



**HAL**  
open science

## Données ouvertes : comment permettre leur réutilisation?

Camille Maumet

### ► To cite this version:

Camille Maumet. Données ouvertes : comment permettre leur réutilisation?. REMI 2021 - Rencontres du Réseau d'Entraide Multicentrique en IRM, Mar 2021, Online, France. pp.1-51. inserm-03174616

**HAL Id: inserm-03174616**

**<https://inria.hal.science/inserm-03174616>**

Submitted on 19 Mar 2021

**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



# Données ouvertes : comment permettre leur réutilisation?

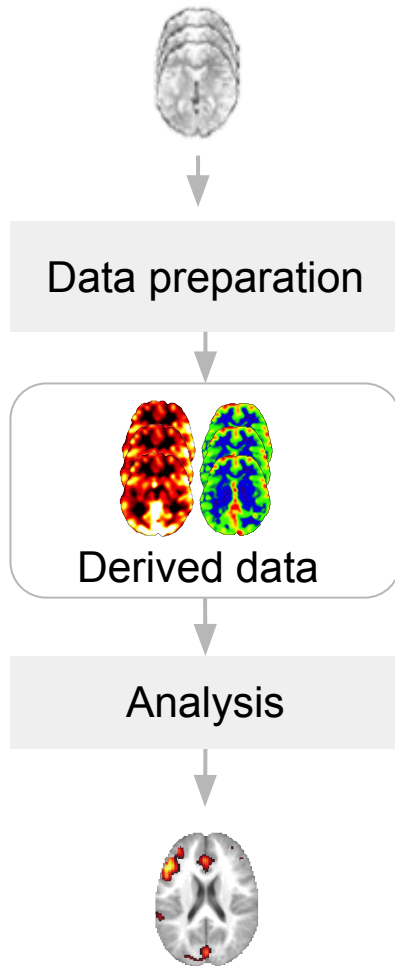


**Camille Maumet**

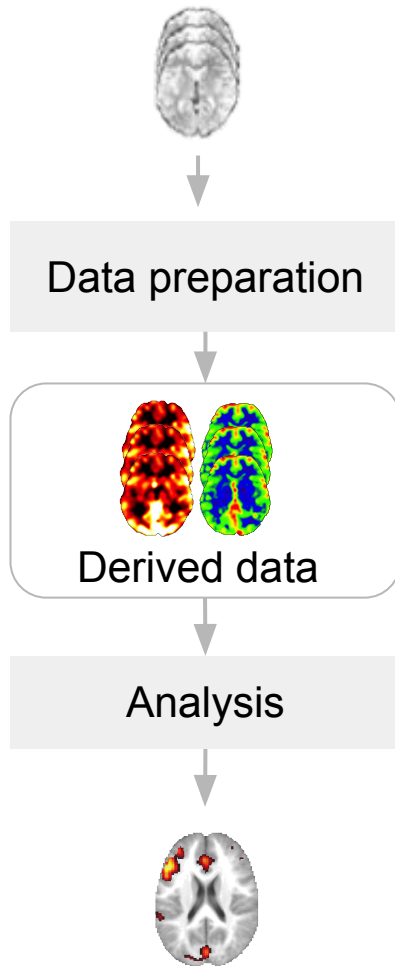
Univ Rennes, Inria, CNRS, Inserm



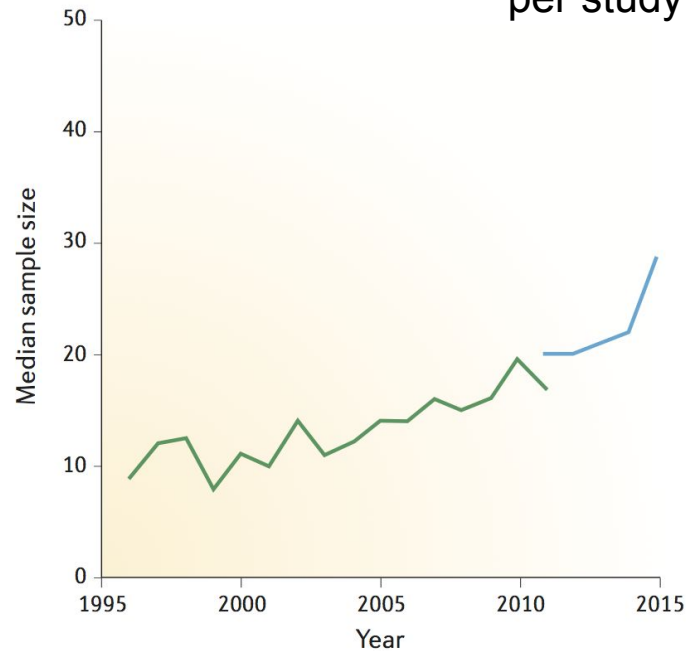
# A brain imaging study



# A brain imaging study



about 30 participants  
per study



(Poldrack et. al, Nat. Neurosci. 2017)

