



HAL
open science

Towards large-scale brain imaging studies

Camille Maumet

► **To cite this version:**

Camille Maumet. Towards large-scale brain imaging studies. workshop Virtual 2020 - Inteligencia Artificial en imagenes medicas, Dec 2020, Santiago (online), Chile. inserm-03058348

HAL Id: inserm-03058348

<https://inria.hal.science/inserm-03058348>

Submitted on 11 Dec 2020

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.

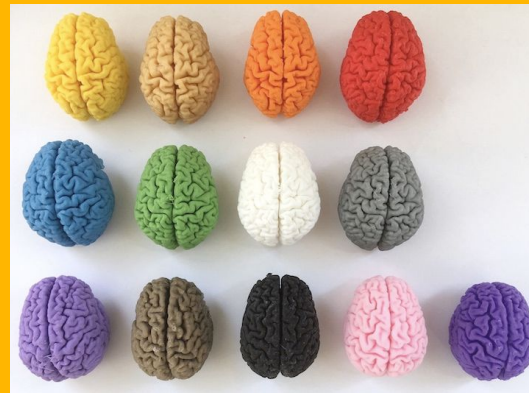
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Workshop | INTELIGENCIA ARTIFICIAL EN IMÁGENES MÉDICAS
Inria Chile - Dec 11, 2020

Towards **large-scale** brain imaging studies

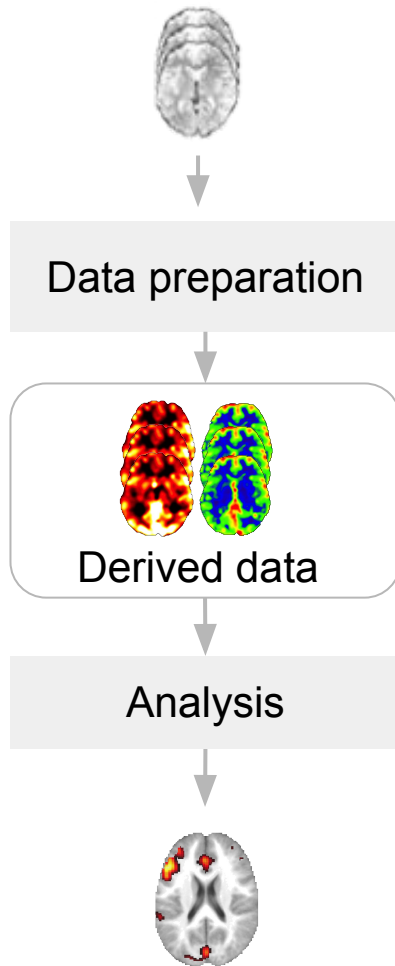


Camille Maumet
Univ Rennes, Inria, CNRS, Inserm

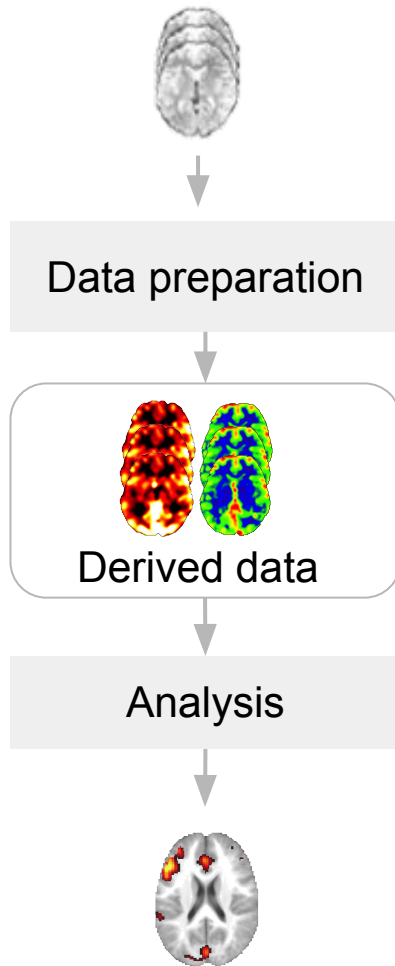


Image credits: Sculpture by Malin Bjornsdotter "Cerebia", [OHBM Brain Art SIG](#)

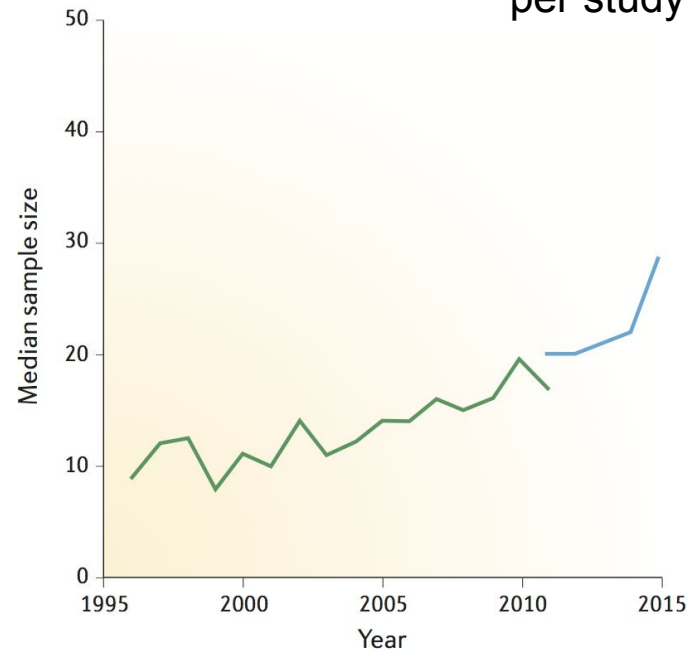
A brain imaging study



A brain imaging study



about 30 participants
per study



[Poldrack et. al, Nature Neuroscience 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

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



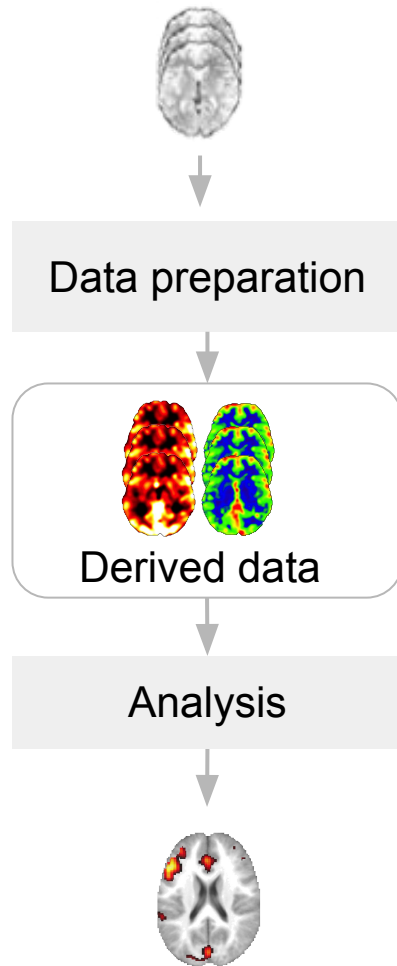
▲ 'Middle-aged runners outperform runners in the 20s ...' Photograph: FatCamera/Getty

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.

Faster in their forties
than twenties?

Selection bias

A brain imaging study



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



Open data

Unique study
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

- + Images
- + Homogenous
- Datasets



Open data

Unique study
30 participants



OpenNEURO

studyforrest.org

Consortium
1000 participants



ABIDE
Autism Brain Imaging
Data Exchange

1000 Functional
Connectomes Project

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

CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY



- + Images
- + Homogenous
- Datasets



Open data

Unique study
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



ABIDE
Autism Brain Imaging
Data Exchange

1000 Functional
Connectomes Project



NITRC

OSF

CORR
CONSORTIUM FOR
RELIABILITY AND
REPRODUCIBILITY



Cohort
100 000 participants



HUMAN
Connectome
PROJECT

biobank^{uk}
Improving the health of future generations

- + Images
- + Homogenous
- Datasets

The New York Times

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images

Lack of diversity of the input data

Algorithmic bias

OPT-OUT CULTURE

Using a fitness app taught me the scary truth about why privacy settings are a feminist issue



The social fitness app Strava is meant to help people compare data on their workouts.

FROM OUR SERIES

Ideas

Our home for bold arguments and big thinkers.



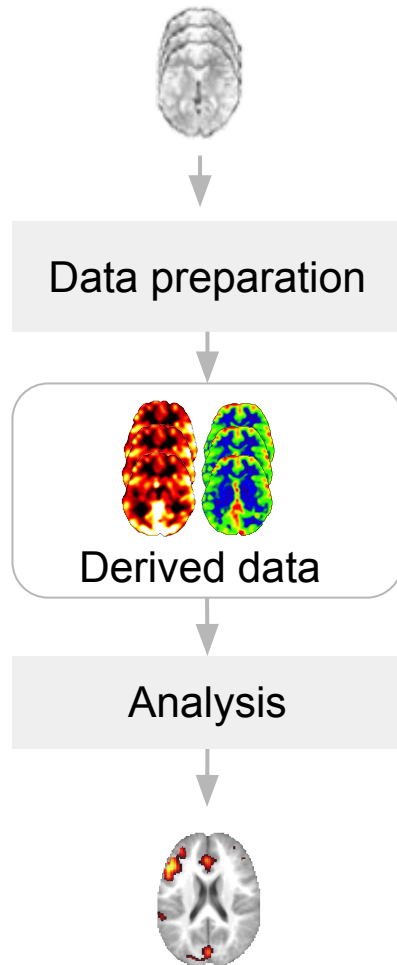
By **Rosie Spinks**

August 1, 2017 • This article is more than 2 years old.

