU Rennes

**13 languages:** Bosnian, Czech, German, Greek, Spanish, Finnish, French, Italian, Dutch, Norwegian, Portuguese, Turkish

# Sharing FAIR datasets



**Interoperable?**



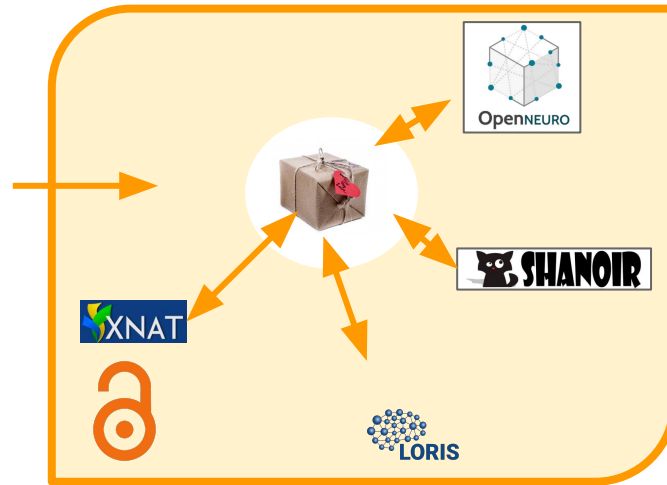
**Findable** we can search for open datasets in common repositories



**Accessible** Direct download (e.g. Openneuro) or with Data User Agreement (e.g. Donders)

# Standards for data sharing

- Data and **meta-data**
- A common “language”





# Community standardization efforts



Chris  
Gorgolewski  
Stanford Uni. (USA)



**Brain Imaging Data Structure**  
File naming conventions + JSON sidecars

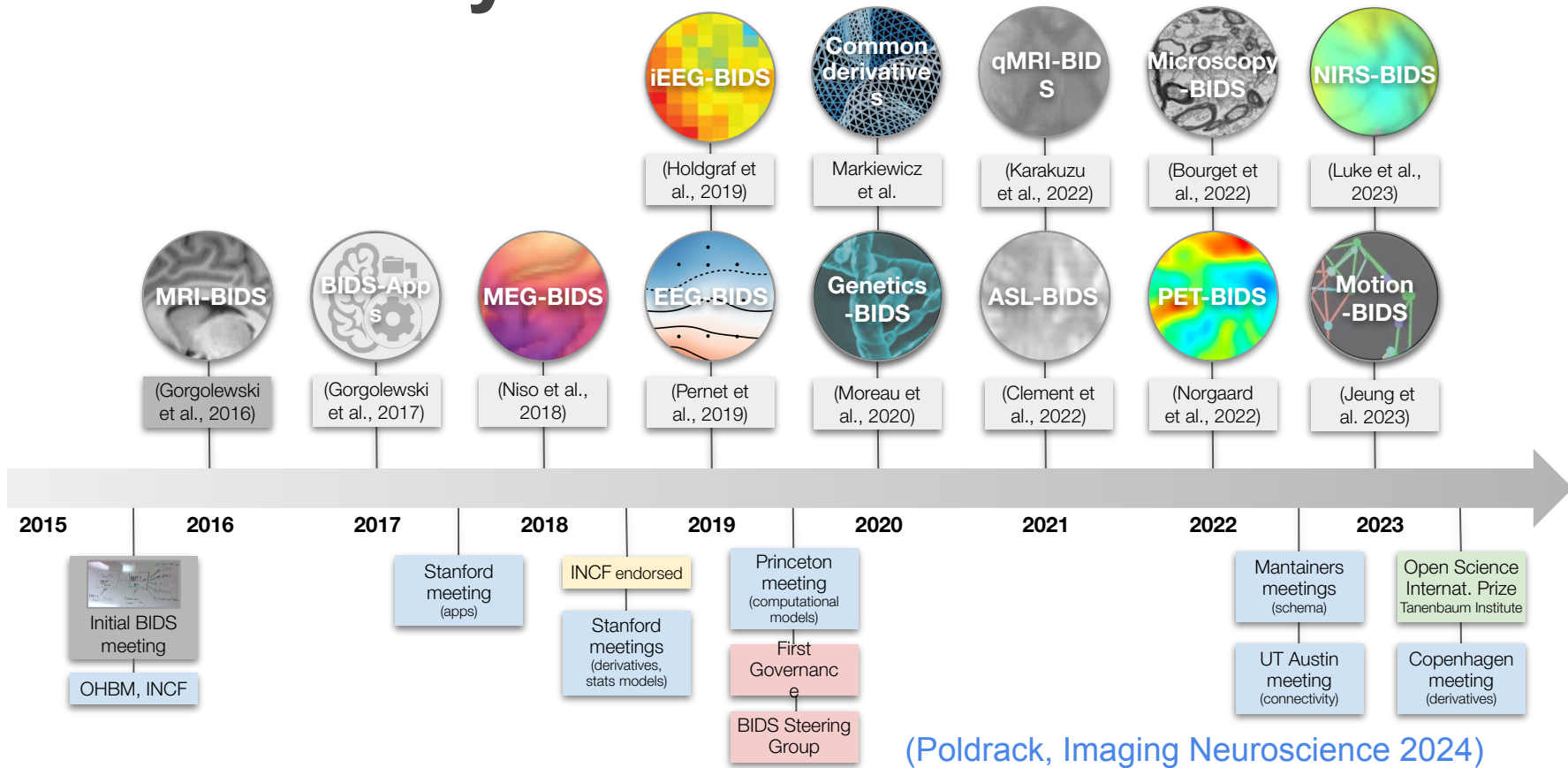
```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   ├── func/  
│   │   ├── sub-01_task-rest_bold.nii.gz  
│   │   └── sub-01_task-rest_bold.json  
│   └── dwi/  
│       ├── sub-01_dwi.nii.gz  
│       ├── sub-01_dwi.json  
│       ├── sub-01_dwi.bval  
│       └── sub-01_dwi.bvec  
├── sub-02/  
└── sub-03/
```