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

**(Ir)-Reproducibility**

## 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.

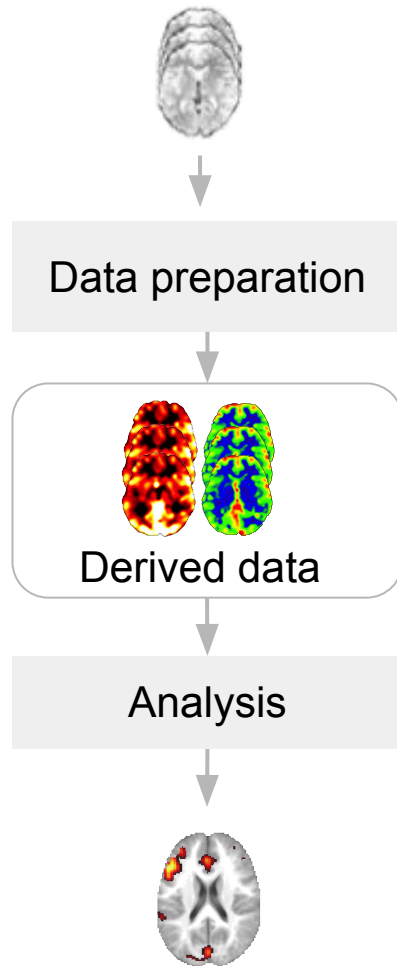
---

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**

# A brain imaging study

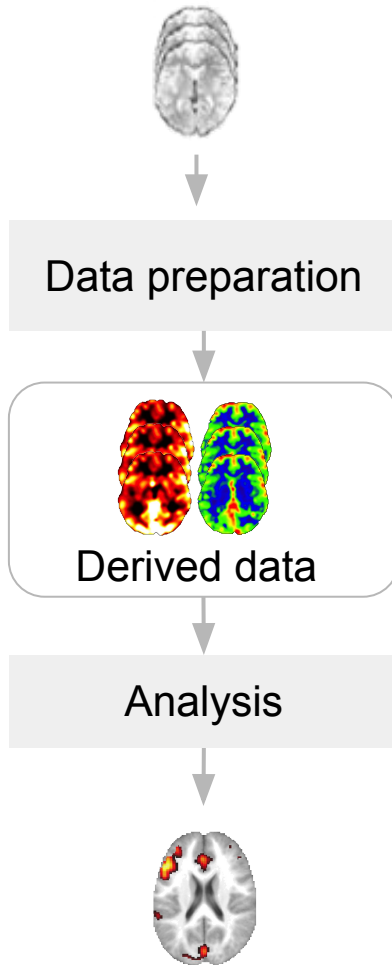


We need **bigger** & more **representative** samples

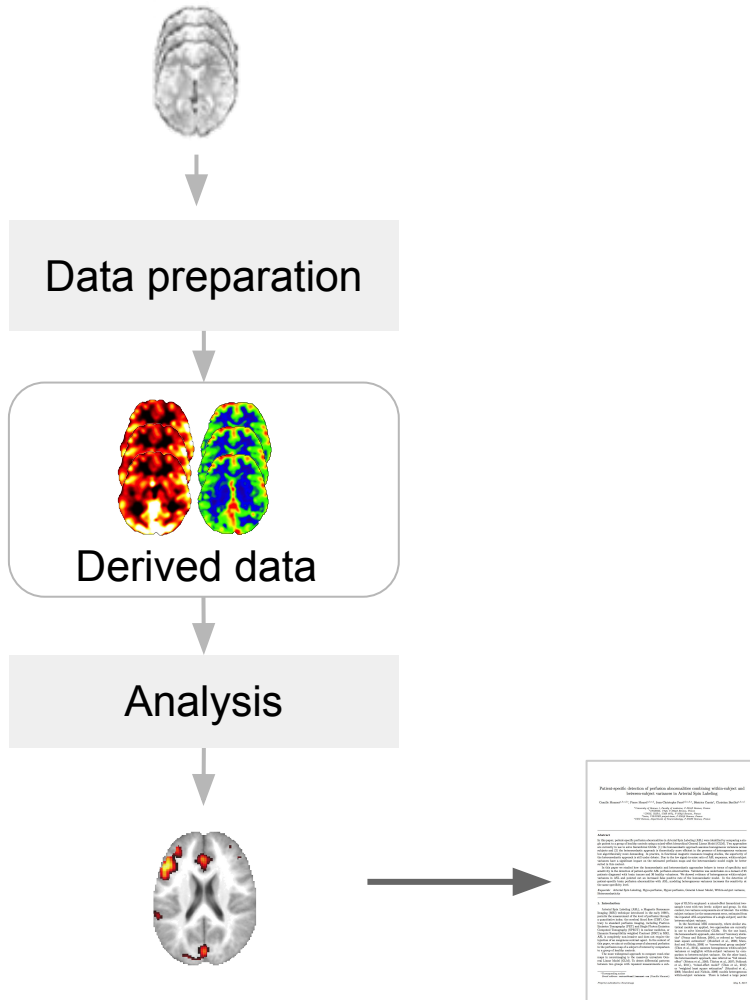
# **Data sharing in** Europe



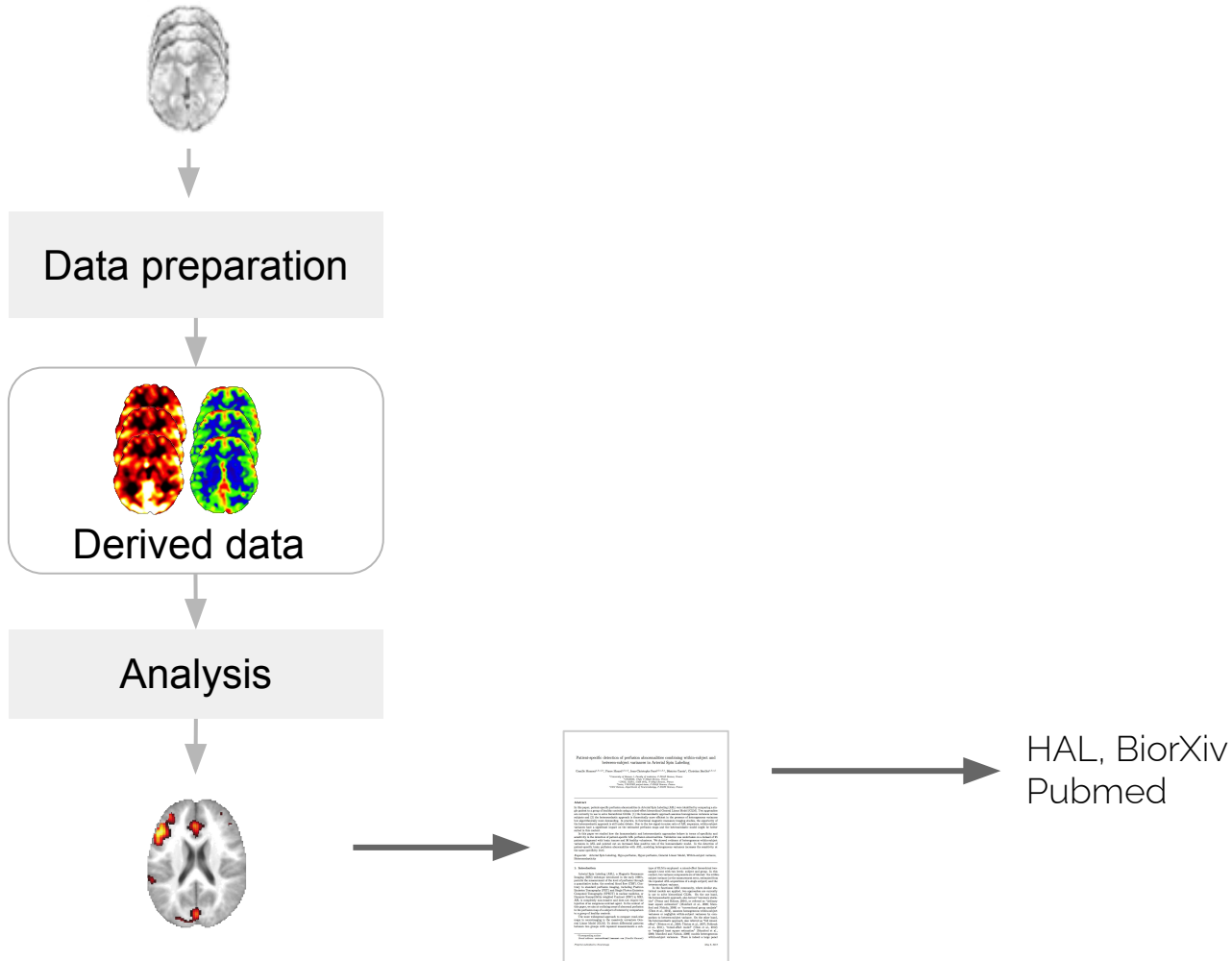
# Neuroimaging data sharing



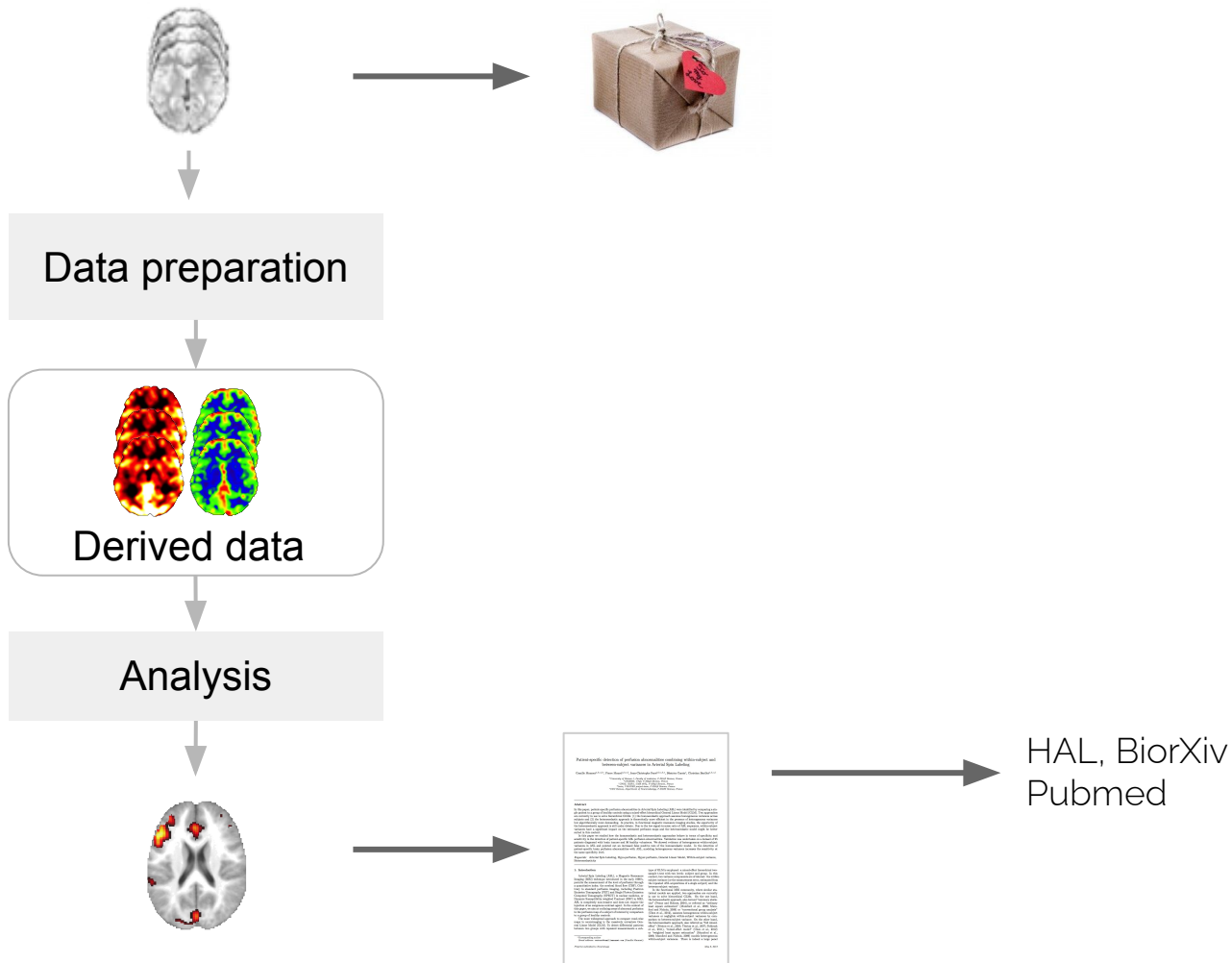
# Neuroimaging data sharing



# Neuroimaging data sharing



# Neuroimaging data sharing



# Brain Imaging Data Structure (BIDS)

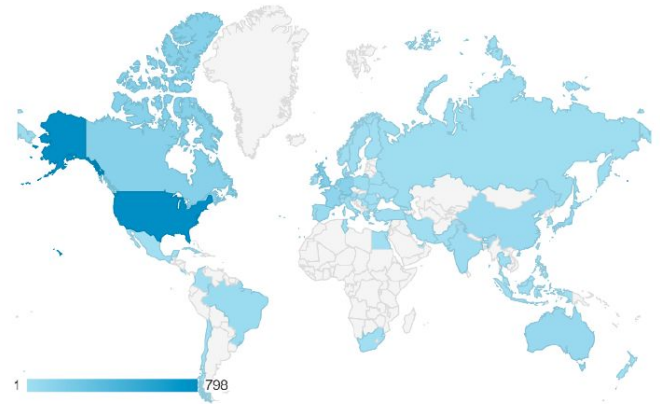
- Used in over 60 labs around the world
- Adopted by: FCP-INDI, Developing Human Connectome, SchizConnect and Donders Data repository.
- Extensions: MEG, iEEG, EEG



Krys  
Gorgolewski

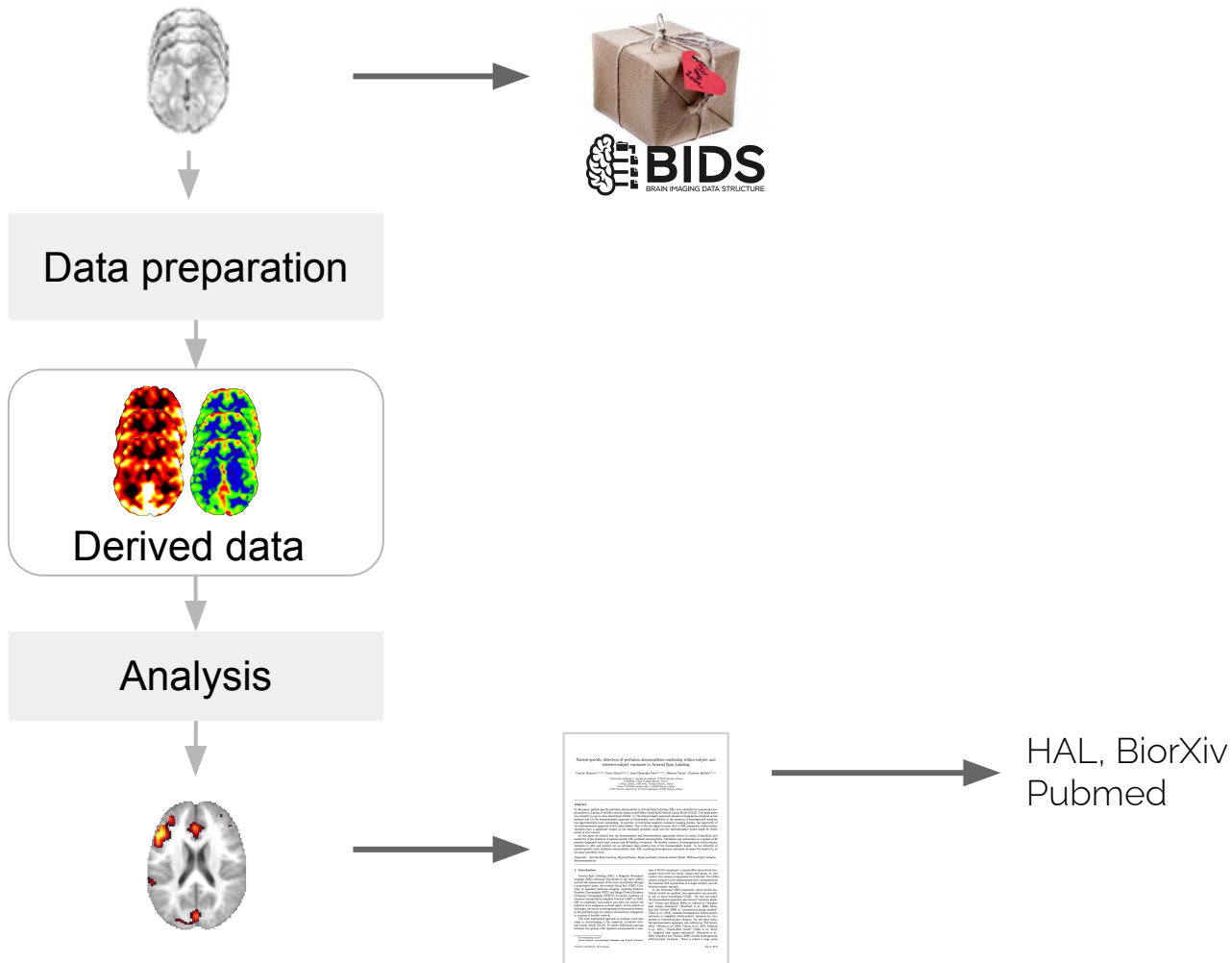
(Gorgolewski et al., *Sci. Data* 2016)

Traffic to [bids.neuroimaging.io](https://bids.neuroimaging.io)

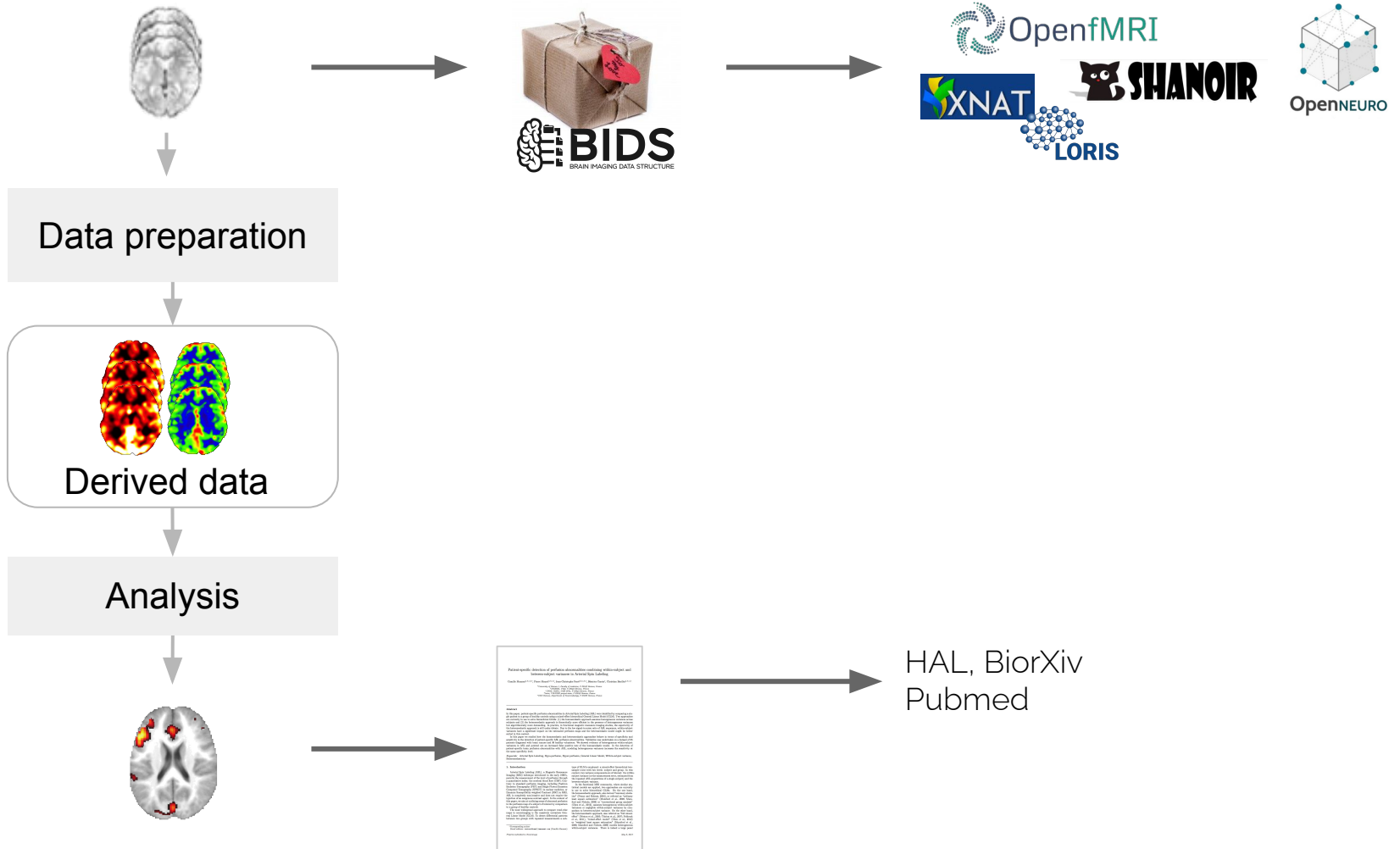


```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   └── func/  
│       ├── sub-01_task-rest_bold.nii.gz  
│       └── sub-01_task-rest_bold.json
```

# Neuroimaging data sharing



# Neuroimaging data sharing



# Data sharing \*and\* privacy



**Open Brain Consent:** Consent forms & Data User Agreement that enable data sharing of neuroimaging data in compliance with GDPR.

(Banner et al., HBM 2020)



Stephan Heunis



Yarik Halchenko

## Translations

- [Bosnian](#)
- [Czech](#)
- [German](#)
- [Greek](#)
- [Spanish](#)
- [Finnish](#)
- [French](#)
- [Italian](#)
- [Dutch](#)
- [Norwegian](#)
- [Turkish](#)

The screenshot shows the Open Brain Consent website interface. On the left, there is a navigation menu with options: Sample consent forms, Recommendations, Ultimate consent form, GDPR edition, Anonymization tools, Contribute, Contact information, and Discussions. Below the menu is a DigitalOcean advertisement. The main content area displays the title 'Make open data sharing a no-brainer for ethics committees.' with a Zenodo badge and a 'Statement of the problem' section. The statement discusses the challenges of open and reproducible science in neuroimaging, particularly regarding privacy and the lack of a common consent form template.



# Data sharing \*and\* privacy



## Formulaire de consentement

[Lien direct](#)

À signer par les participant.e.s.

Version française



Anne Hespel



Elise Bannier

“[...] Il vous est proposé de participer à une étude. Dans ce cadre, des données à caractère personnel vous concernant seront collectées, utilisées et stockées. Si vous l’acceptez, **vos données peuvent également être utilisées pour d’autres projets de recherche futurs dans le domaine des neurosciences médicales et cognitives dans une mission d’intérêt public.** [...]”

## License d'utilisation

[Lien direct](#)

À signer par les scientifiques souhaitant accéder aux données

“[...] 2. **Je ne poursuivrai jamais**, par quelque moyen que ce soit, la finalité **visant à établir ou à retrouver l'identité des personnes** concernées par le traitement des données à caractère personnel, ci-devant les participants à l'étude [...]