Lack of diversity of the team

Data privacy

A brain imaging study

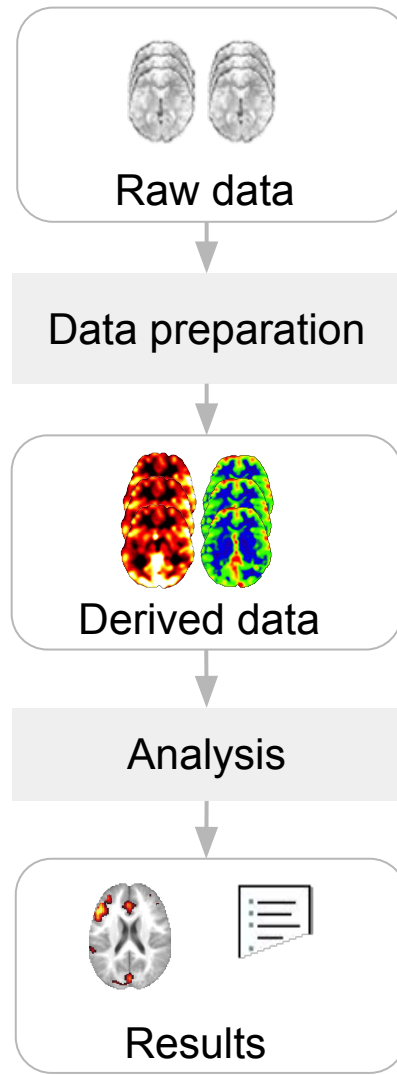


We need **bigger** and **representative** samples

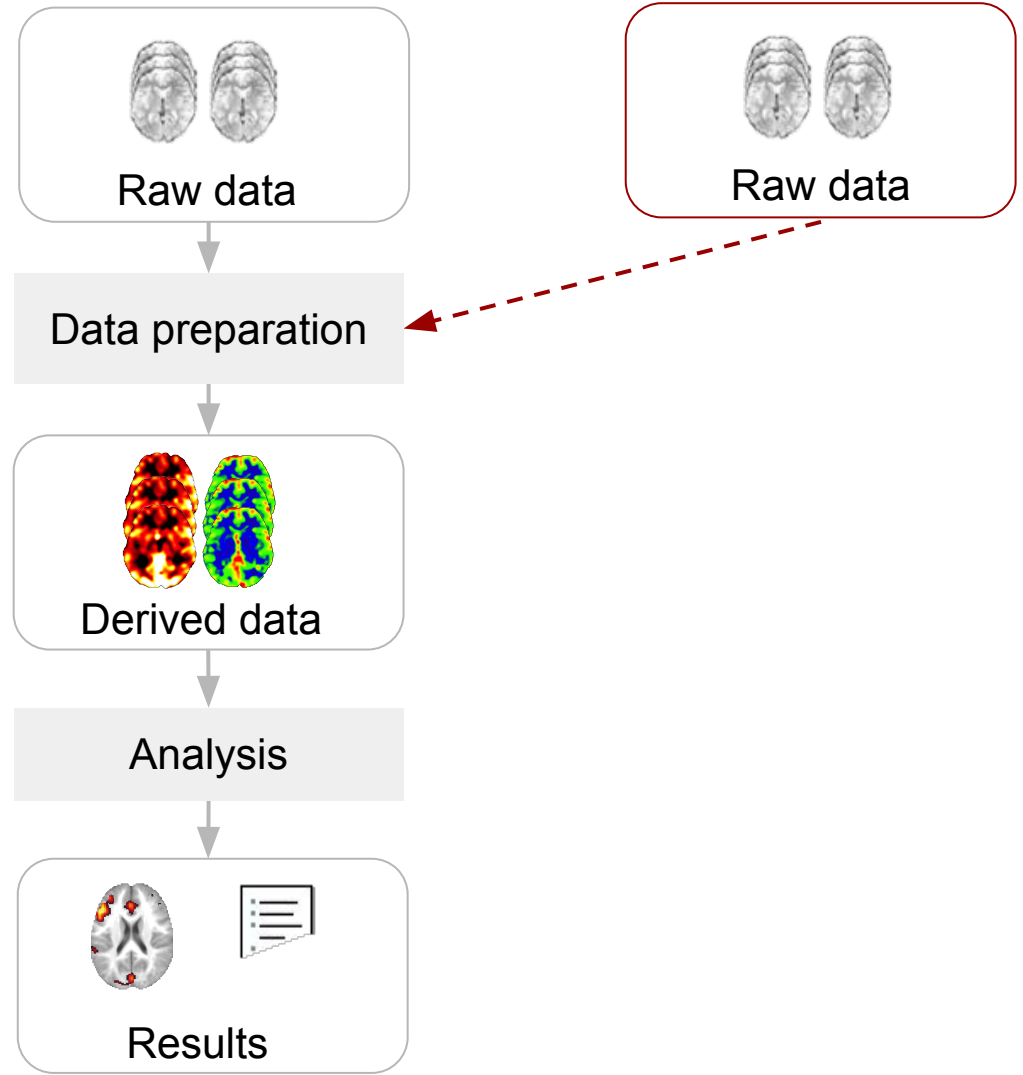
We need **diversity in datasets** *and* in **leadership teams**