>1 million BIDS-compliant  
images in 2024

Source: [MRIQC Api](#)

(Gorgolewski, Sci. Data 2016)

# BIDS history timeline



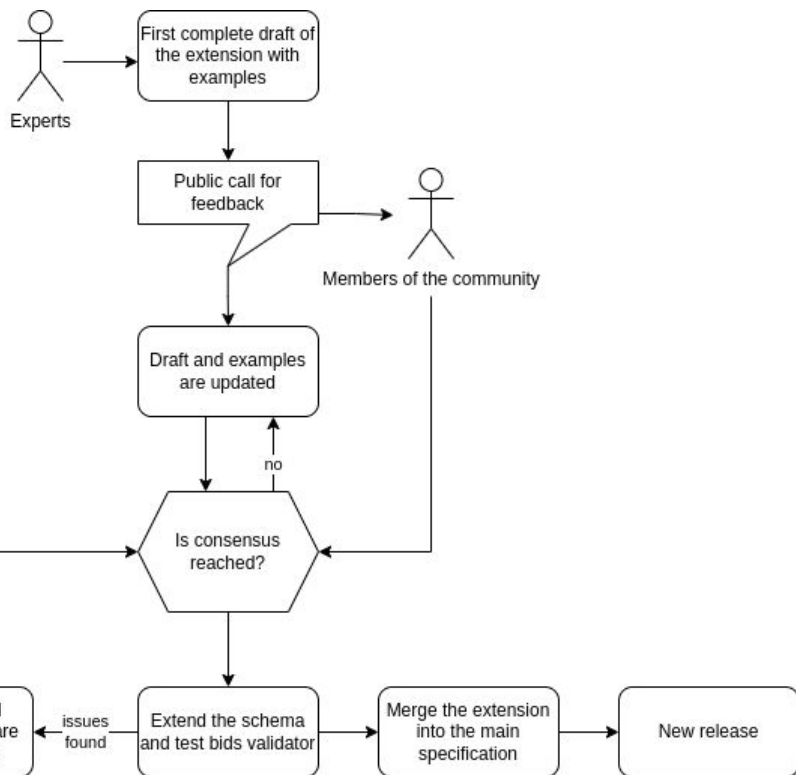
(Poldrack, Imaging Neuroscience 2024)