3. Je ne diffuserai pas ces données.  
[...]

# Shared (open) data



Unique study  
30 participants



OpenNEURO

studyforrest.org



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

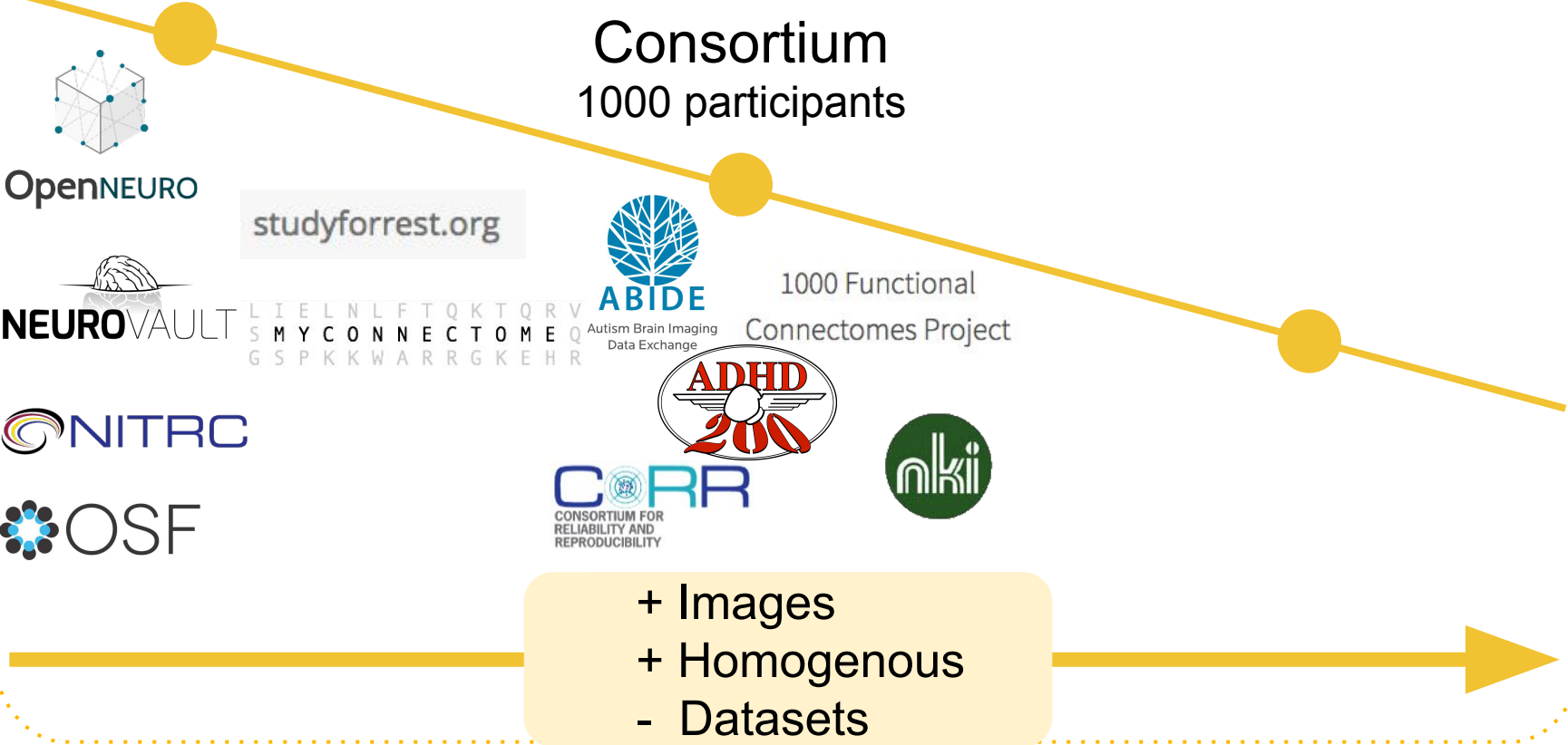
+ Images  
+ Homogenous  
- Datasets



# Shared (open) data

Unique study  
30 participants

Consortium  
1000 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

ABIDE  
Autism Brain Imaging  
Data Exchange

1000 Functional  
Connectomes Project

ADHD  
200

NITRC

OSF

CORR  
CONSORTIUM FOR  
RELIABILITY AND  
REPRODUCIBILITY

nkii

+ Images  
+ Homogenous  
- Datasets

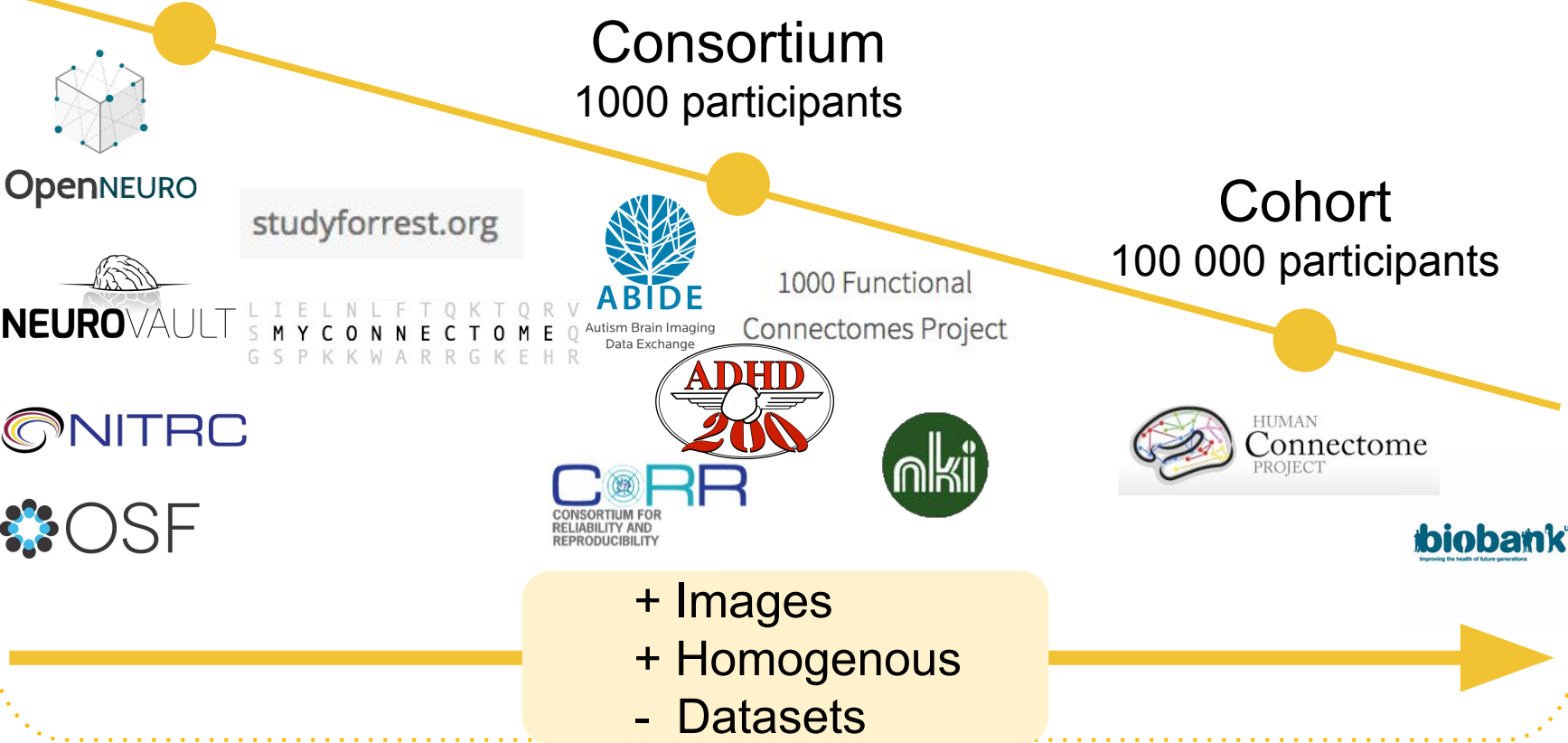


# Shared (open) data

Unique study  
30 participants

Consortium  
1000 participants

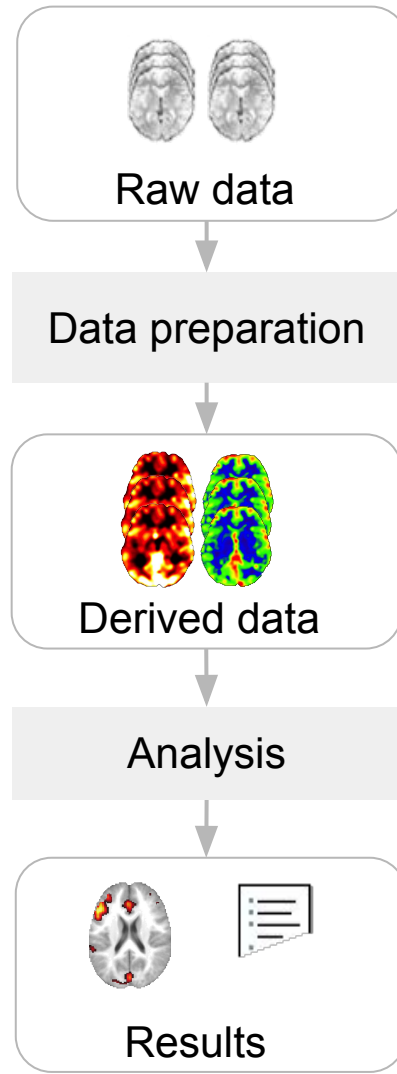
Cohort  
100 000 participants



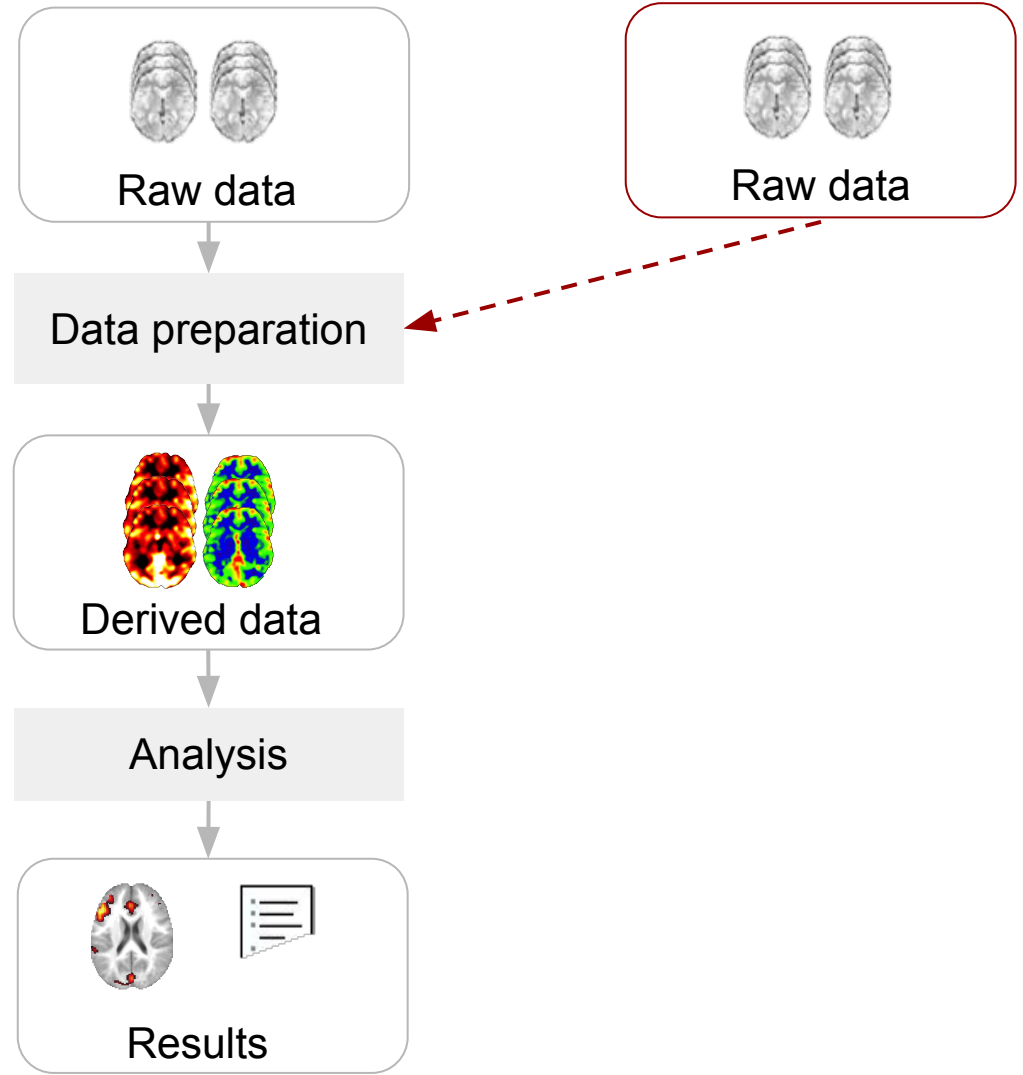
Crédits : Cerveaux, Neil Conway, [Flickr](#) (CC BY 2.0)