Data integration

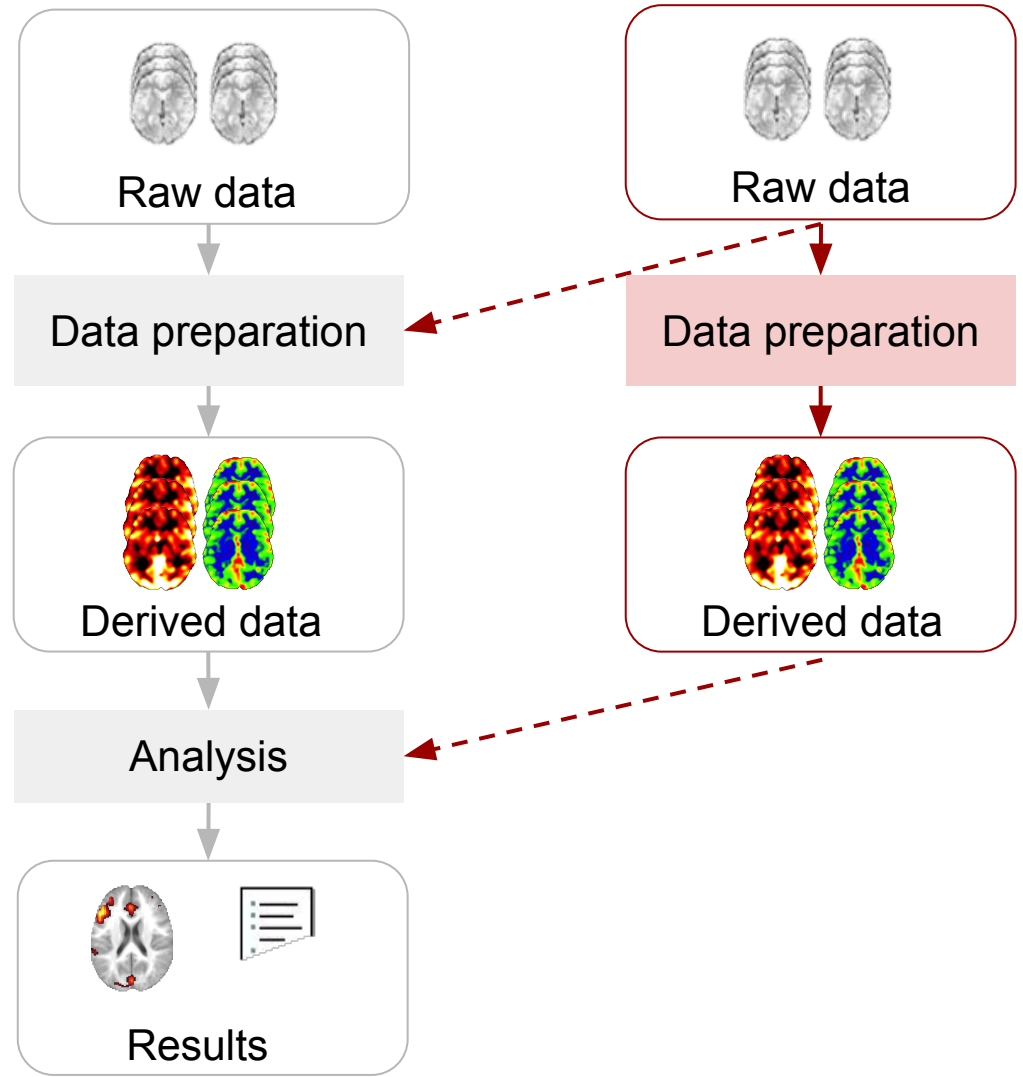
Data integration



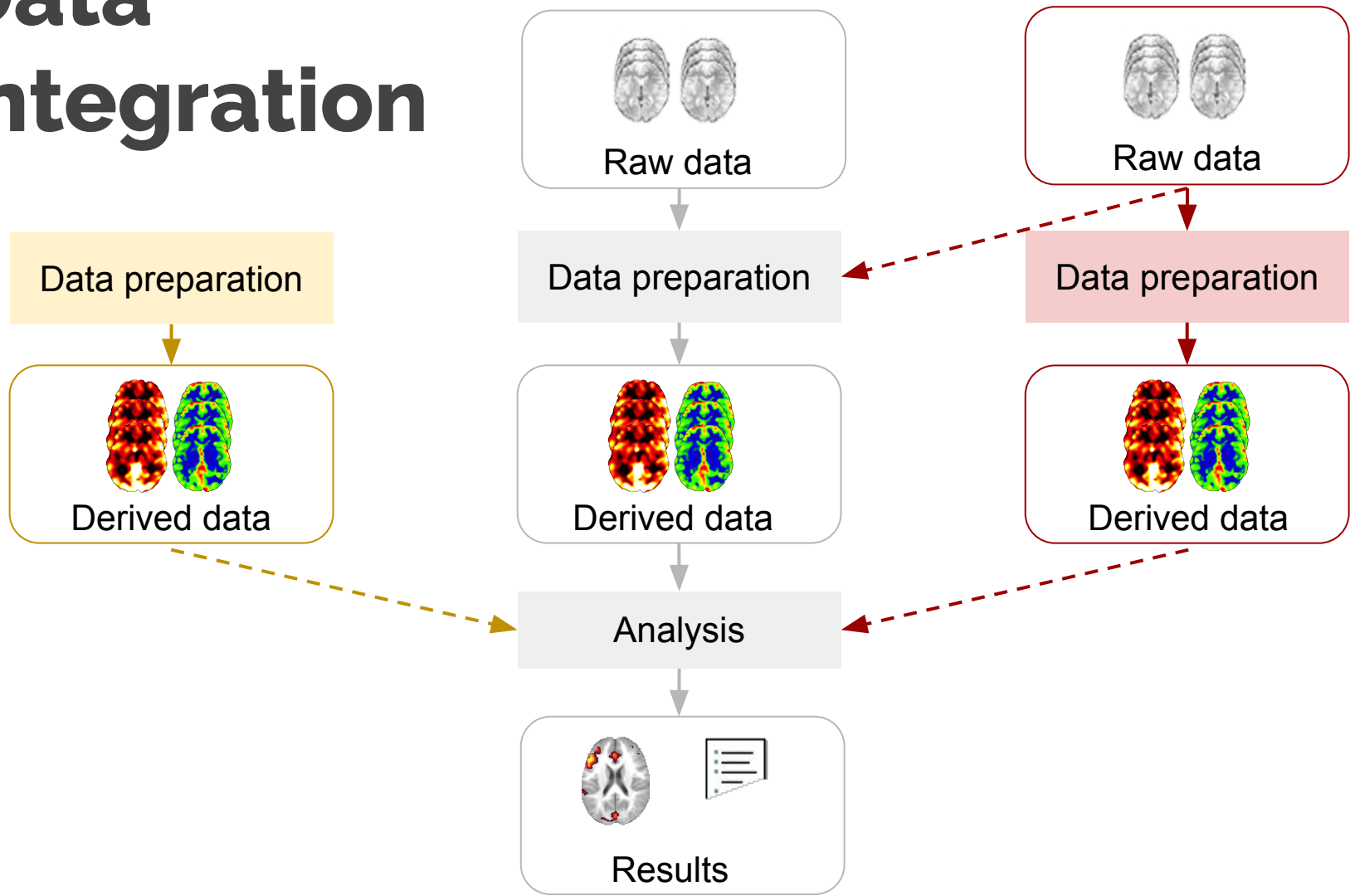
Data integration



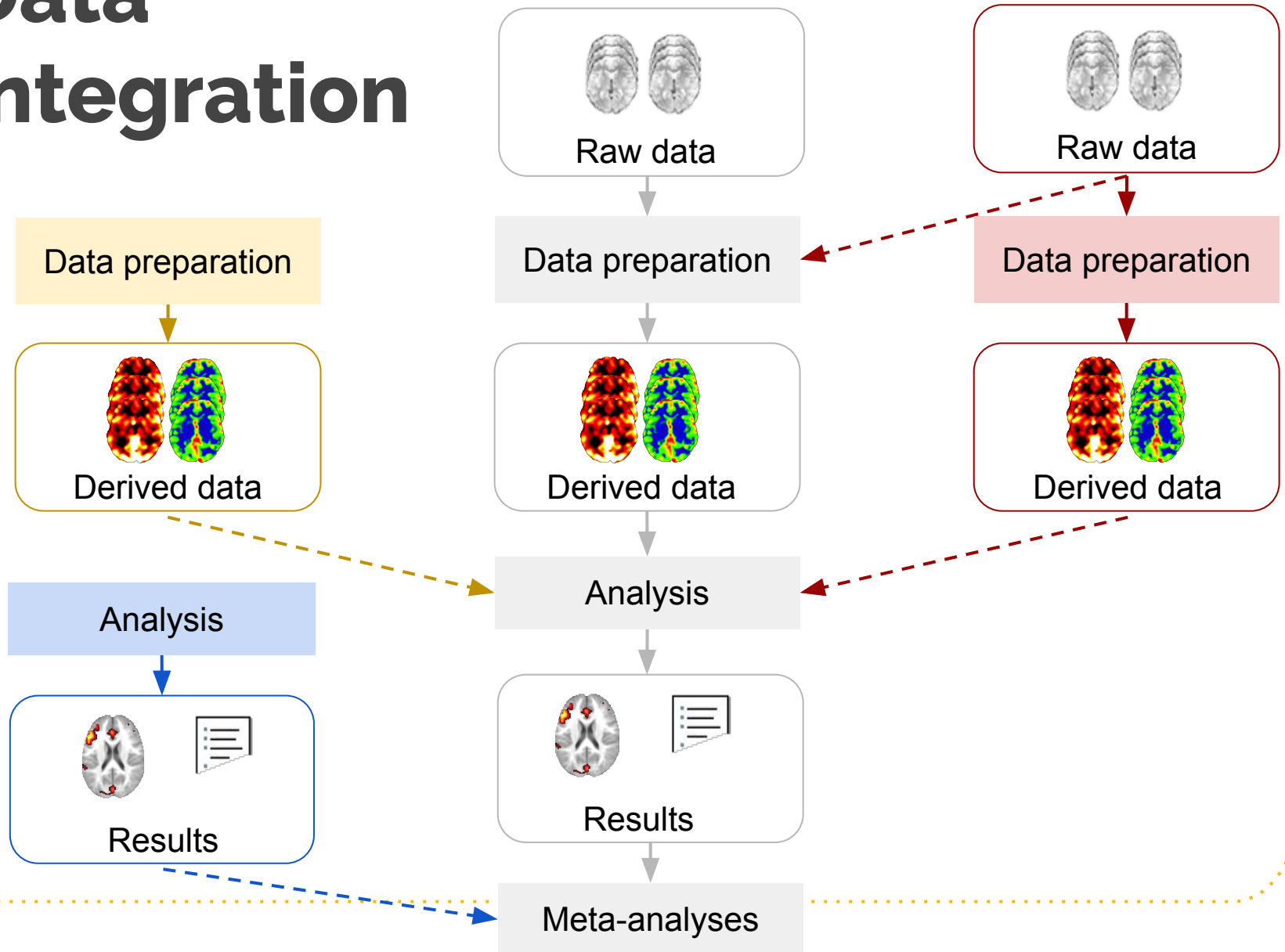
Data integration



Data integration



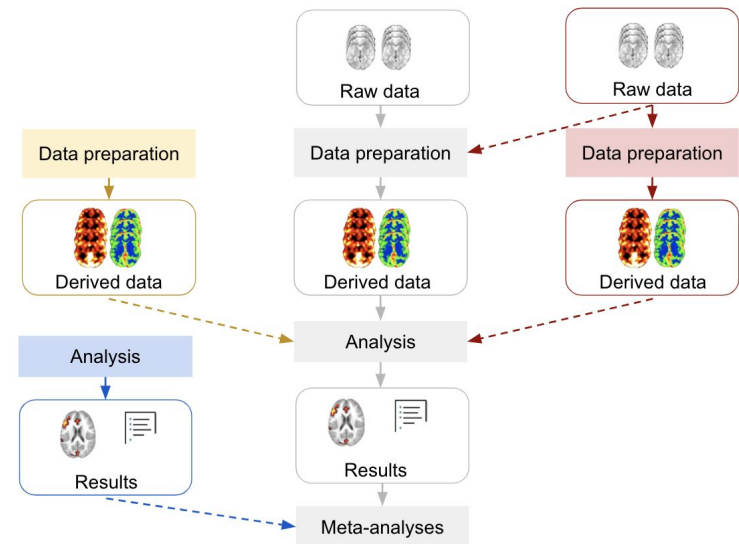
Data integration



A new level of
variability

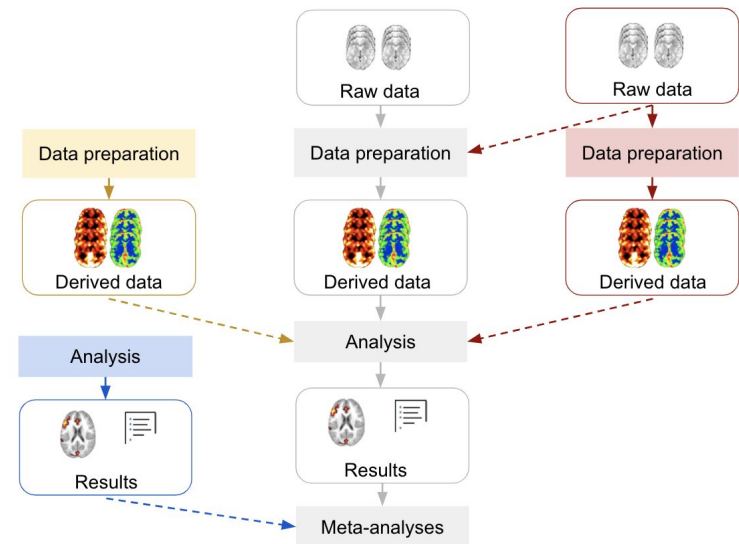
Levels of variability

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



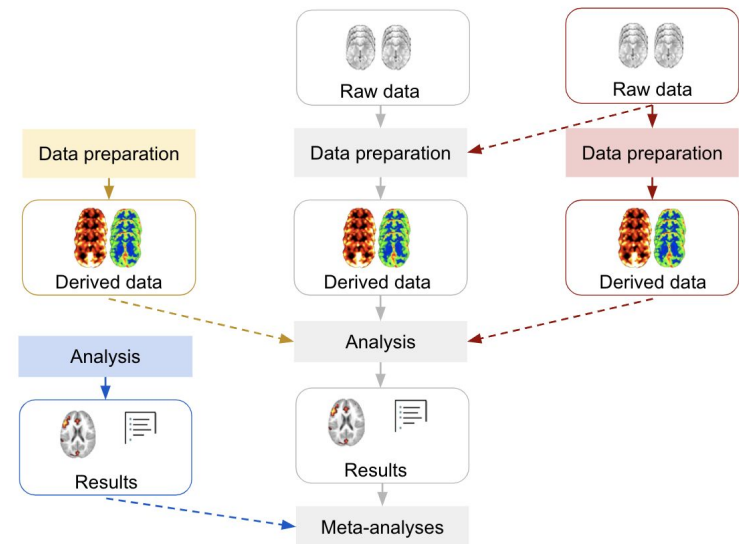
Levels of variability

- 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



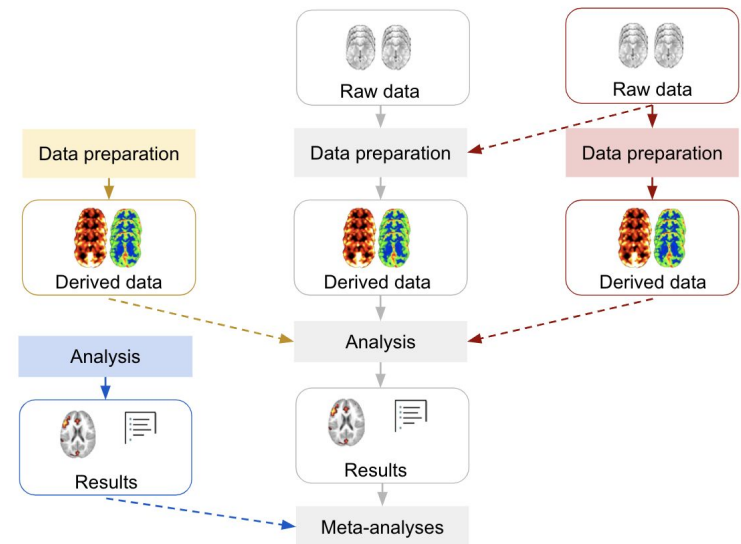
Levels of variability