# Extending BIDS

## BIDS maintainers



## BIDS steering



<https://bids-extensions.readthedocs.io/en/latest/guide/>

# Sharing FAIR datasets

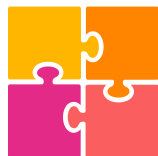


OpenNEURO



LORIS

*Donders Repository*



**Interoperable?**



**Findable** we can search for open datasets in common repositories



**Accessible** Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

# Sharing FAIR datasets

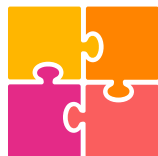


OpenNEURO



LORIS

*Donders Repository*



**Interoperable** BIDS can act as a common languages across tools



**Findable** we can search for open datasets in common repositories



**Accessible** Direct download (e.g. *Openneuro*) or with Data User Agreement (e.g. *Donders*)

# Sharing FAIR datasets

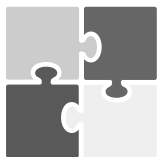


OpenNEURO



LORIS

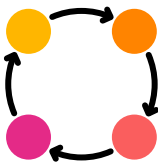
*Donders Repository*



**Interoperable** BIDS can act as a common languages across tools



**Findable** we can search for open datasets in common repositories



**Reusable** Metadata in BIDS were selected to enable reuse. *e.g. onsets of the task for task-based fMRI.*



**Accessible** Direct download (*e.g. Openneuro*) or with Data User Agreement (*e.g. Donders*)

# Sharing FAIR datasets



OpenNEURO

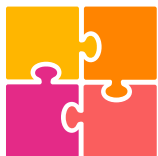


LORIS

*Donders Repository*



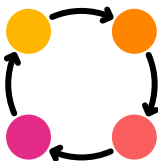
Findable **A**ccessible **I**nteroperable **R**eusable



**Interoperable** BIDS can act as a common languages across tools



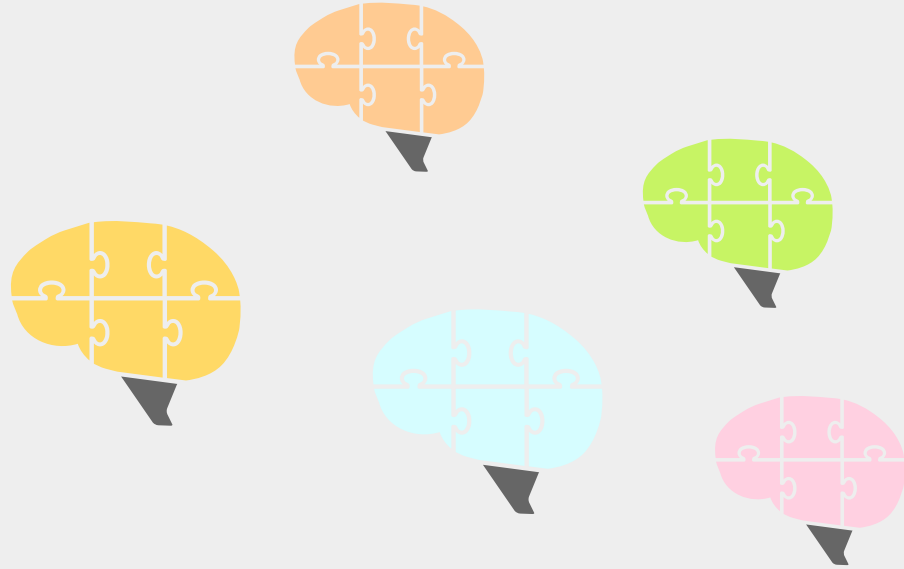
**Findable** we can search for open datasets in common repositories