**Data** integration

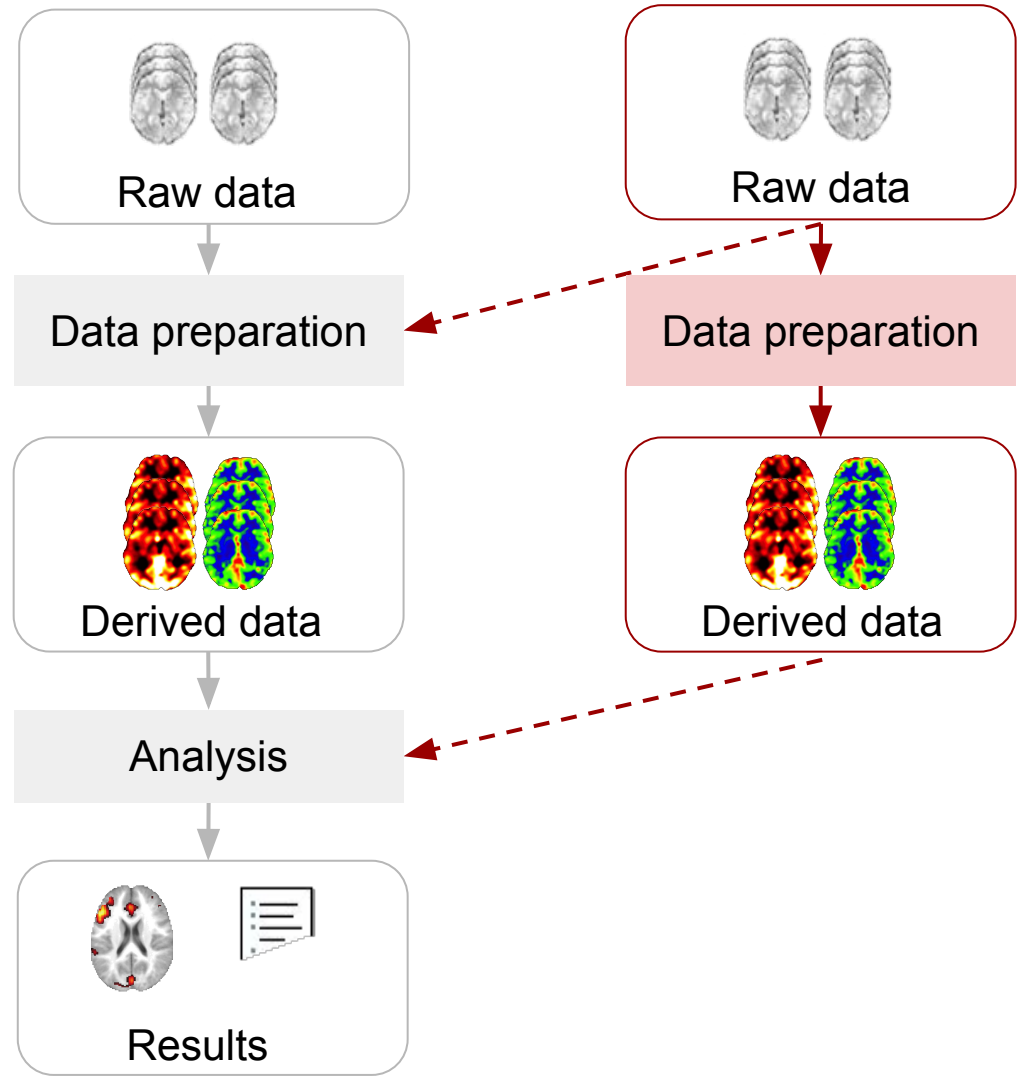
# Data integration



# Data integration

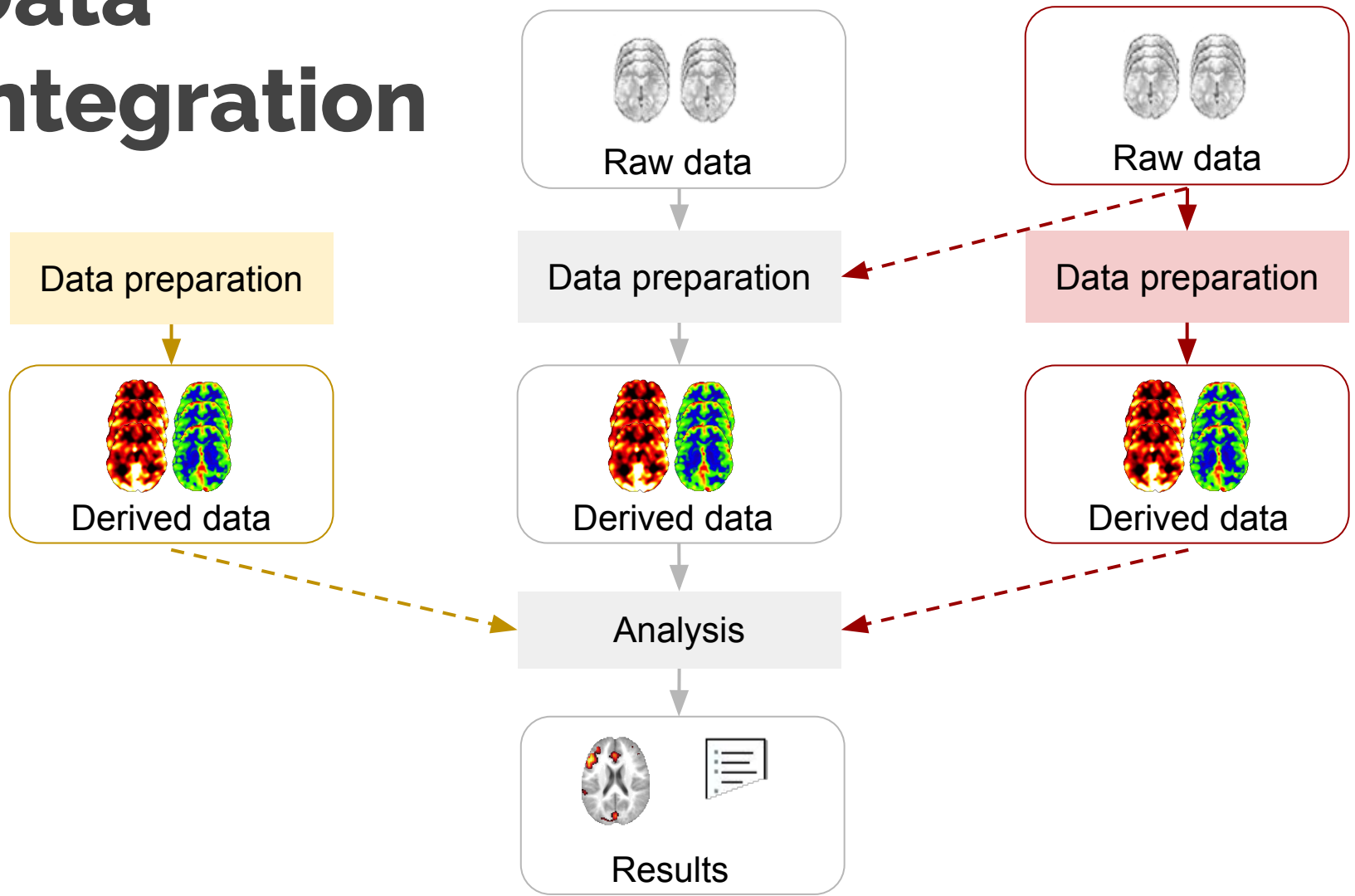


# Data integration

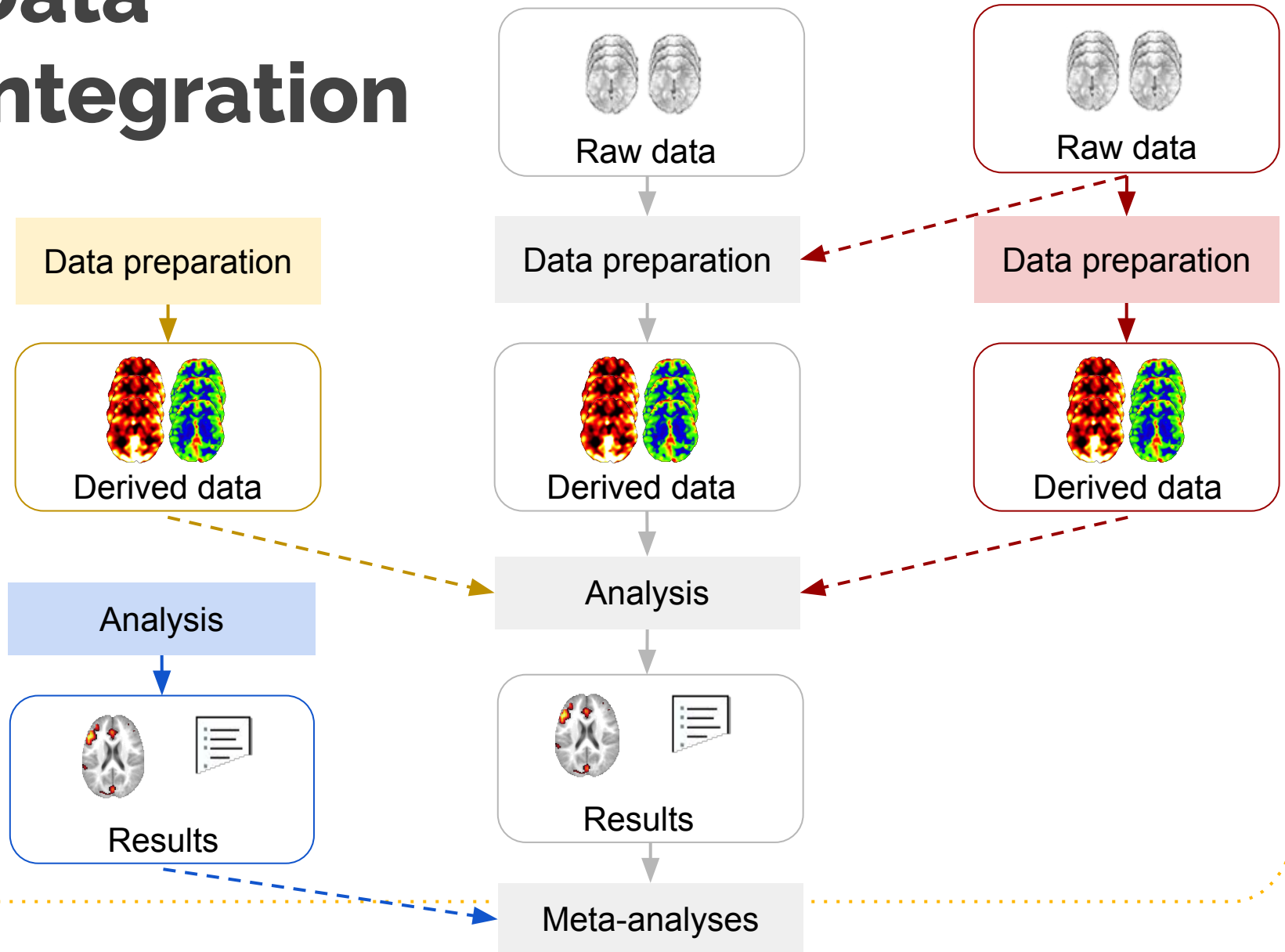




# Data integration



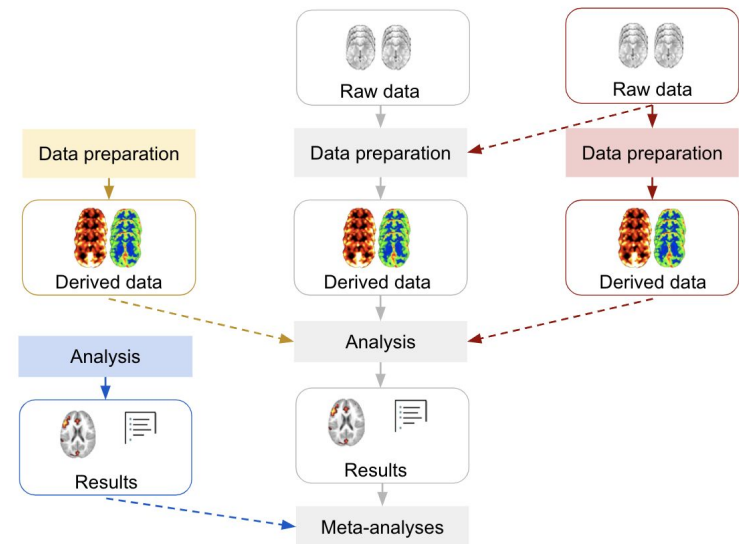
# Data integration



A new level of  
**variability**

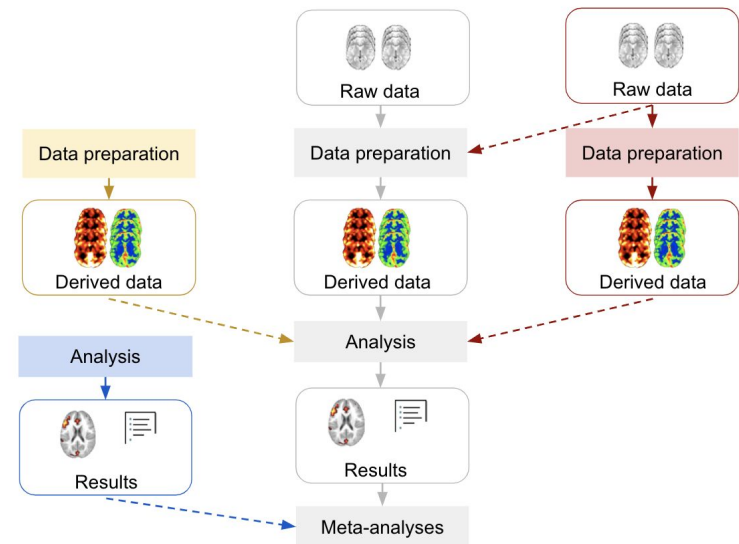
# Levels of variability

- Variability in **single-lab** studies
  - Across subjects
  - Across scans



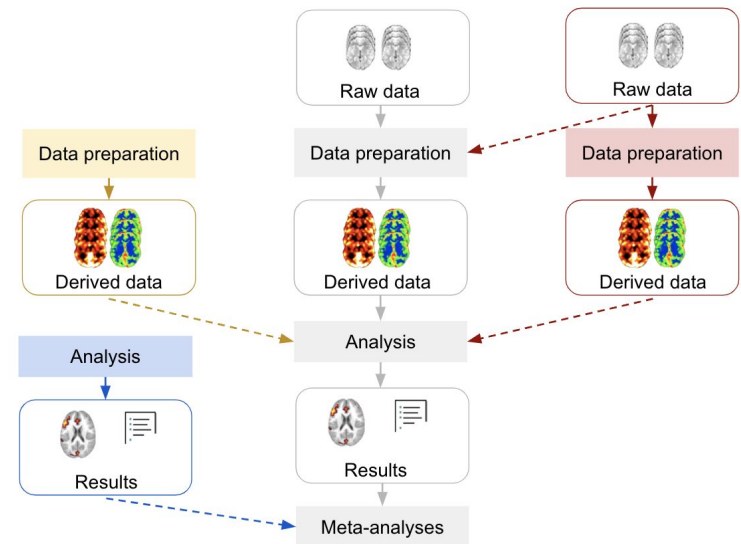
# Levels of variability

- Variability in **single-lab** studies
  - Across subjects
  - Across scans
- Variability in **multi-lab** studies
  - All of the above plus
  - Across machines
  - Across sites
  - Across acquisition protocols



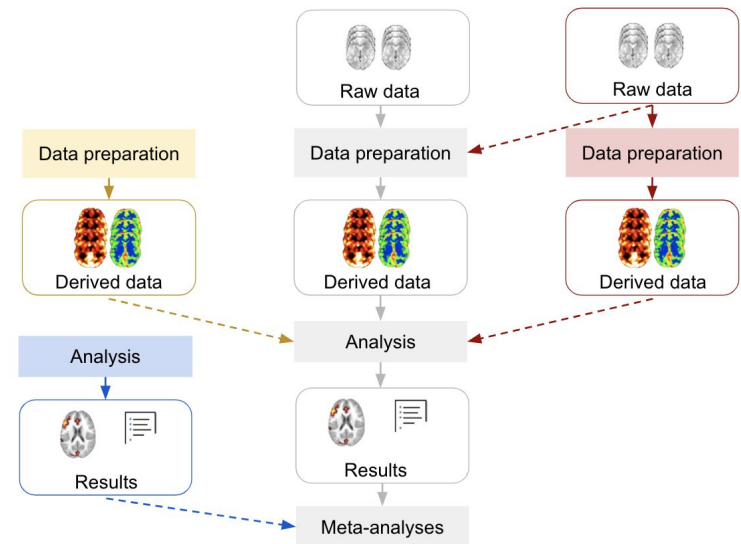
# Levels of variability

- Variability in **single-lab** studies
  - Across subjects
  - Across scans
- Variability in **multi-lab** studies
  - All of the above plus
  - Across machines
  - Across sites
  - Across acquisition protocols
- Variability in **data integration** studies
  - All of the above plus
  - Across preparation protocols
  - Across analysis protocols



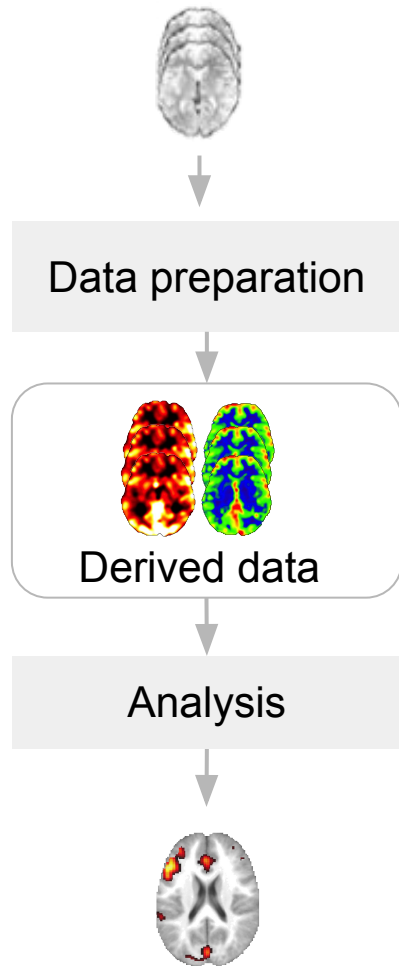
# Levels of variability

- Variability in **single-lab** studies
  - Across subjects
  - Across scans
- Variability in **multi-lab** studies
  - All of the above plus
  - Across machines
  - Across sites
  - Across acquisition protocols
- Variability in **data integration** studies
  - All of the above plus
  - Across preparation protocols
  - Across analysis protocols