- 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



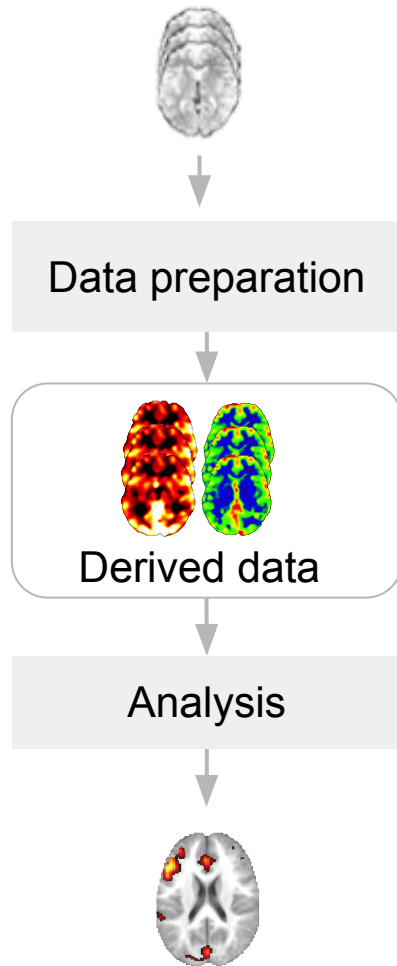
Levels of variability

- 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



Analytic Variability

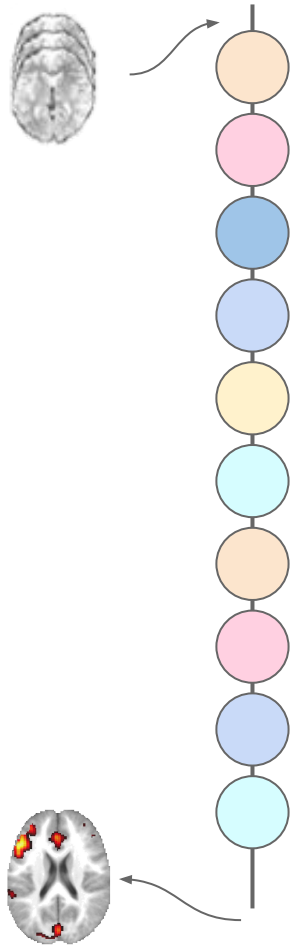
Analytic variability



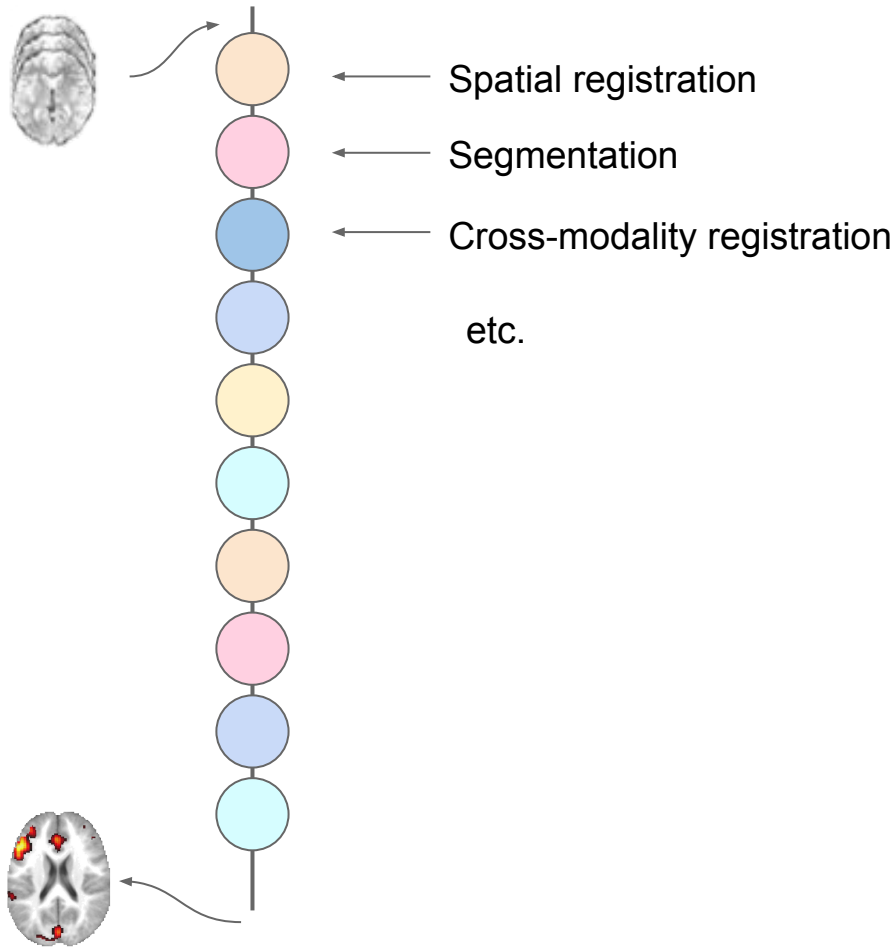
“Different **acceptable** analysis methods”

Carp et al. (2012)

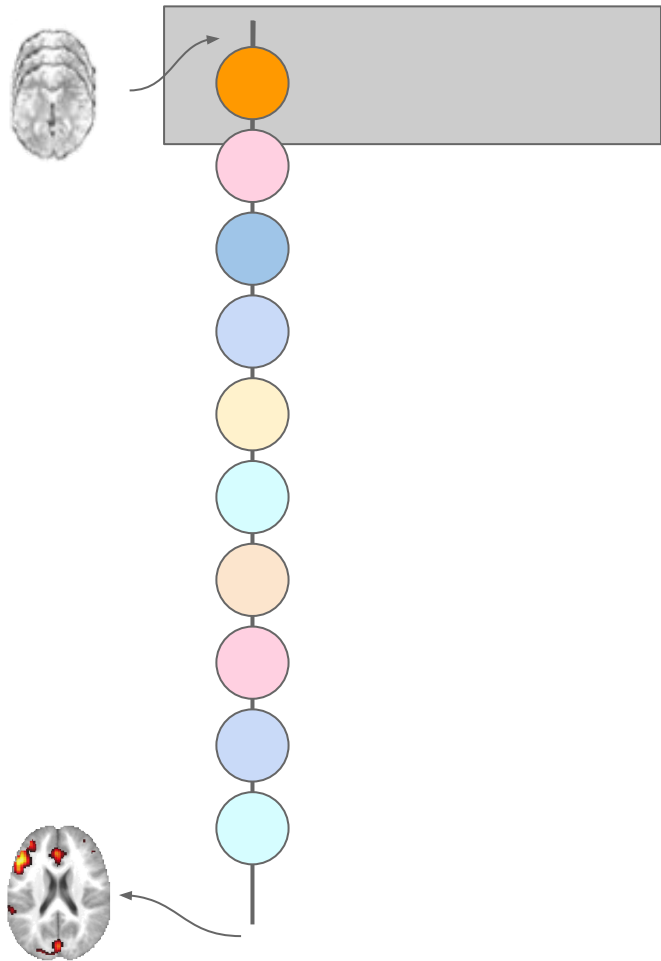
Analytic variability



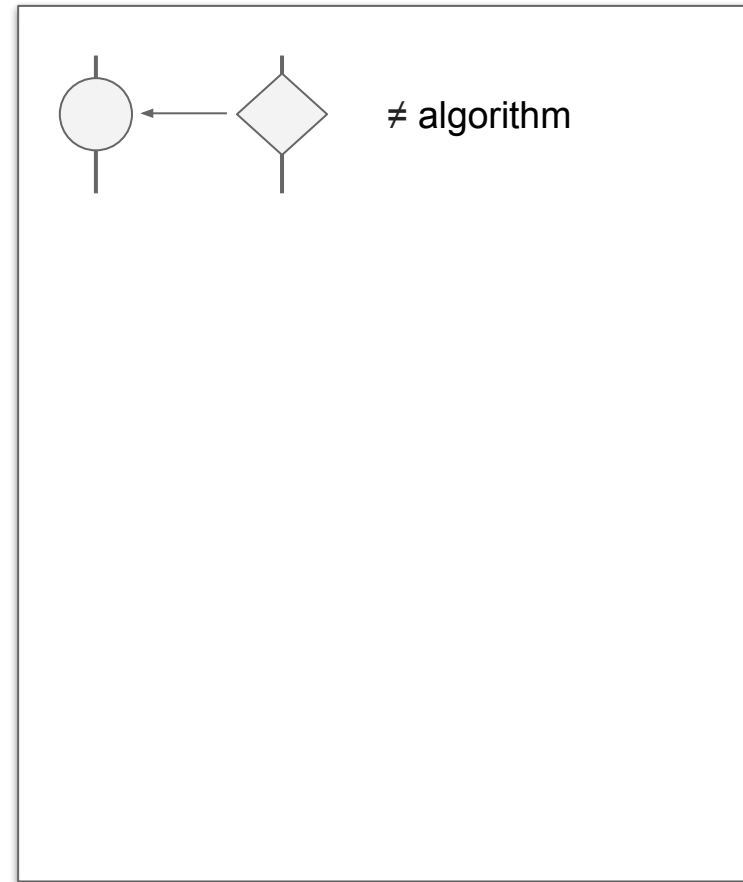
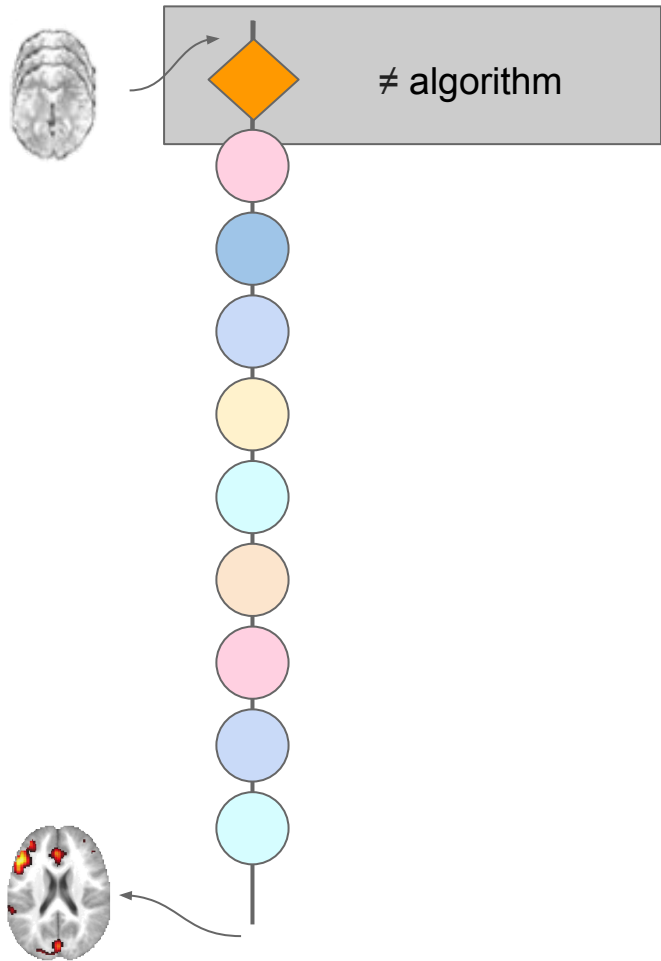
Analytic variability



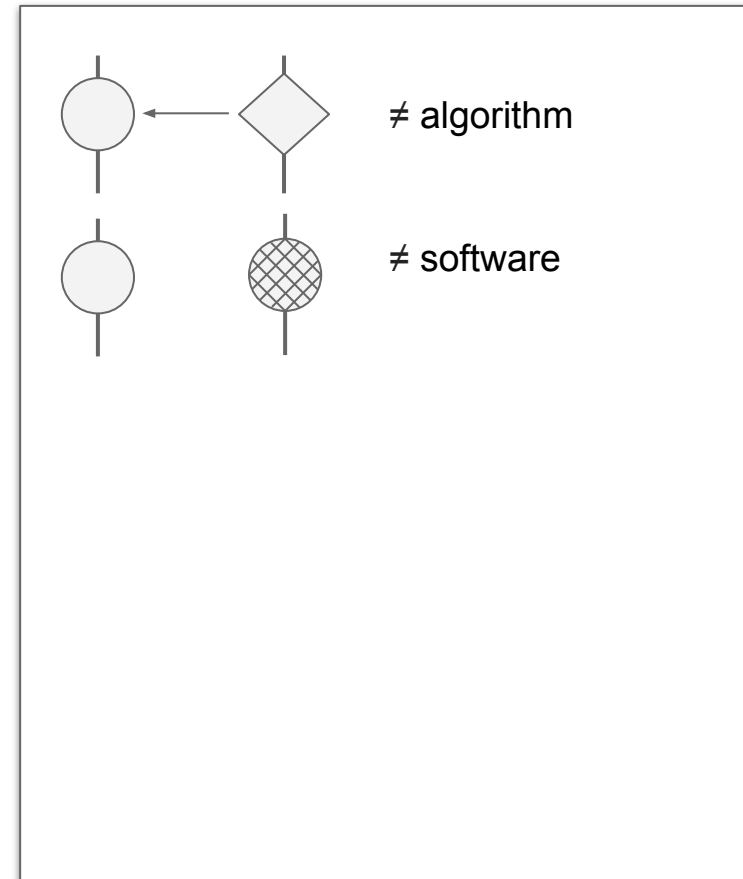
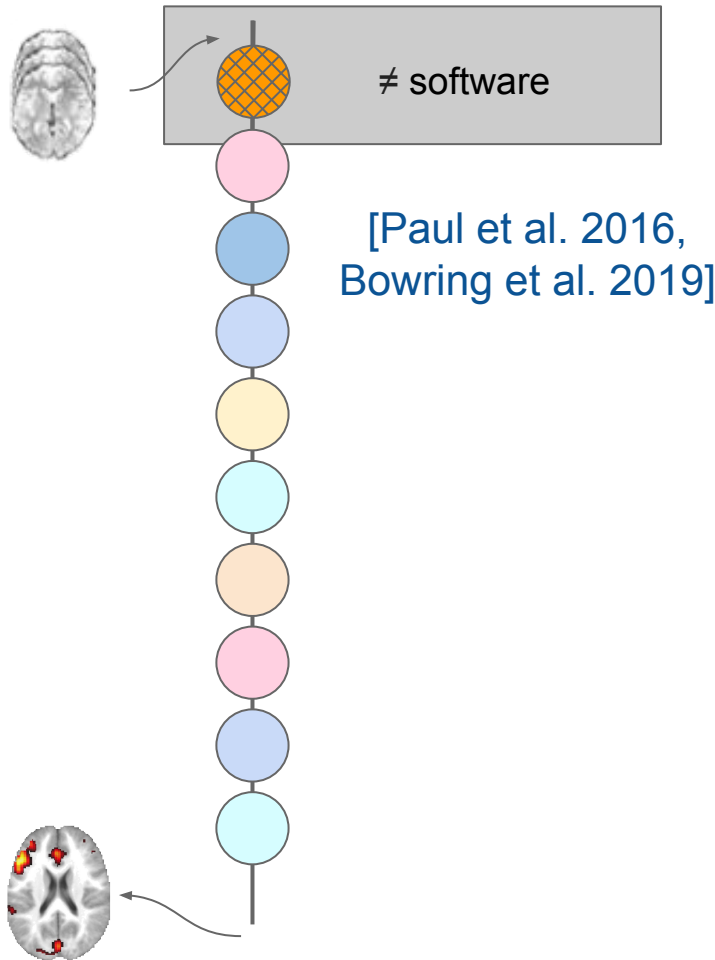
Analytic variability