**Reusable** Metadata in BIDS were selected to enable reuse. *e.g. onsets of the task for task-based fMRI.*



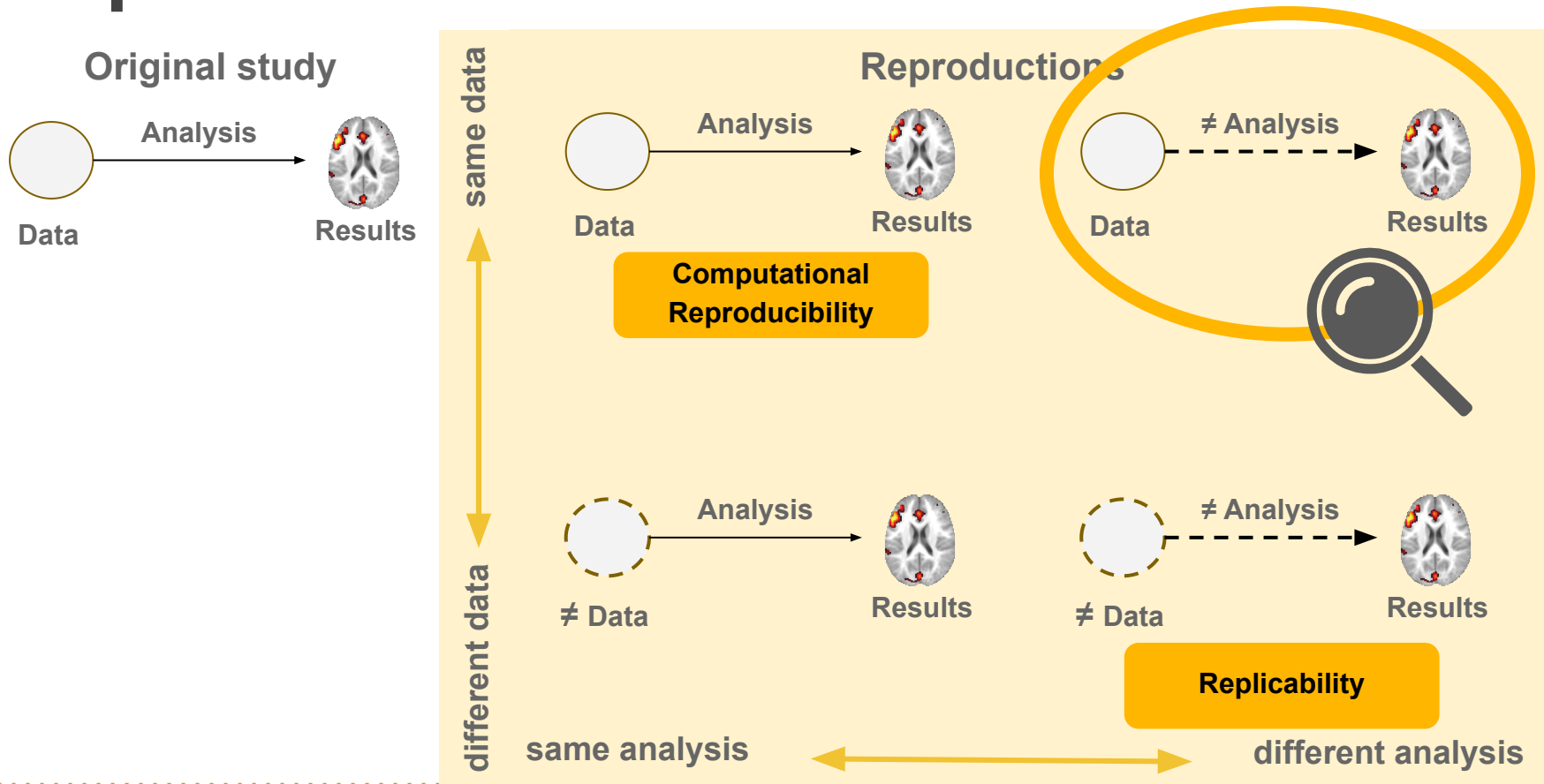
**Accessible** Direct download (*e.g. Openneuro*) or with Data User Agreement (*e.g. Donders*)