*Analytical Variability*

# Analytic variability

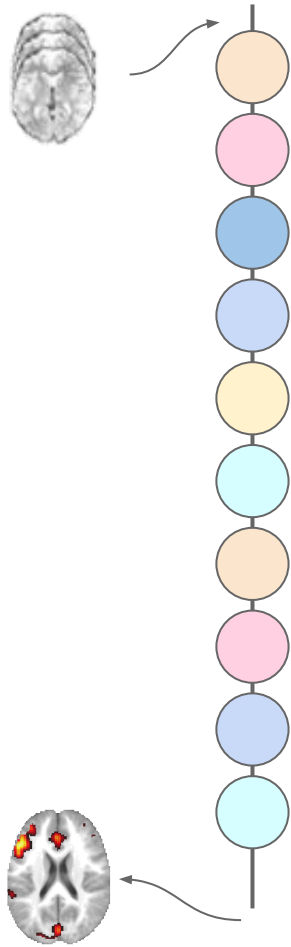


“Different **acceptable** analysis methods”

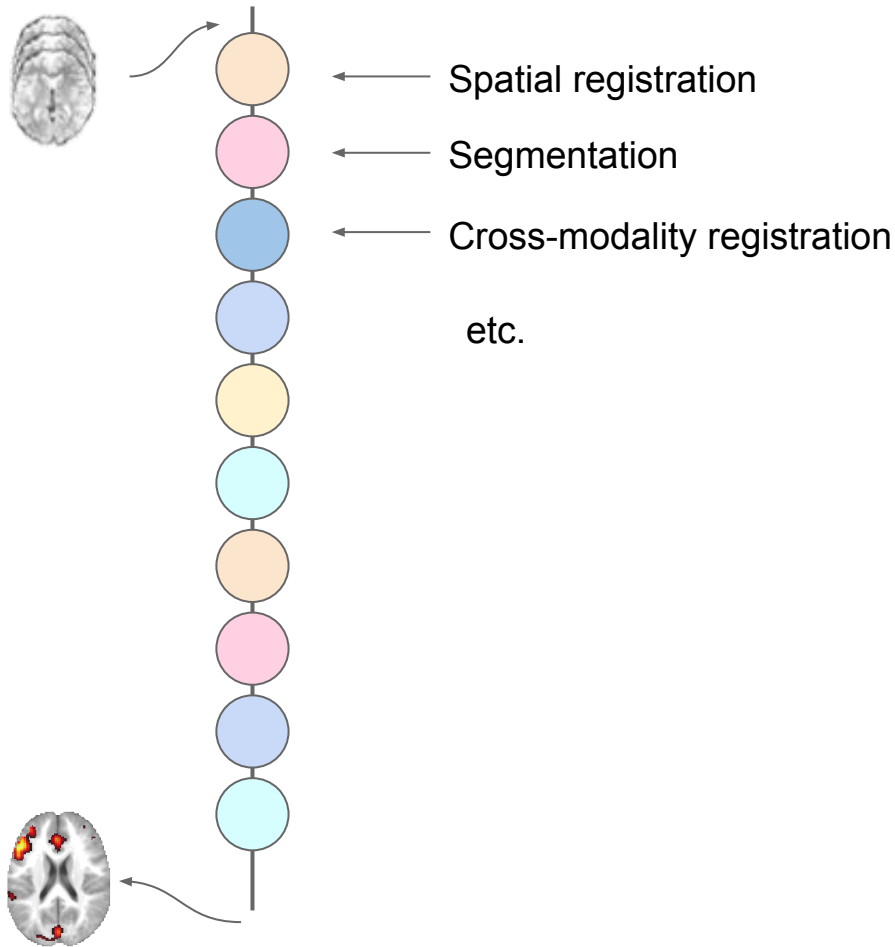
(Carp, Front. Neurosci. 2012)



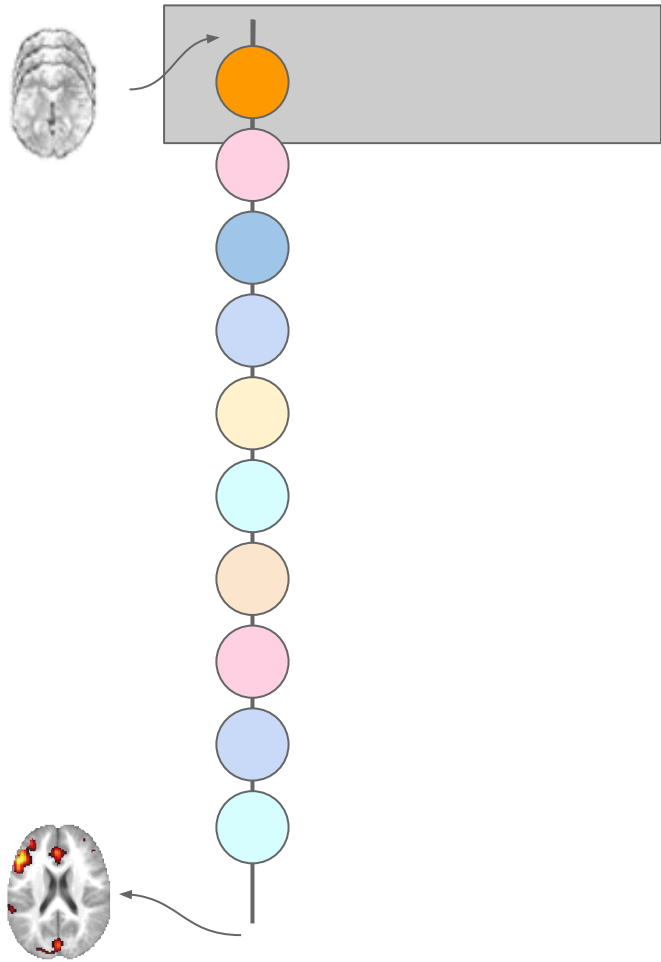
# Analytic variability



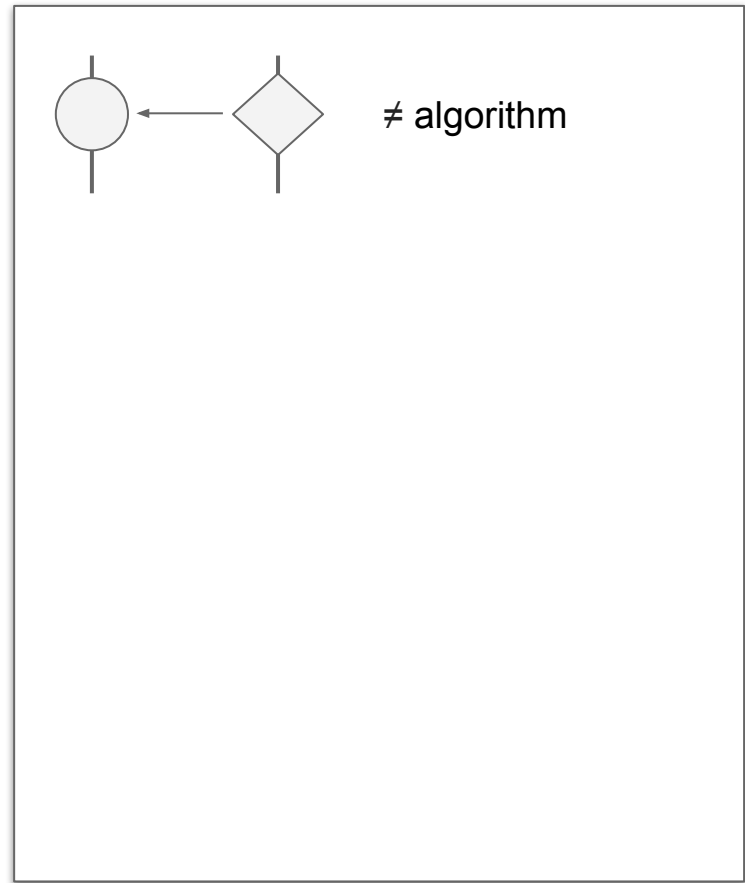
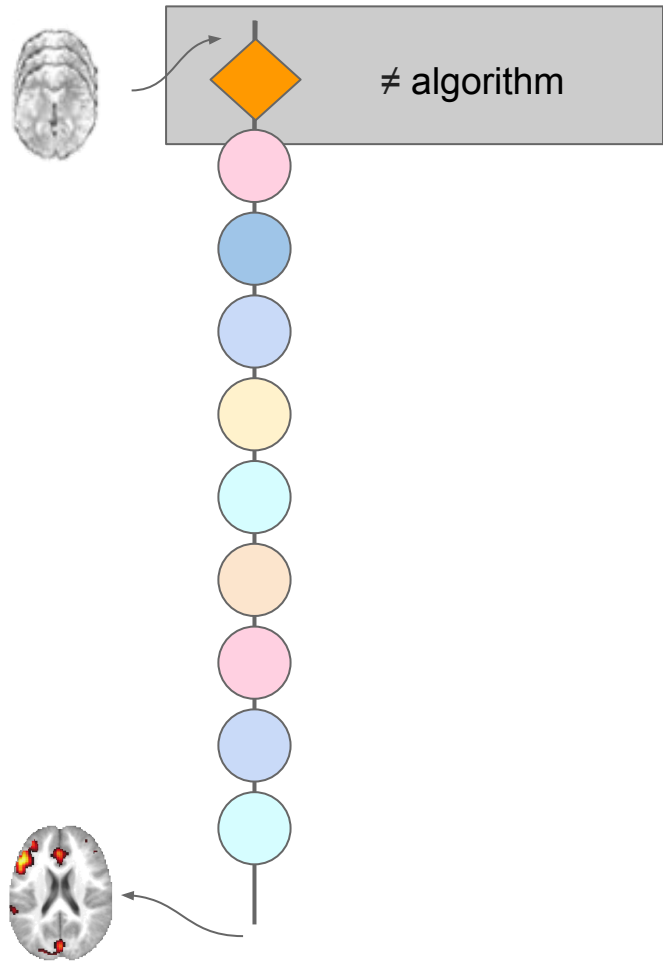
# Analytic variability