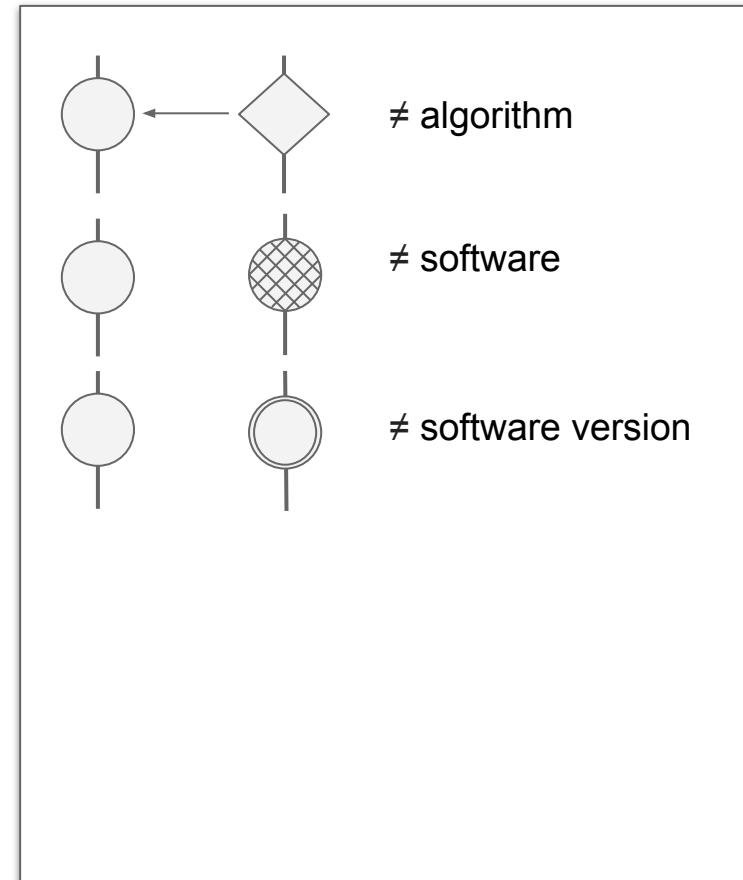
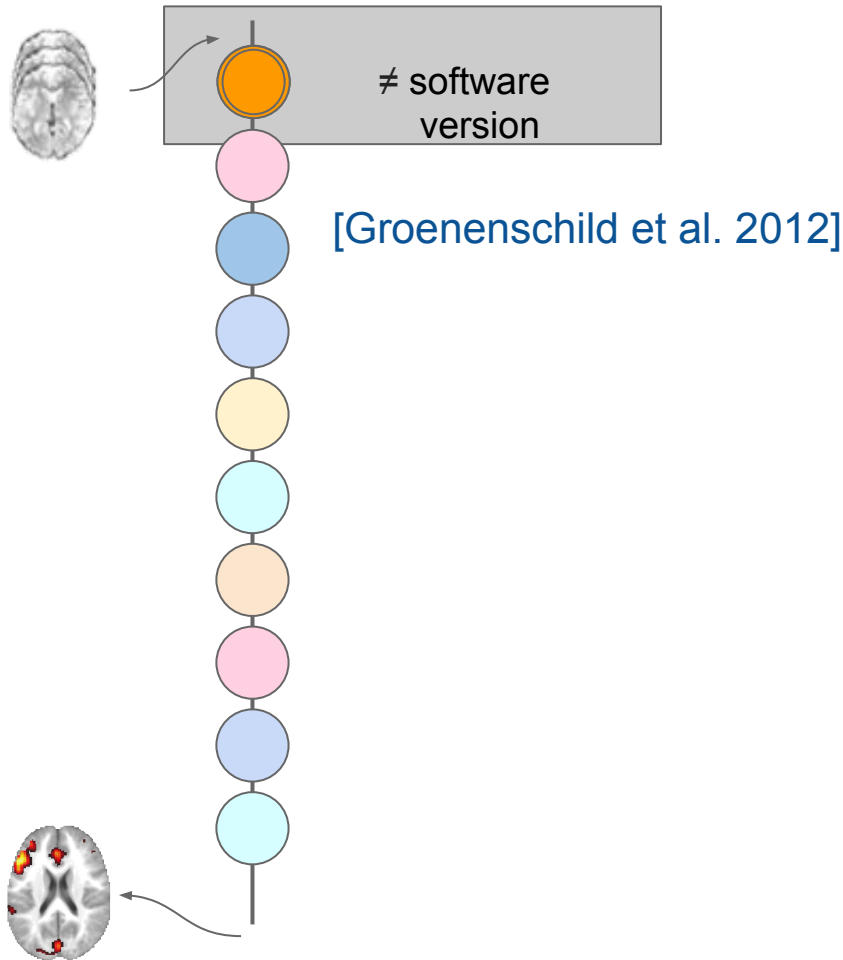
Analytic variability



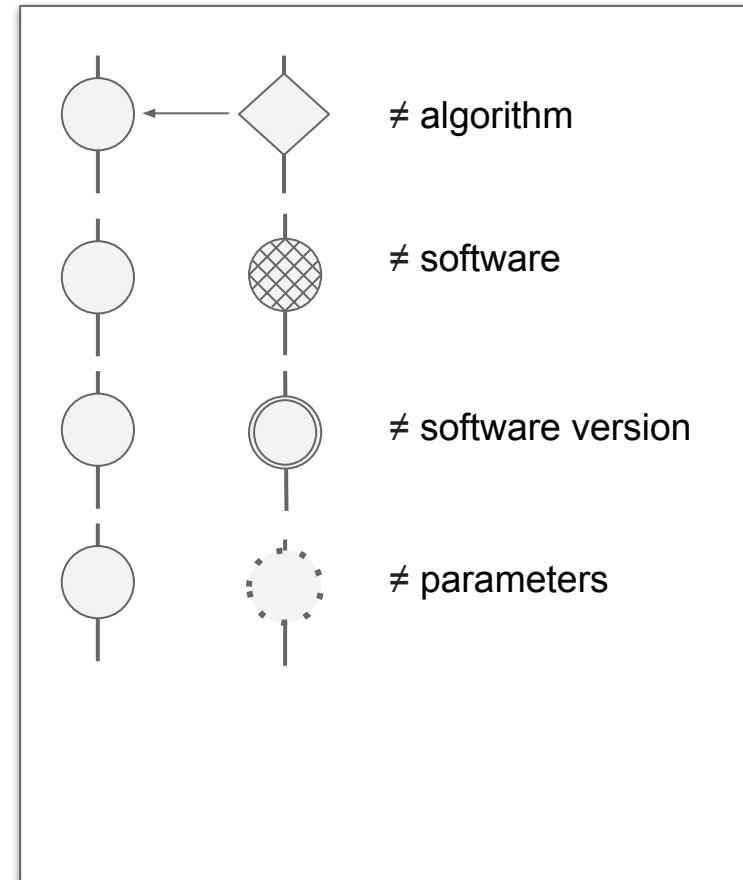
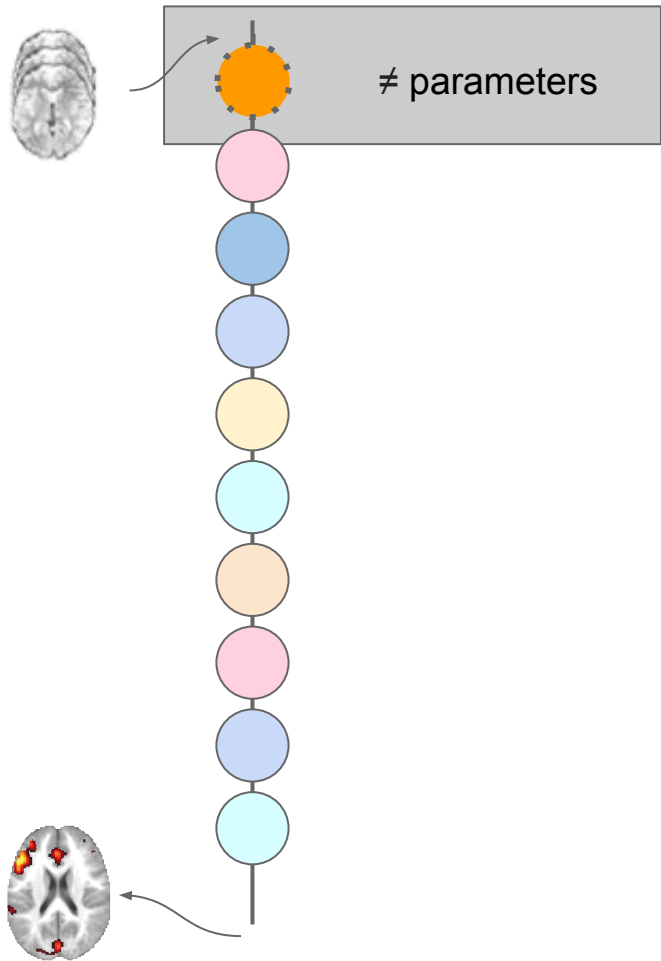
Analytic variability