**More on** reproducibility...



# Reproducible research

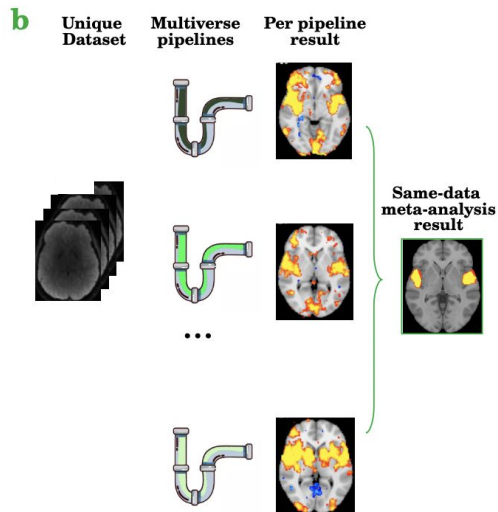


# Dealing with analytical variability

## Multiverse analyses



Jérémie  
Lefort-Besnard

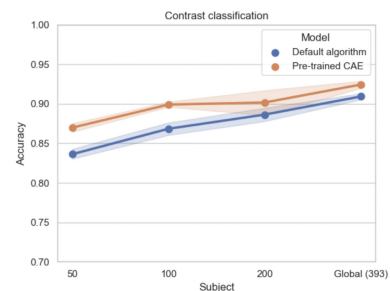
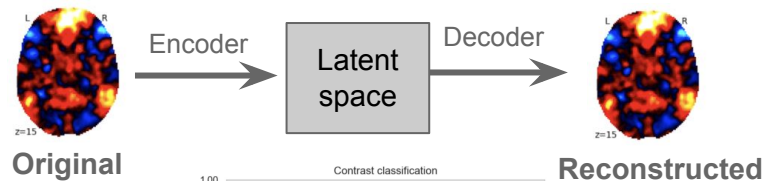


(Lefort-Besnard, OHBM 2024)

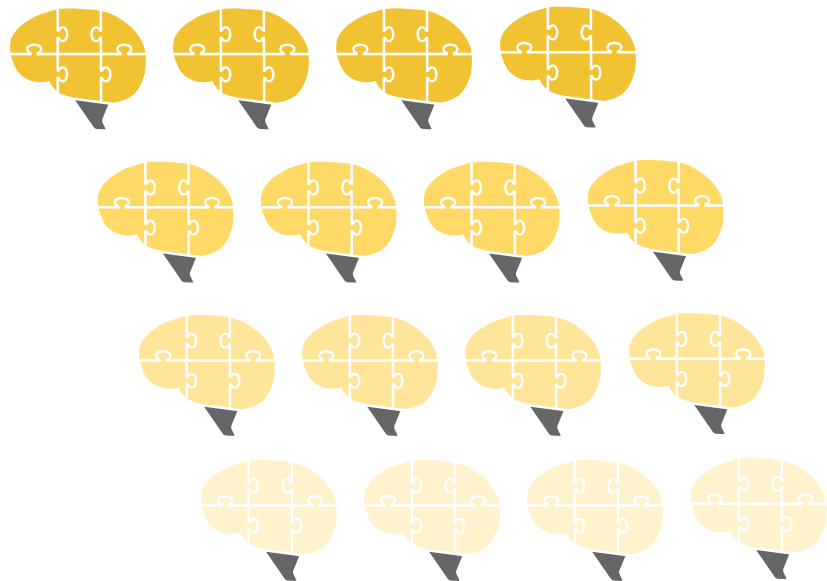
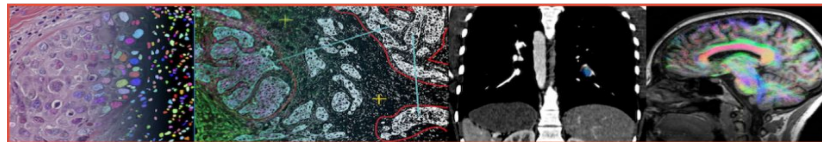
## Leveraging large and open datasets



Elodie  
Germani



(Germani, Gigascience 2023)



Data sharing towards  
building large **open**  
**datasets** in brain imaging

**Camille Maumet**

Thank you!

**Supported by:** Inria Aex, Région Bretagne, ANR JCJC

**Credit:** Presentation template and icons by  
[SlidesCarnival](#), adapted

 [@cmaumet@fediscience.org](mailto:@cmaumet@fediscience.org)

 [@cmaumet.bsky.social](#)