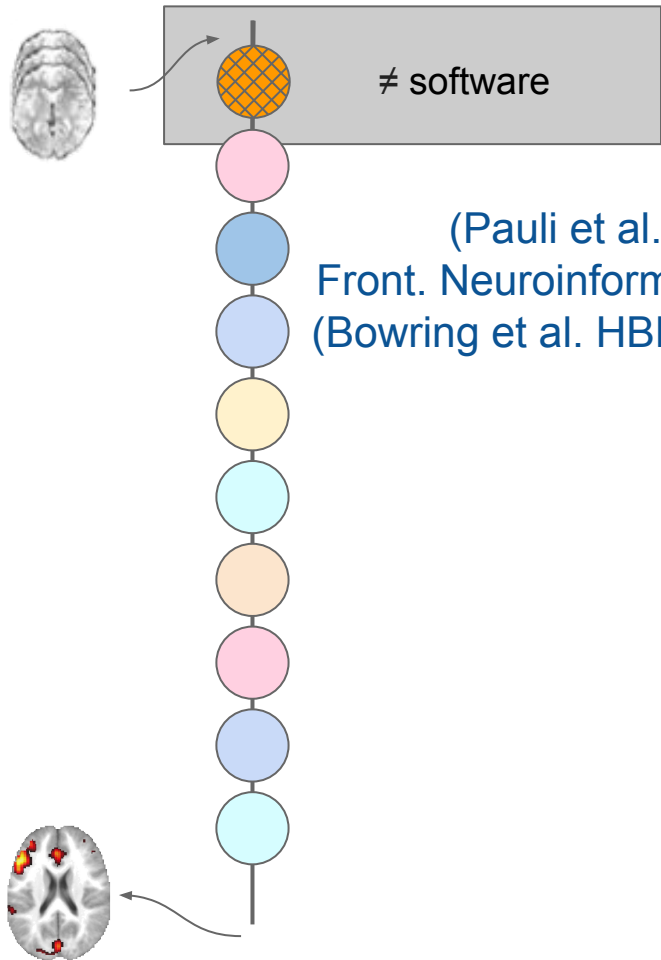
# Analytic variability



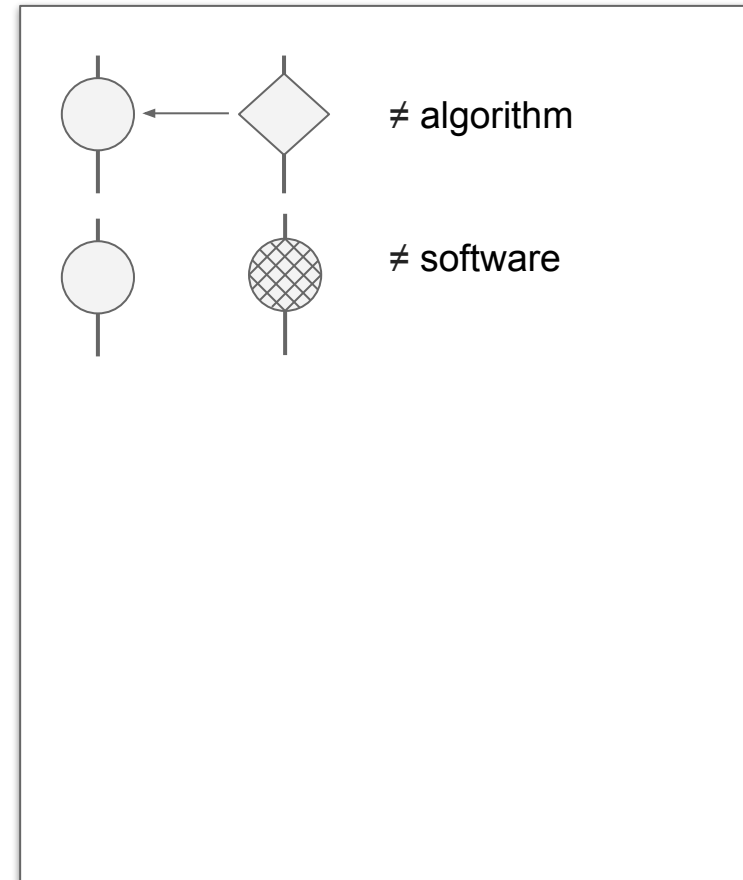
# Analytic variability



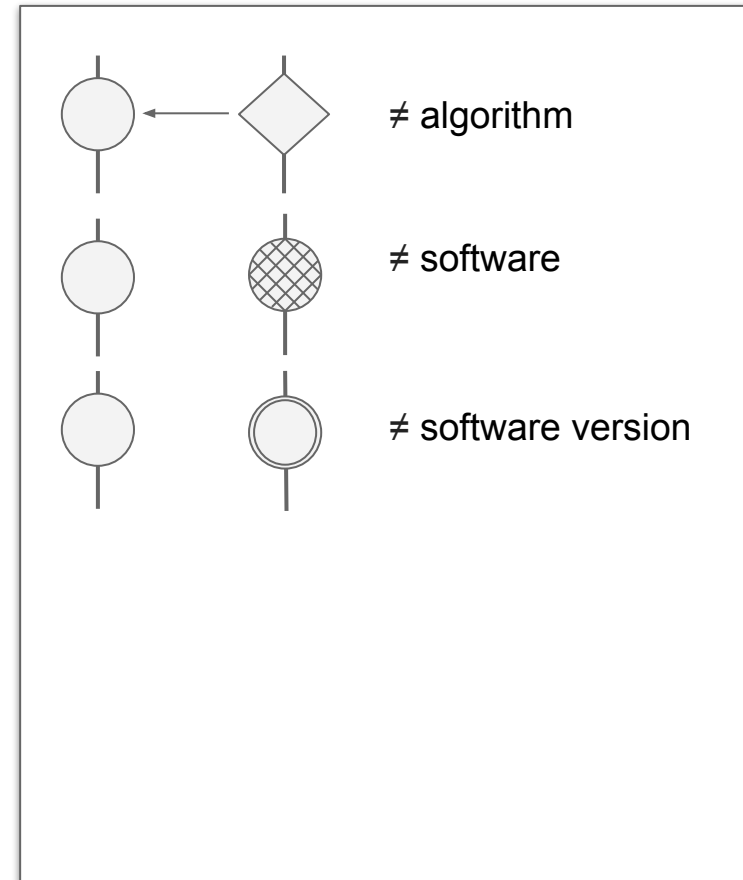
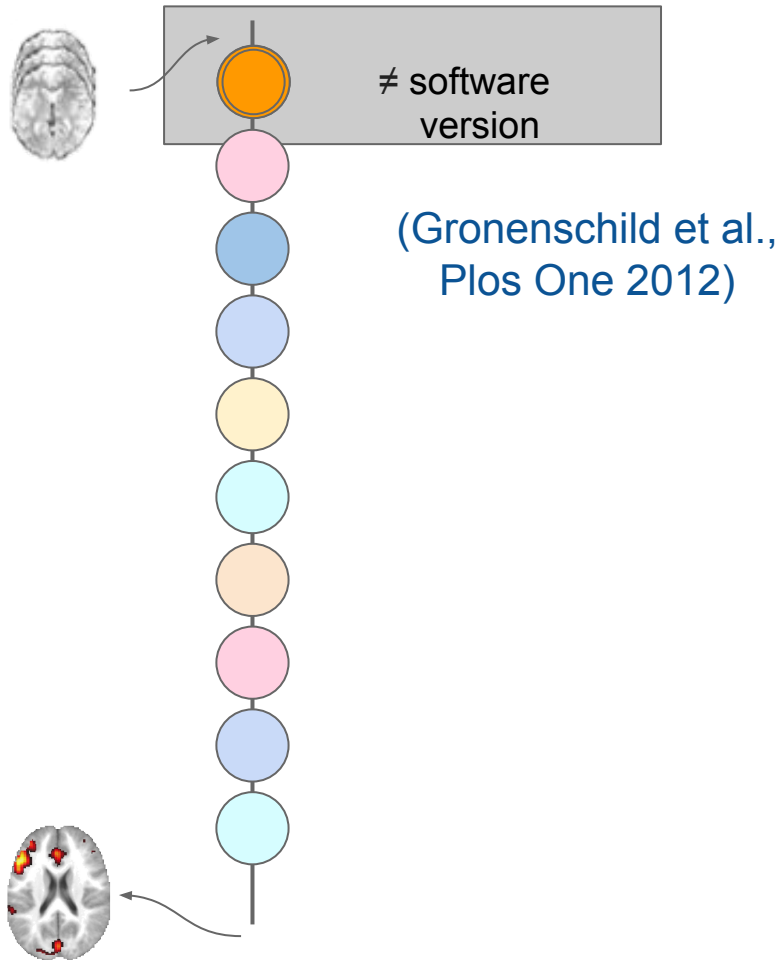
# Analytic variability



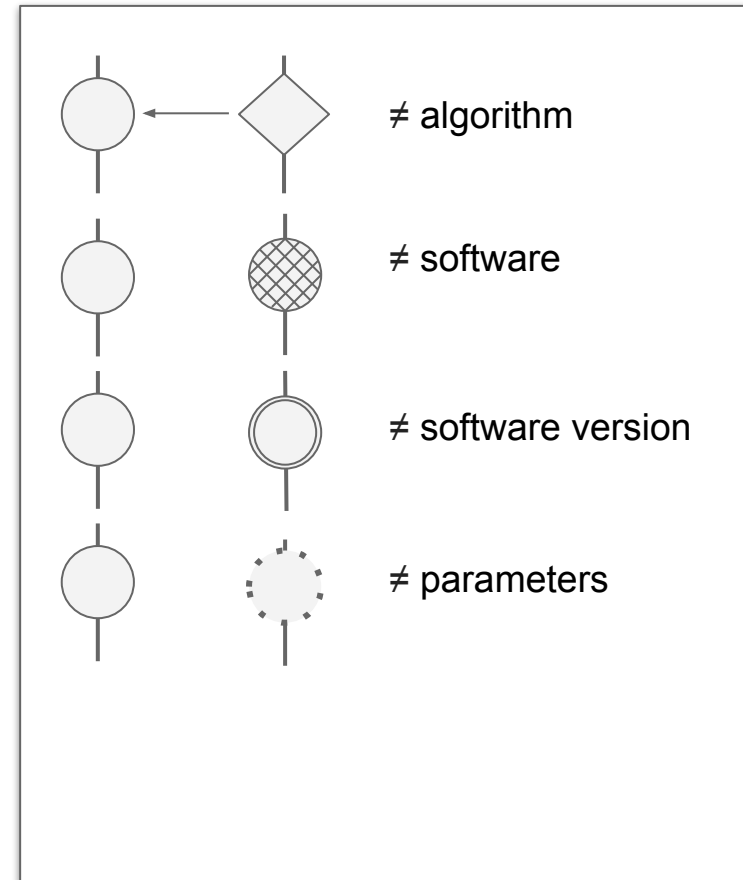
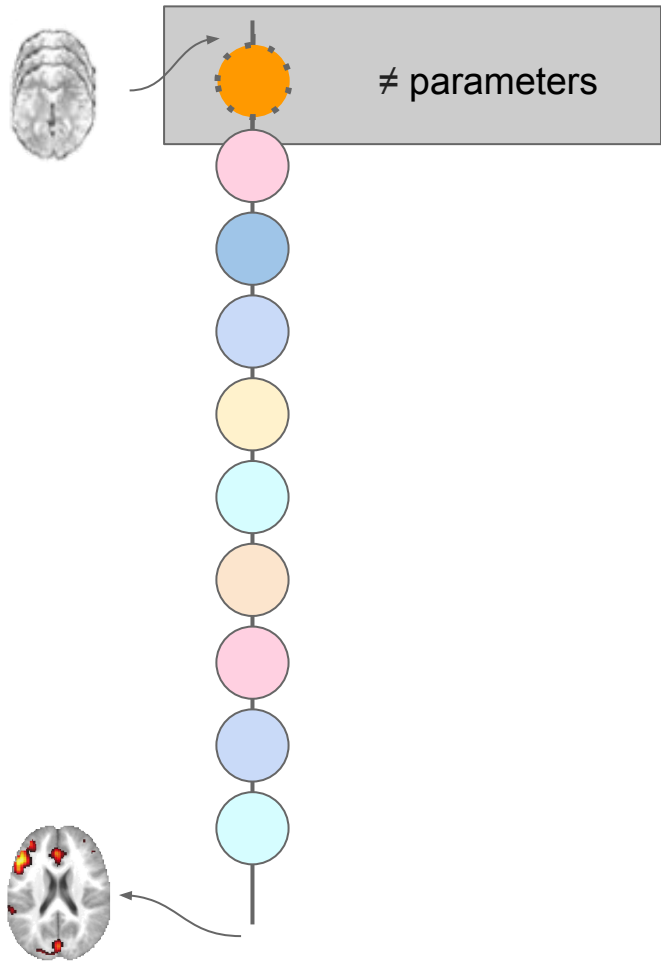
(Pauli et al.,  
Front. Neuroinform. 2016)  
(Bowring et al. HBM 2019)



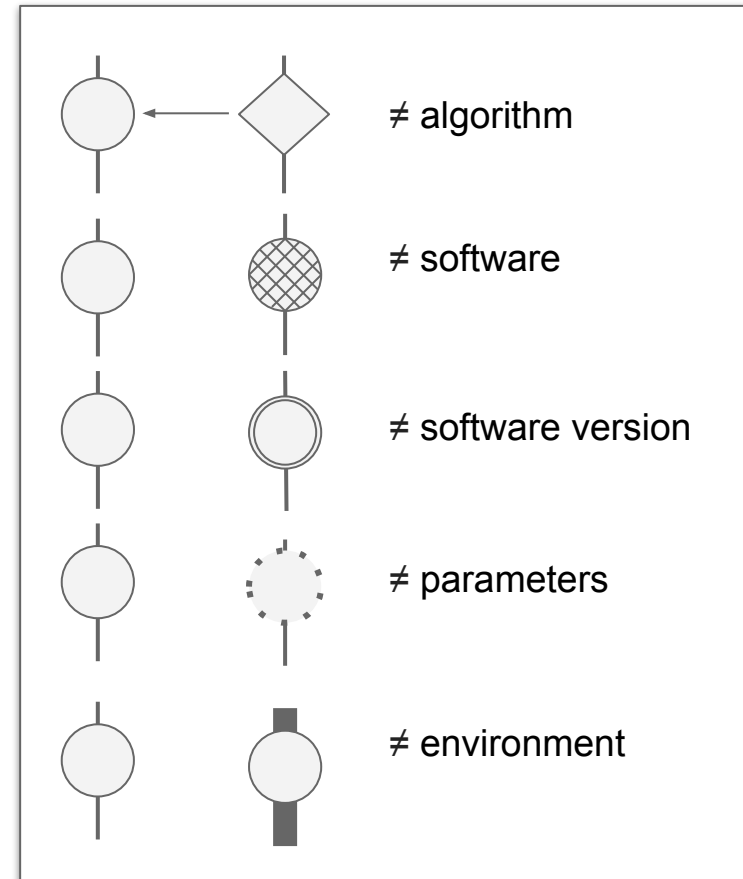
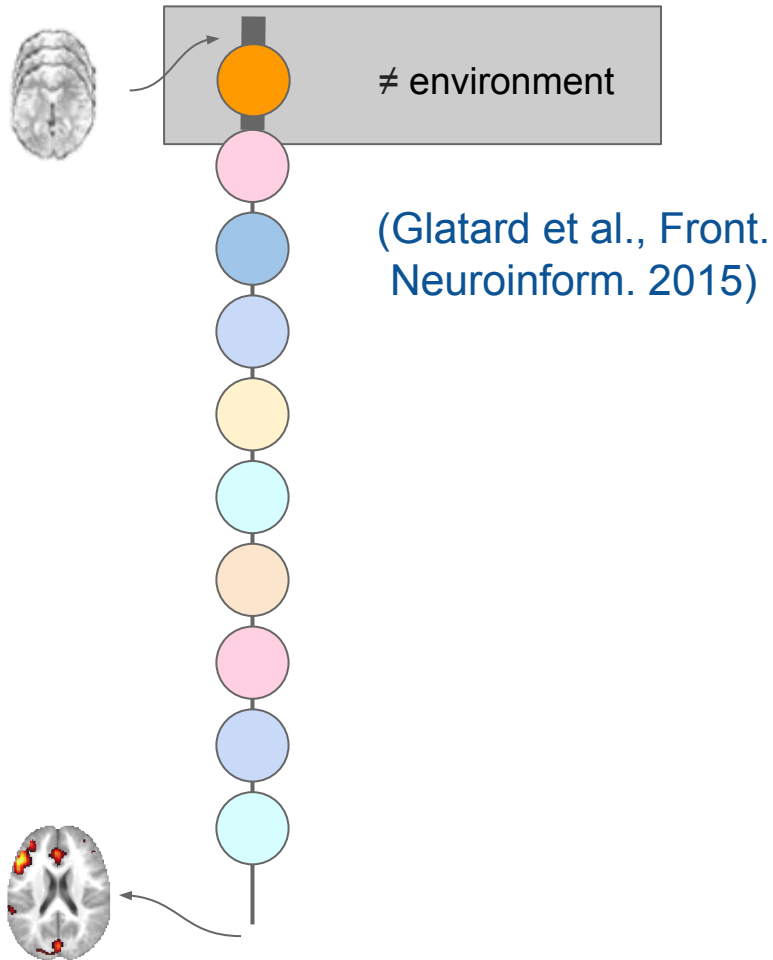
# Analytic variability