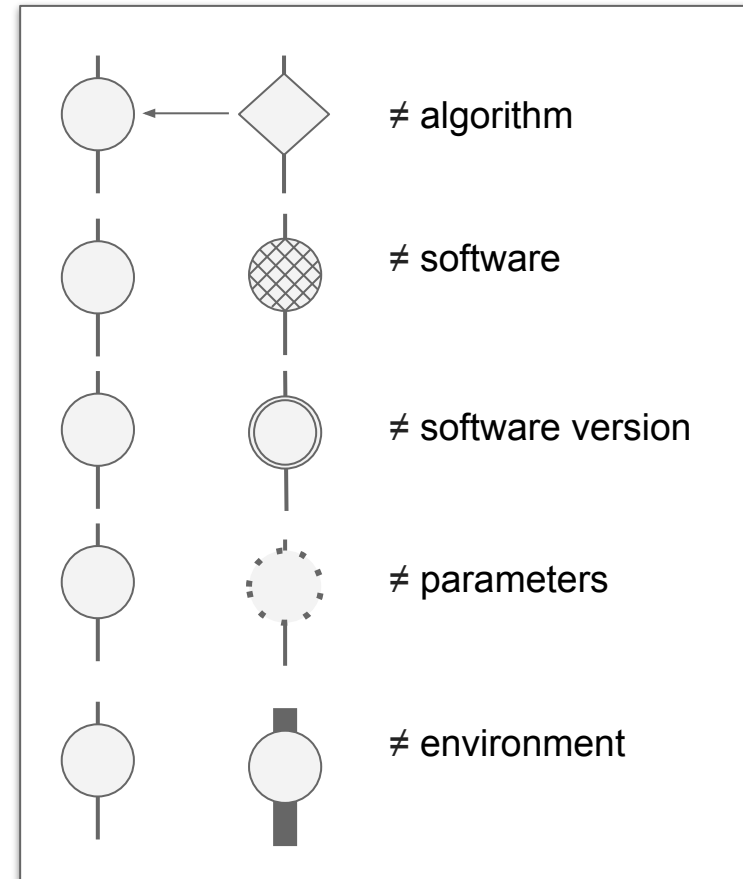
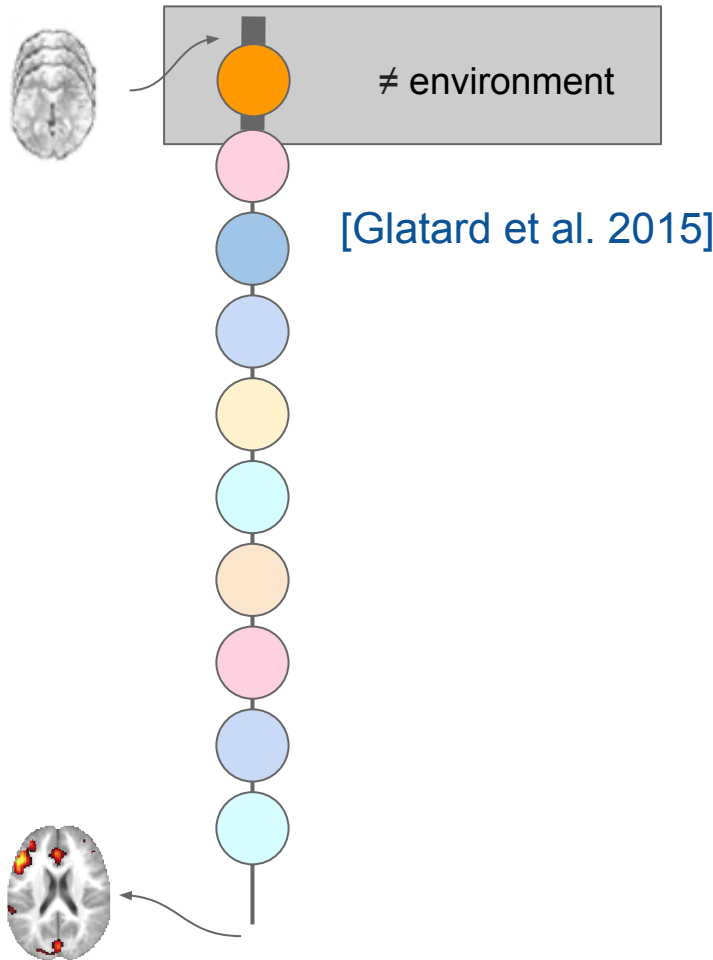
Analytic variability



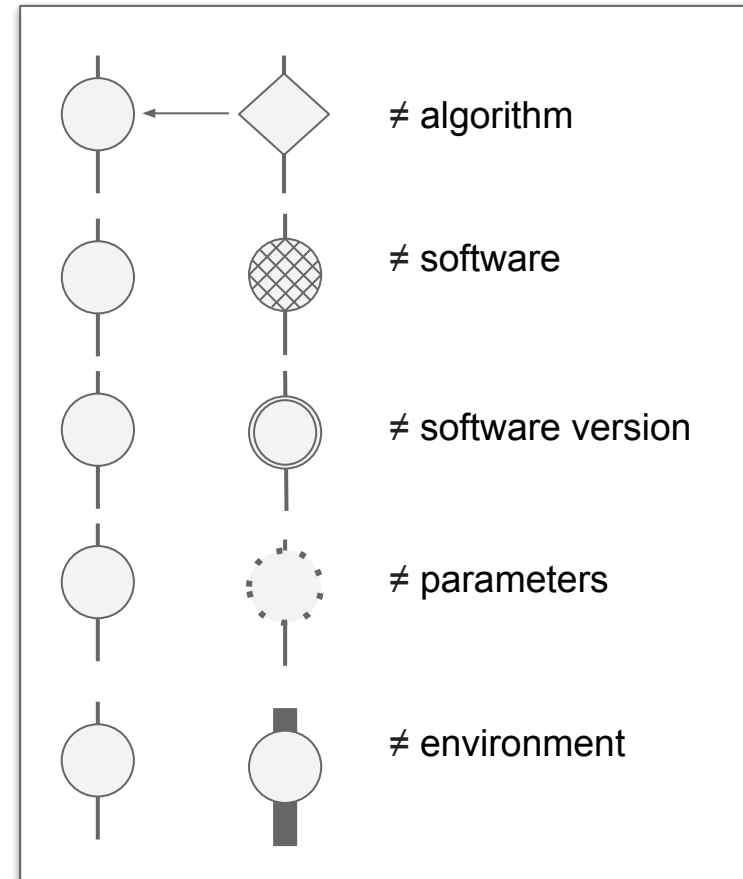
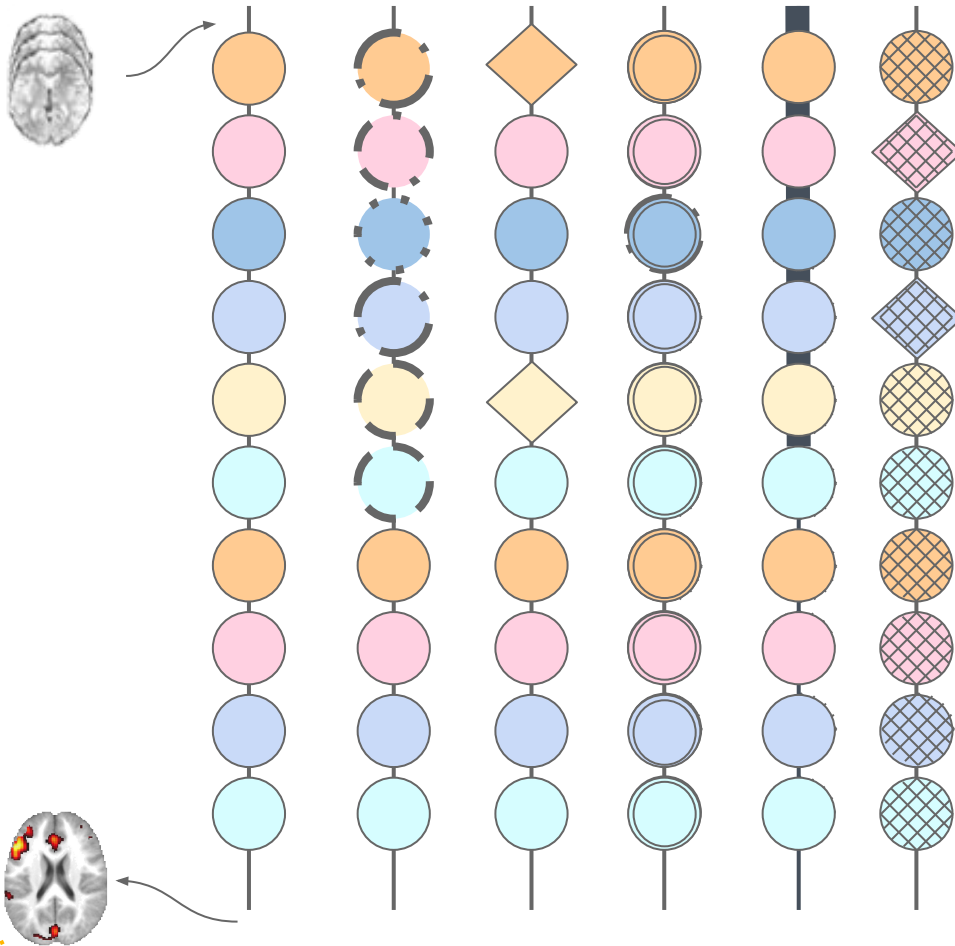
Analytic variability



Analytic variability



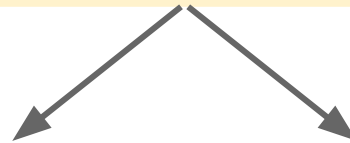
Analytic variability



**Towards data
integration** in brain
imaging



Towards **data integration**
in brain imaging studies

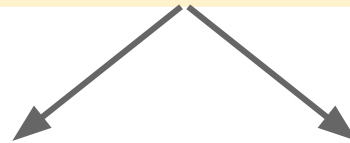


Standards for
data sharing

Statistical modelling
of analytic variability



Towards **data integration**
in brain imaging studies



Standards for
data sharing

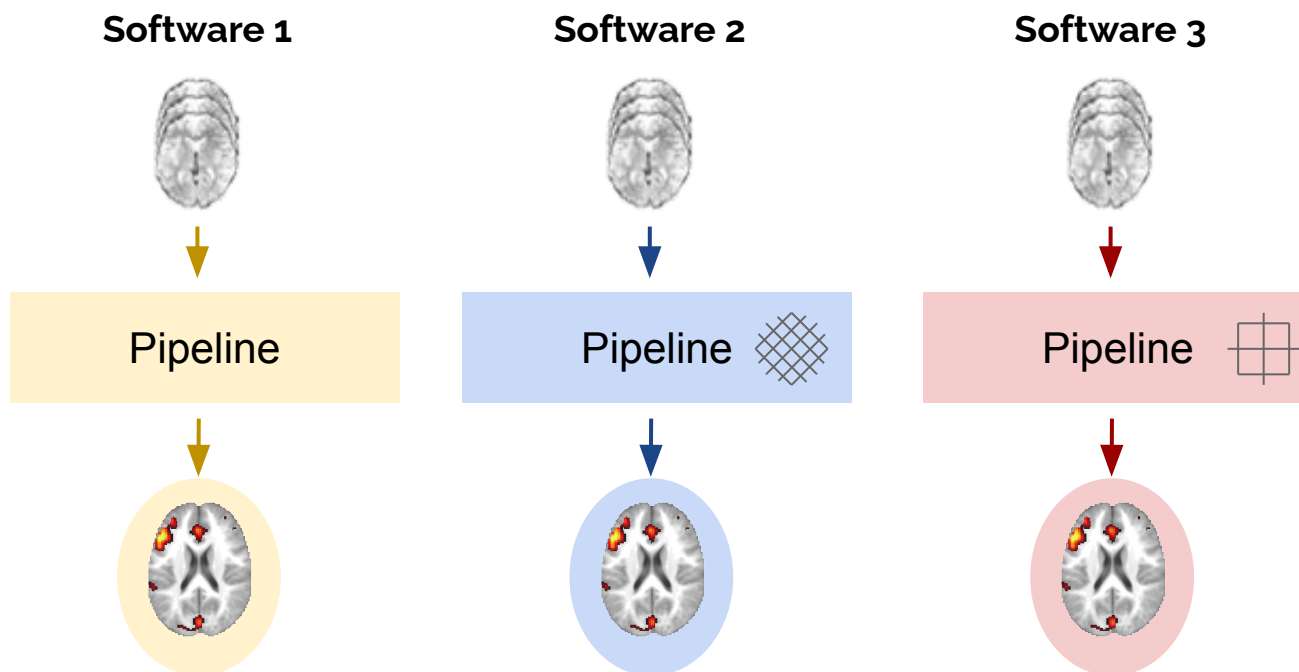
Statistical modelling
of analytic variability

Variations across software

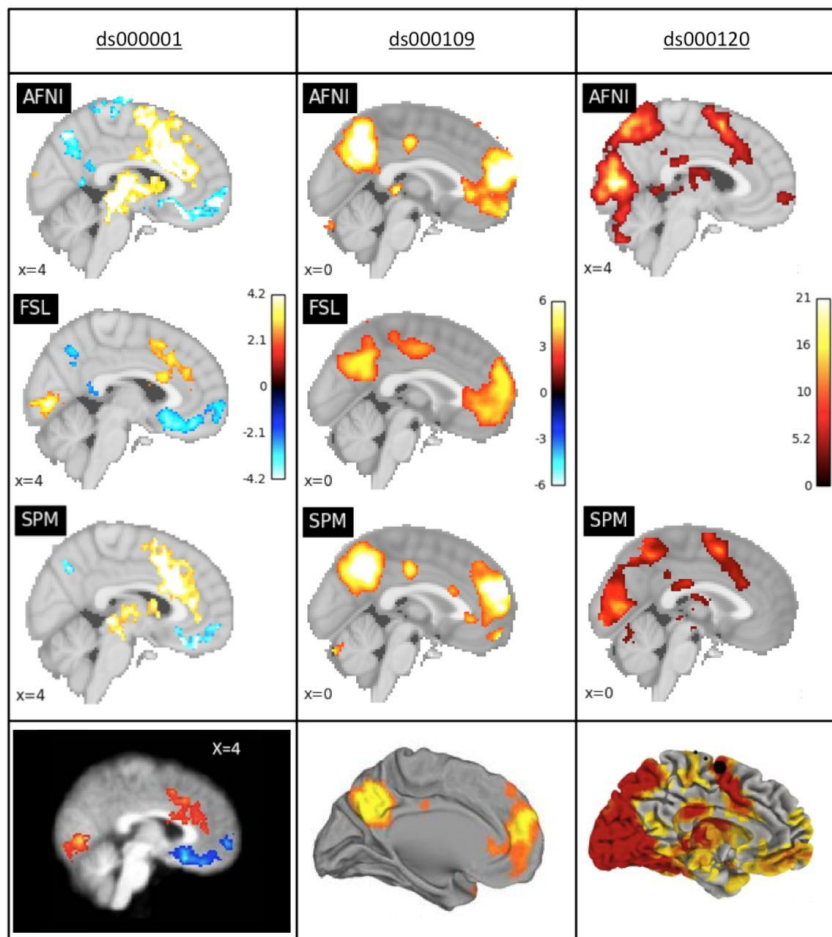
- Reproduced 3 published functional MRI studies
- Using 3 different software



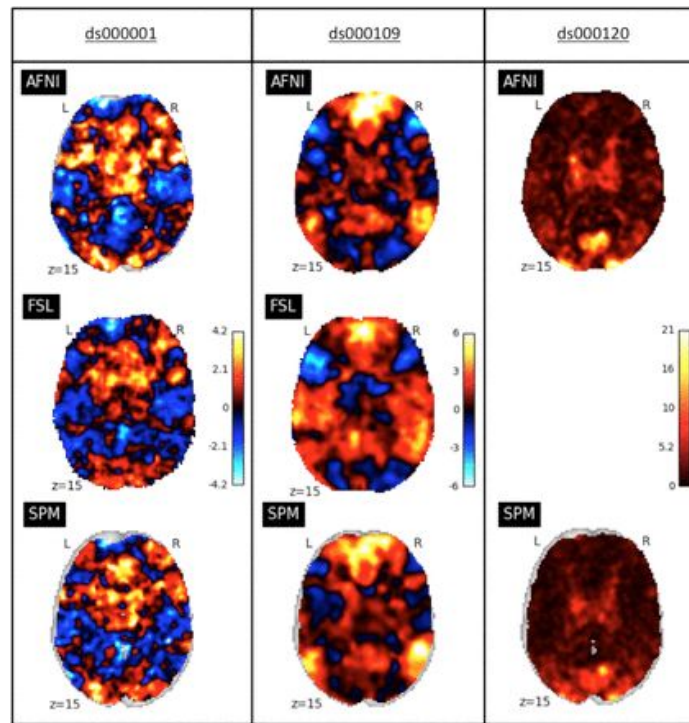
Alex Bowring Tom Nichols



Variations across software



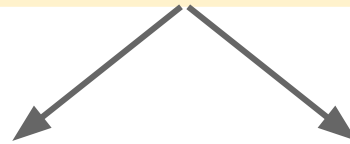
Comparison of the final results



Comparison of the statistic maps



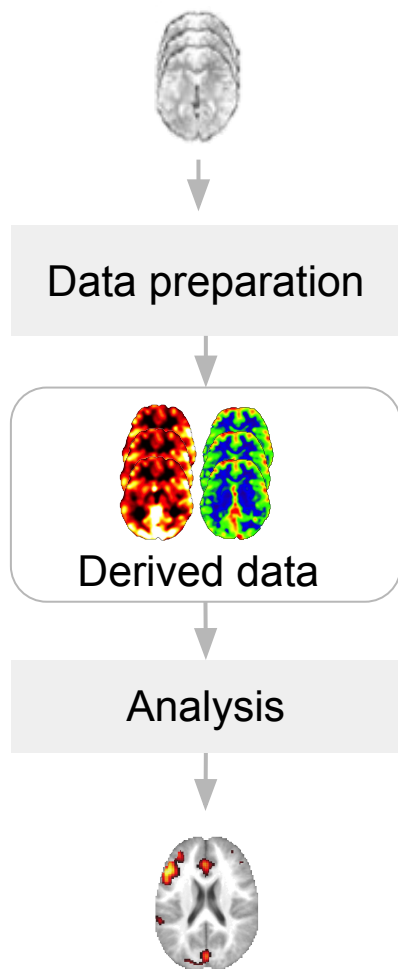
Towards **data integration**
in brain imaging studies