# Analytic variability

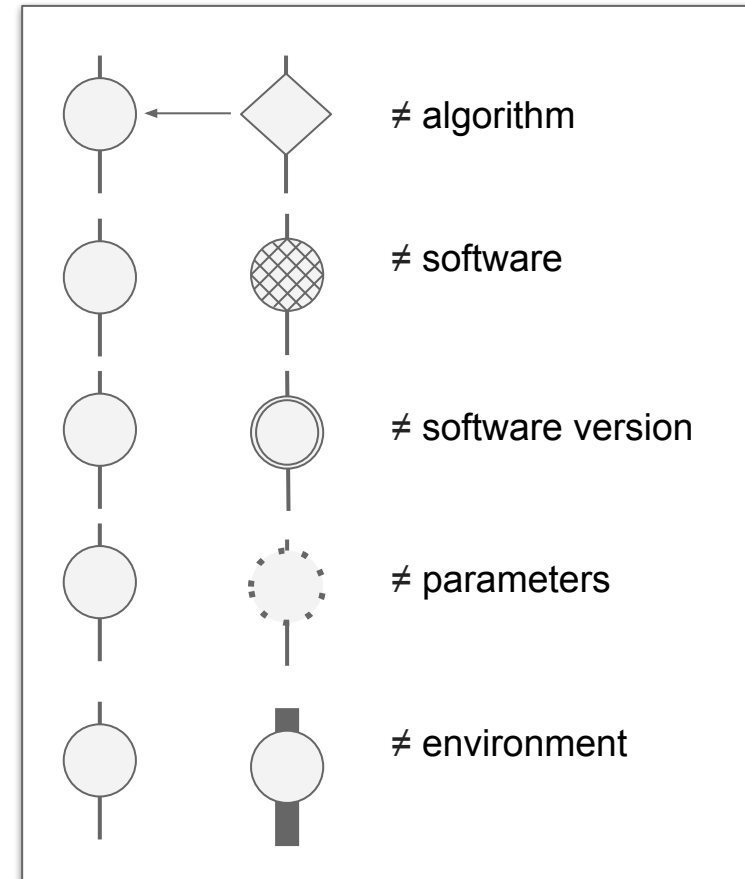
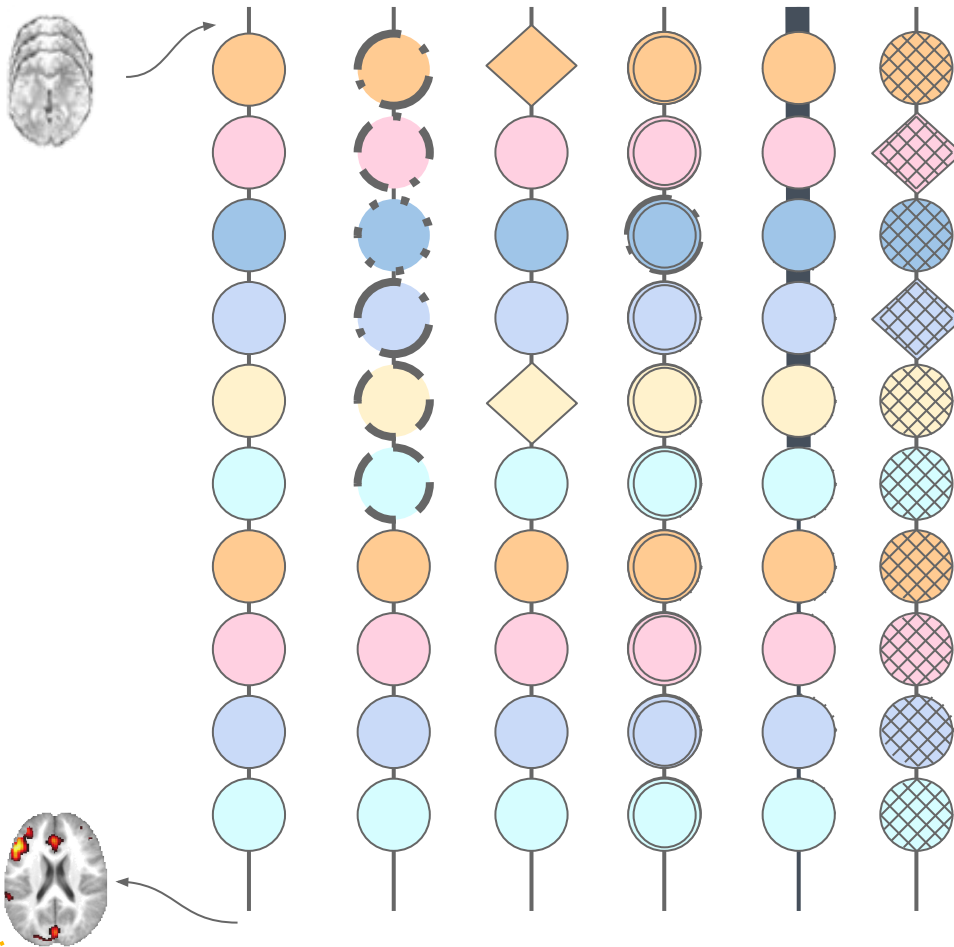


# Analytic variability





# Analytic variability



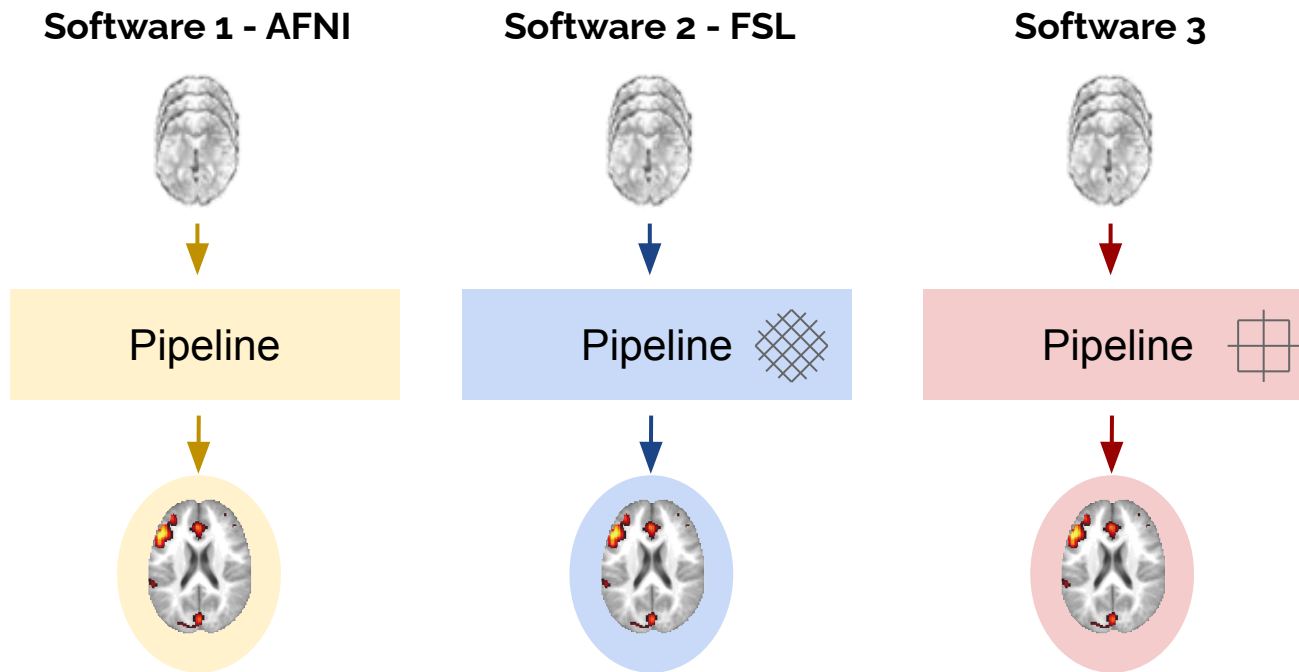
(Carp et al. Front. Neurosci. 2012; Nørgaard et al. JCBFM 2019;  
Botvinik-Nezer et al. Nature 2020)

# Variations across software

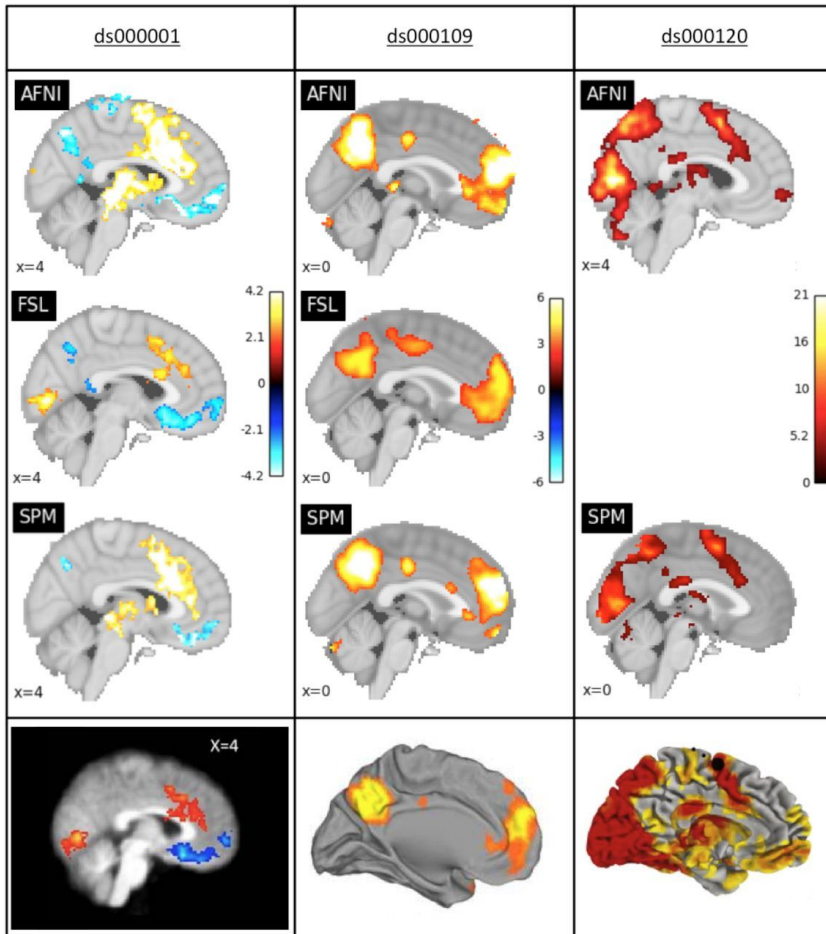
- Reproduced 3 published functional MRI studies
- Using 3 different software  
(Bowring et. al, HBM 2019)



Alex Bowring Tom Nichols

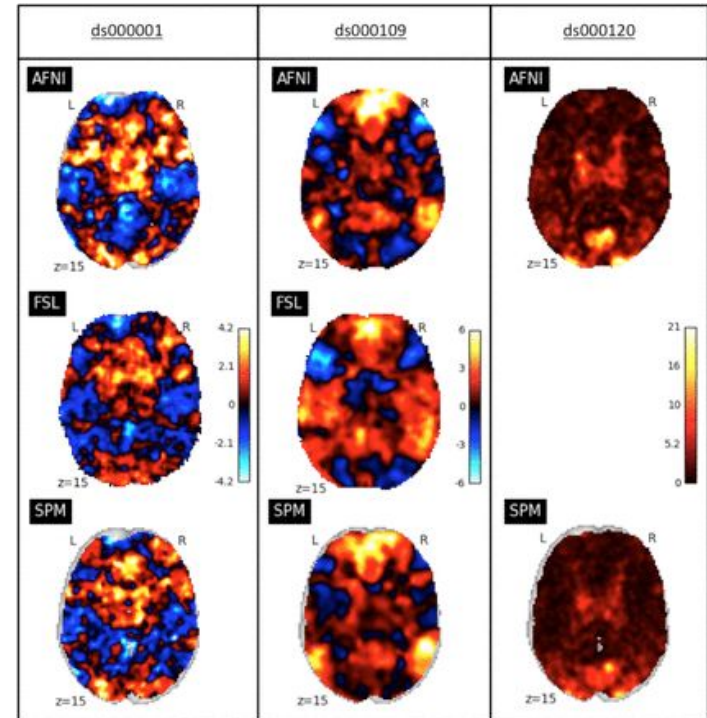


# Variations across software



Comparison of the final results

(Bowring et. al, HBM 2019)

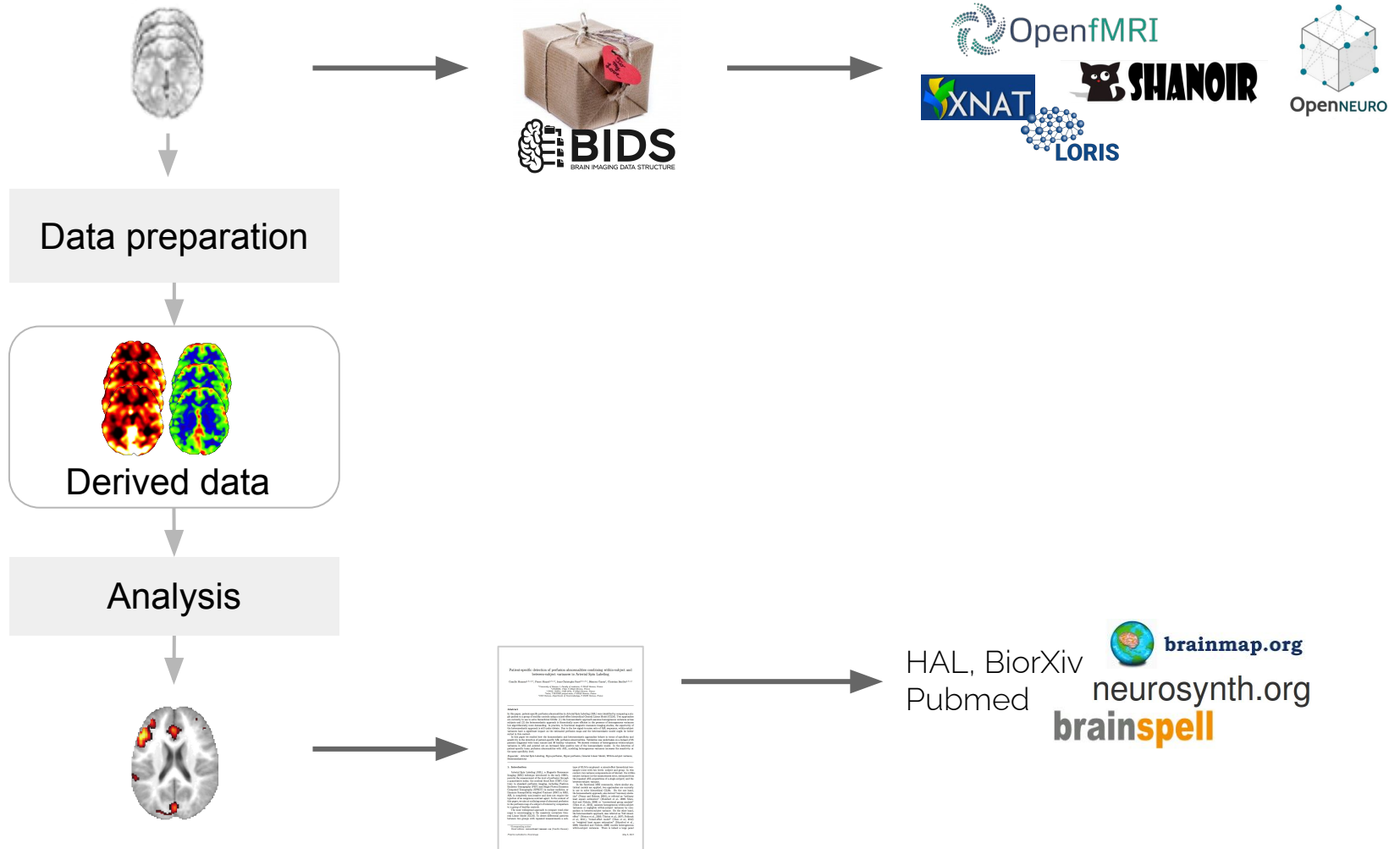


Comparison of the statistic maps

- 1 dataset, 70 teams, 9 pre-defined research questions (yes/no)  
(Botvinik-Nezer et. al, Nature 2020)

Sharing more  
**research outputs**

# Sharing more research outputs



# Sharing more research outputs

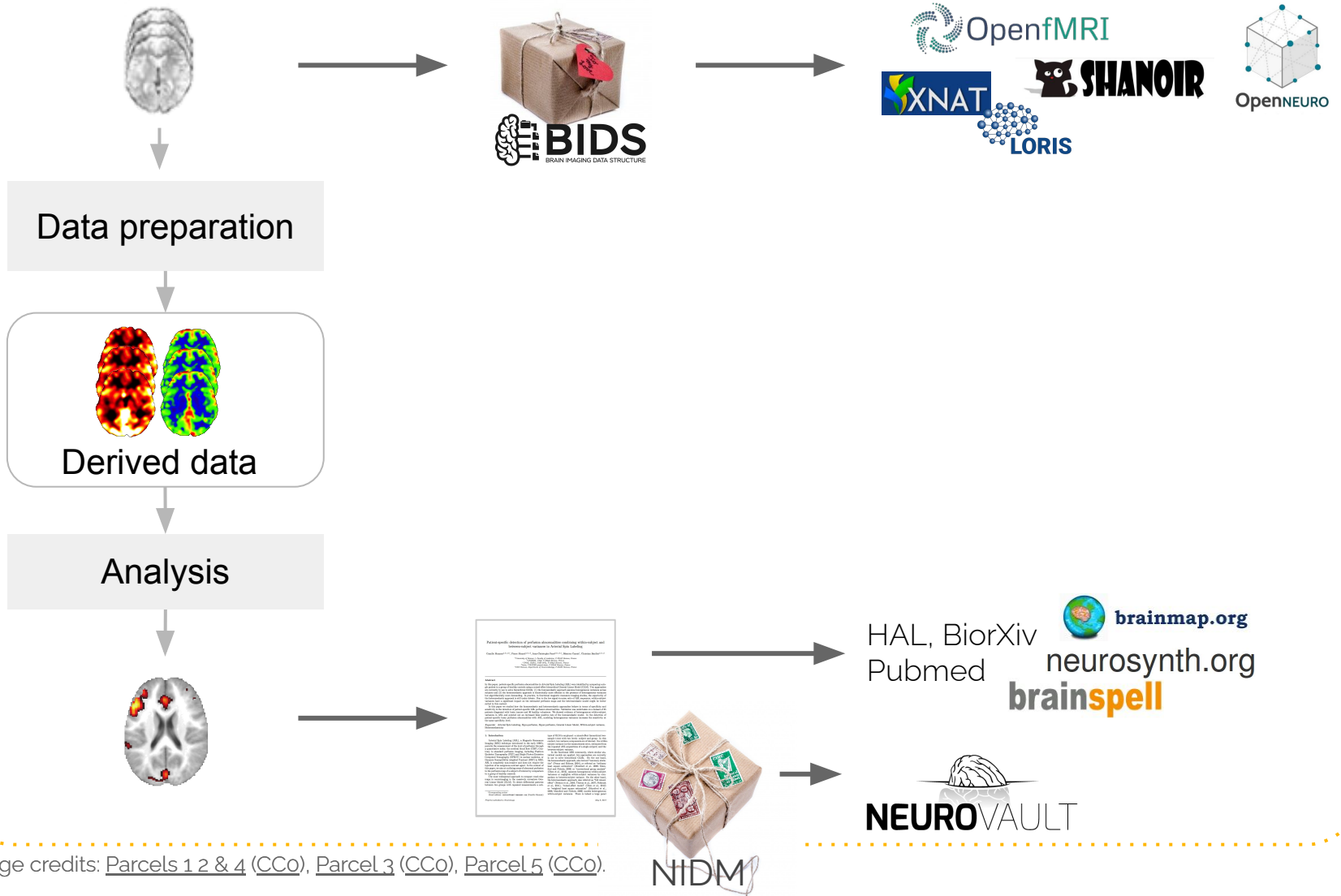
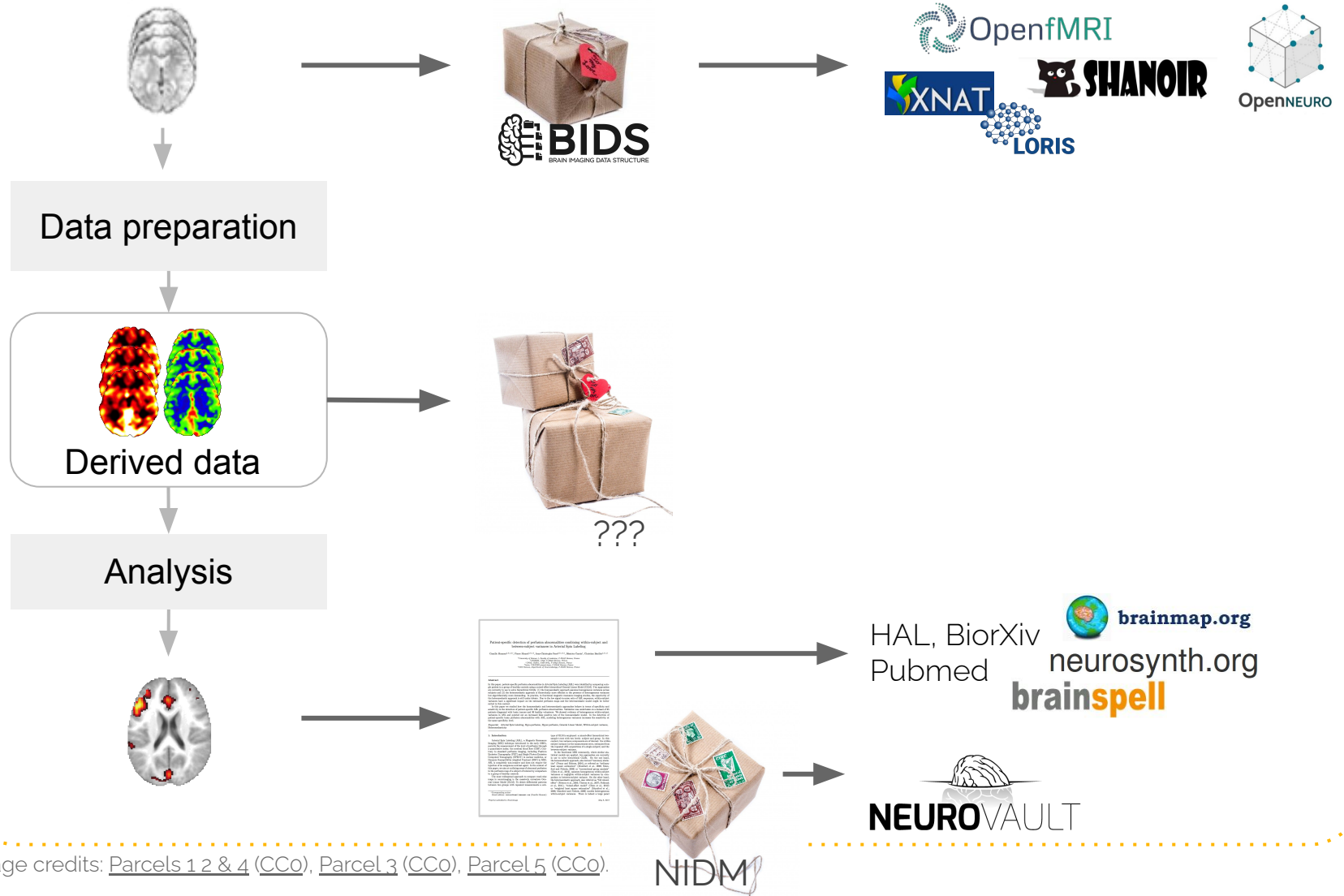


Image credits: [Parcels 1 2 & 4 \(CC0\)](#), [Parcel 3 \(CC0\)](#), [Parcel 5 \(CC0\)](#).

# Sharing more research outputs



# BIDS Prov **Bridging BIDS & NIDM**

## BIDS

```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   └── func/
```

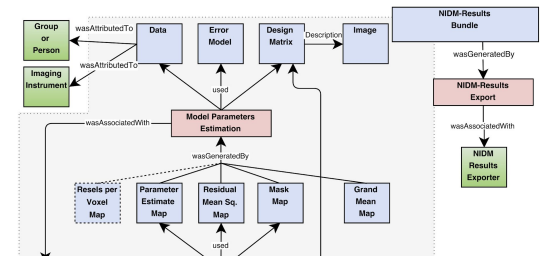
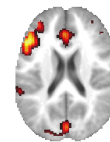


File naming convention + JSON files

—  
60 labs worldwide



## NIDM

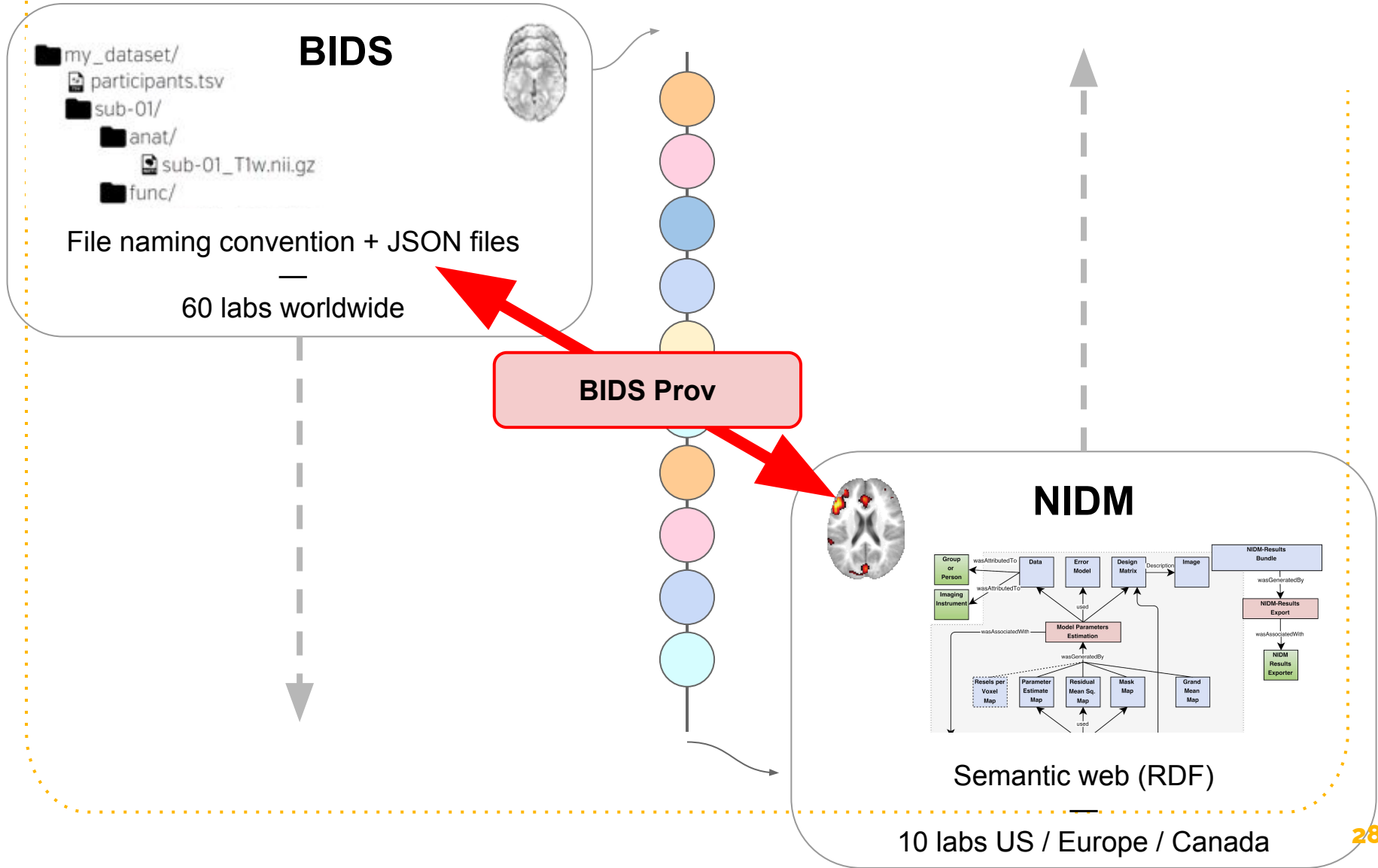


Semantic web (RDF)

10 labs US / Europe / Canada



# BIDS Prov Bridging BIDS & NIDM



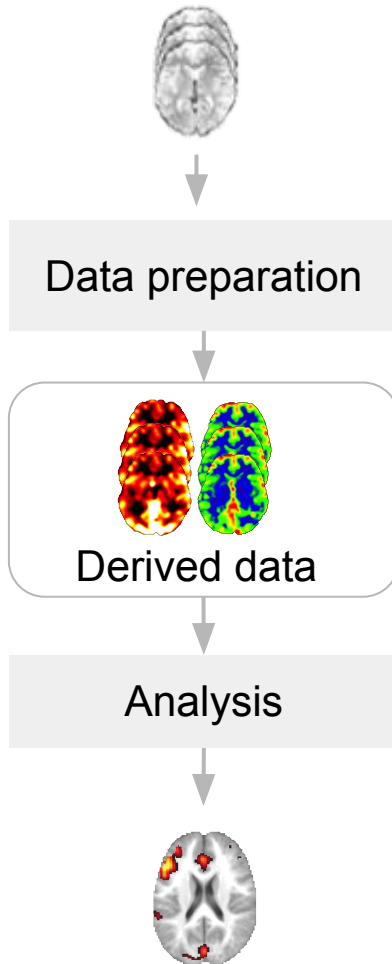
# BIDS Provenance



Satra Ghosh



Rémi Adon



BEP028	Provenance	Satra Ghosh and Camille Maumet	new BEP, actively looking for contributors and community input.
--------	------------	--------------------------------	---



Looking for contributors and community input

<https://github.com/bids-standard/BEP028> BIDSprov

# References

- Botvinik-Nezer**, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, ... Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Bowring**, A., Maumet\*, C., & Nichols\*, T. E. (2019). Exploring the impact of analysis software on task fMRI results. *Human Brain Mapping*, 40(11), 3362–3384. <https://doi.org/10.1002/hbm.24603>
- Carp**, J. (2012). On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments. *Frontiers in Neuroscience*, 6. <https://doi.org/10.3389/fnins.2012.00149>
- Glatard**, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, ... Evans, A. C. (2015). Reproducibility of neuroimaging analyses across operating systems. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00012>
- Gorgolewski**, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- Gronenschild**, E. H. B. M., Habets, P., Jacobs, H. I. L., Mengelers, R., Rozendaal, ... Marcelis, M. (2012). The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements. <https://doi.org/10.1371/journal.pone.0038234>
- Nørgaard**, M., Ganz, M., Svarer, C., Frokjaer, V. G., Greve, ... Knudsen, G. M. (2020). Different preprocessing strategies lead to different conclusions: A [11C]DASB-PET reproducibility study. *Journal of Cerebral Blood Flow & Metabolism*, 40(9), 1902–1911. <https://doi.org/10.1177/0271678X19880450>
- Pauli**, R., Bowring, A., Reynolds, R., Chen, G., Nichols, T. E., & Maumet, C. (2016). Exploring fMRI Results Space: 31 Variants of an fMRI Analysis in AFNI, FSL, and SPM. *Frontiers in Neuroinformatics*, 10. <https://doi.org/10.3389/fninf.2016.00024>
- Poldrack**, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, 18(2), 115–126. <https://doi.org/10.1038/nrn.2016.167>

# Données ouvertes : comment permettre leur réutilisation?

Camille Maumet



Sculpture by Malin Bjornsdotter "Cerebia", [OHBM Brain Art SIG](#)

# Merci