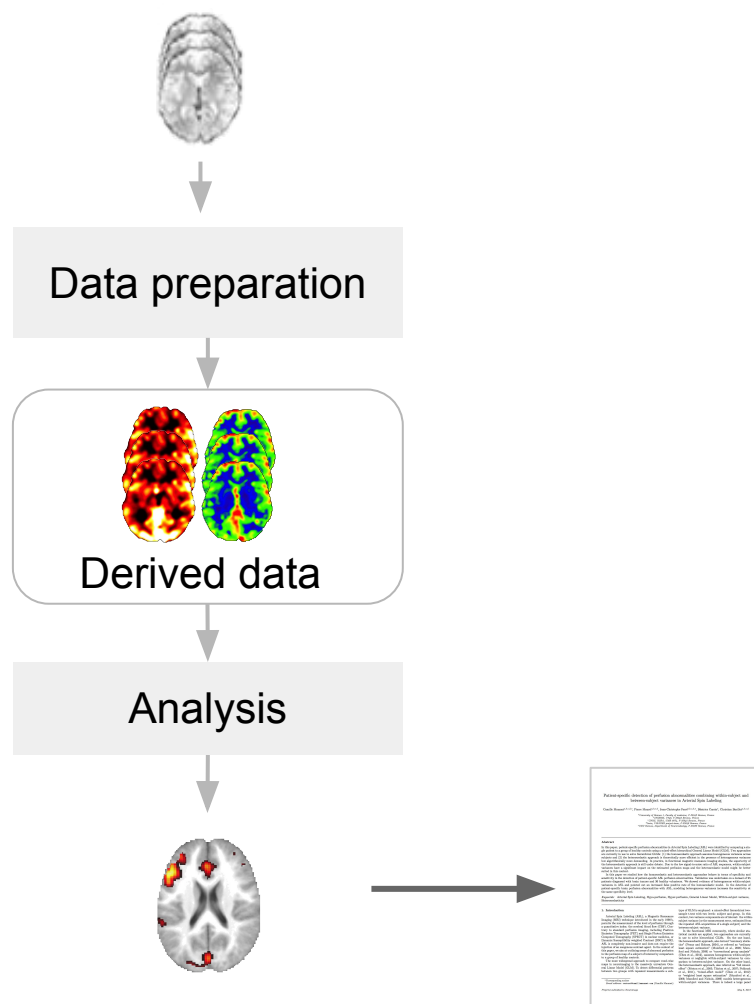
Standards for
data sharing

Statistical modelling
of analytic variability

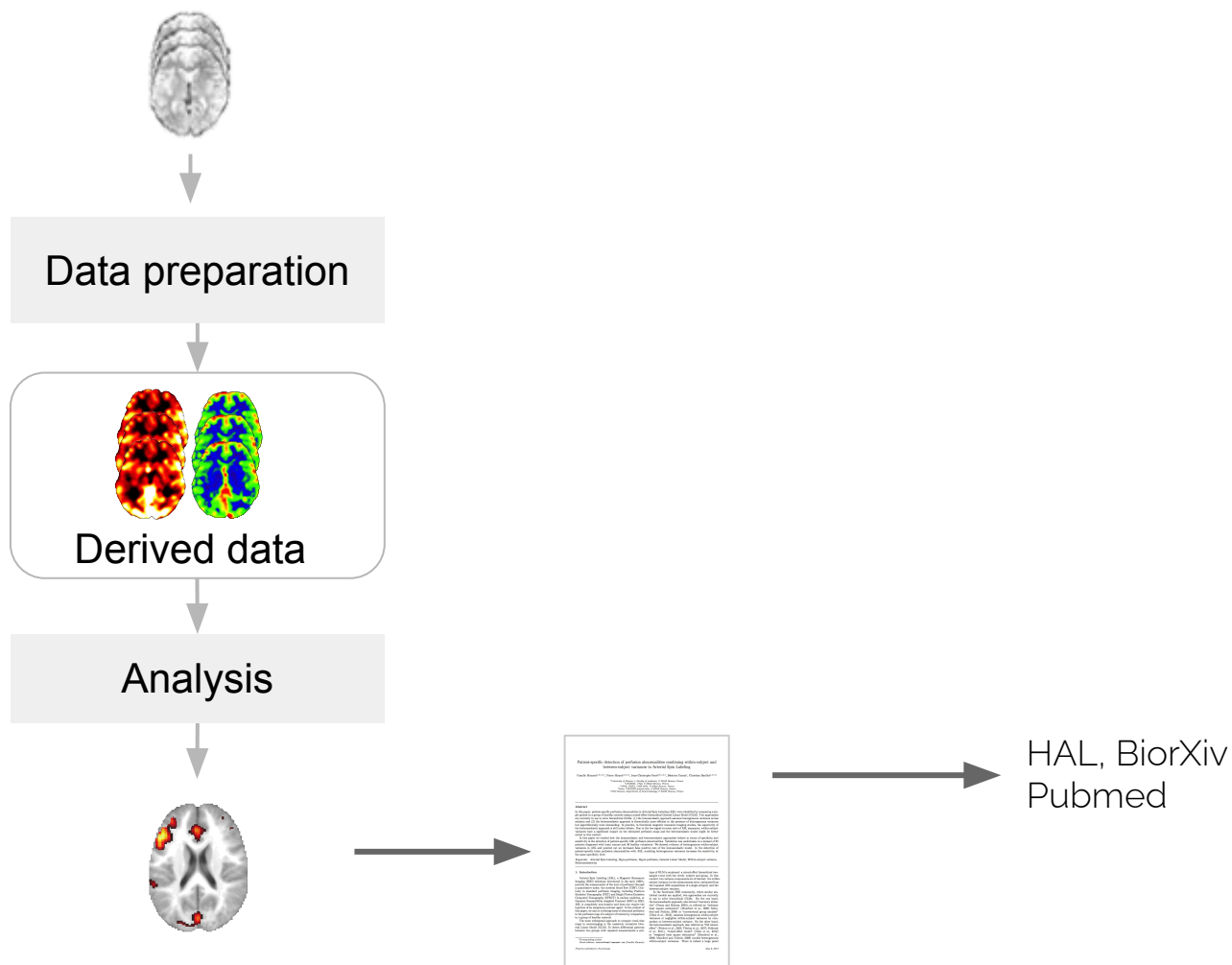
Sharing more research outputs



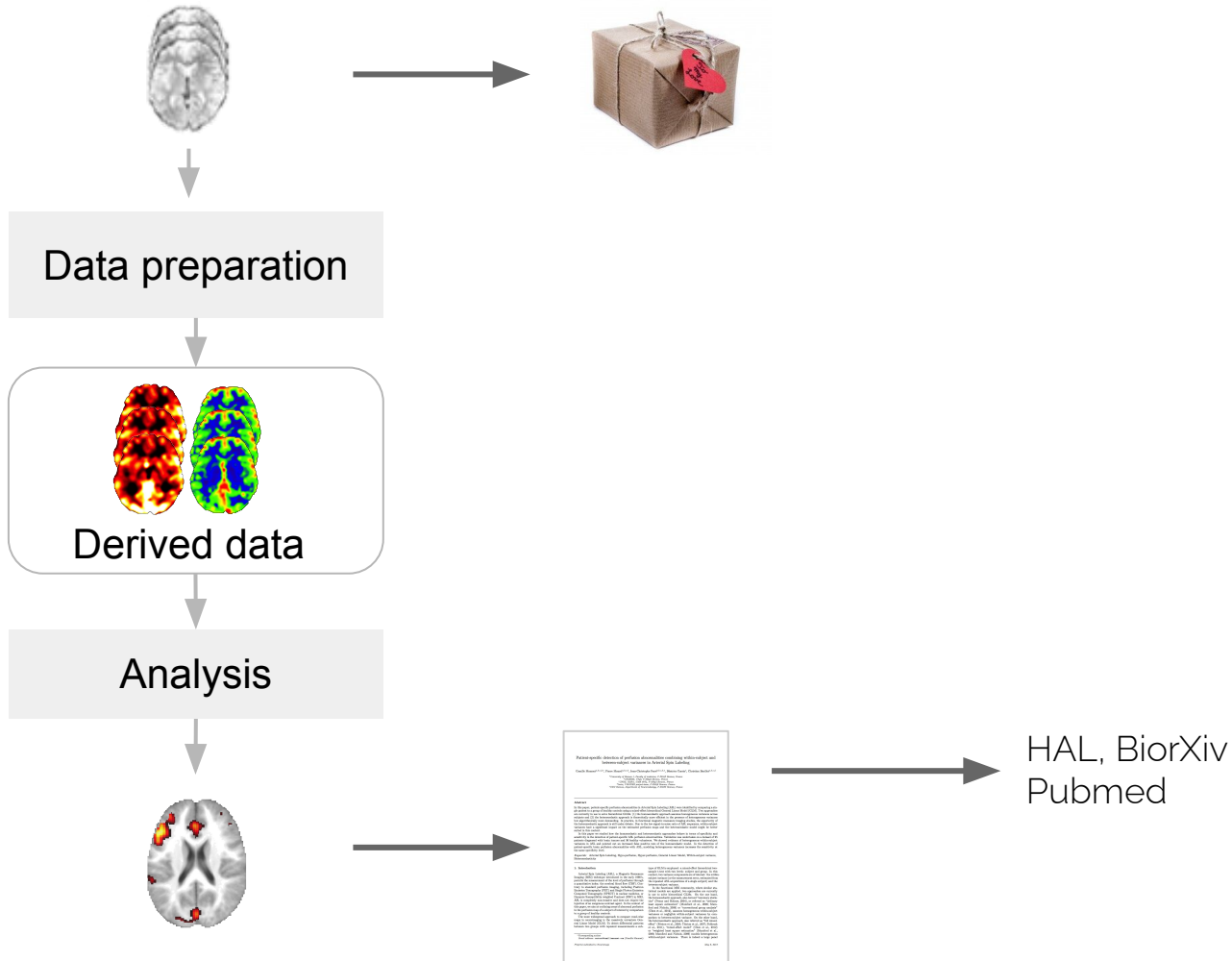
Sharing more research outputs



Sharing more research outputs



Sharing more research outputs



Brain Imaging Data Structure (BIDS)

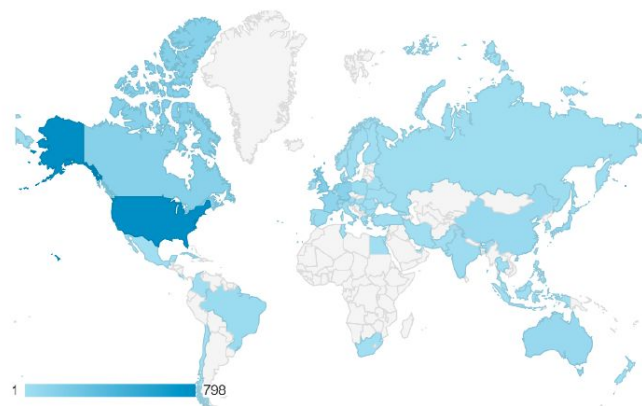
- Used in over 60 labs around the world
- Adopted by: FCP-INDI, Developing Human Connectome, SchizConnect and Donders Data repository.
- Extensions: MEG, iEEG, EEG



Krys
Gorgolewski

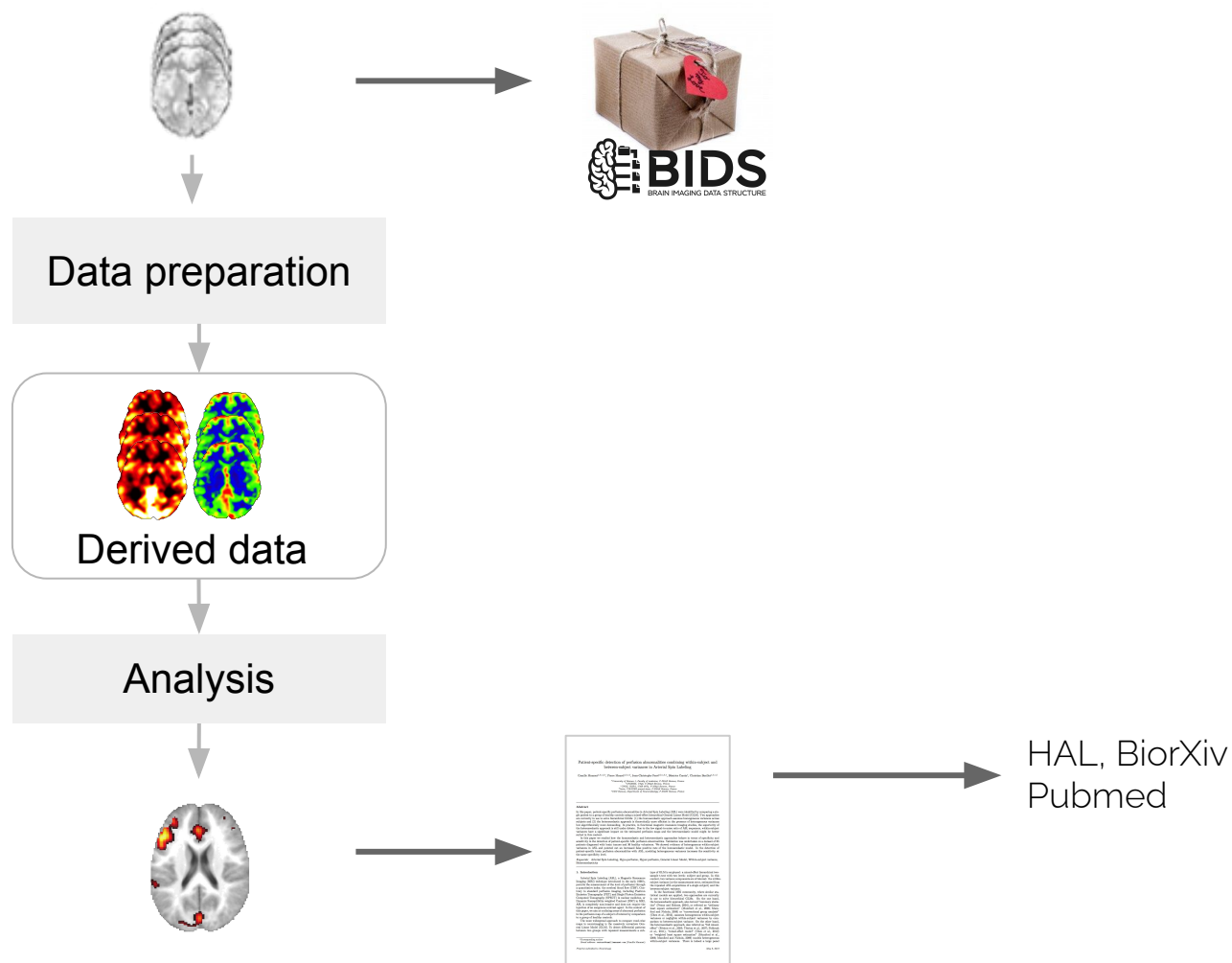
[Gorgolewski et al., *Sci. Data* 2016]

Traffic to bids.neuroimaging.io

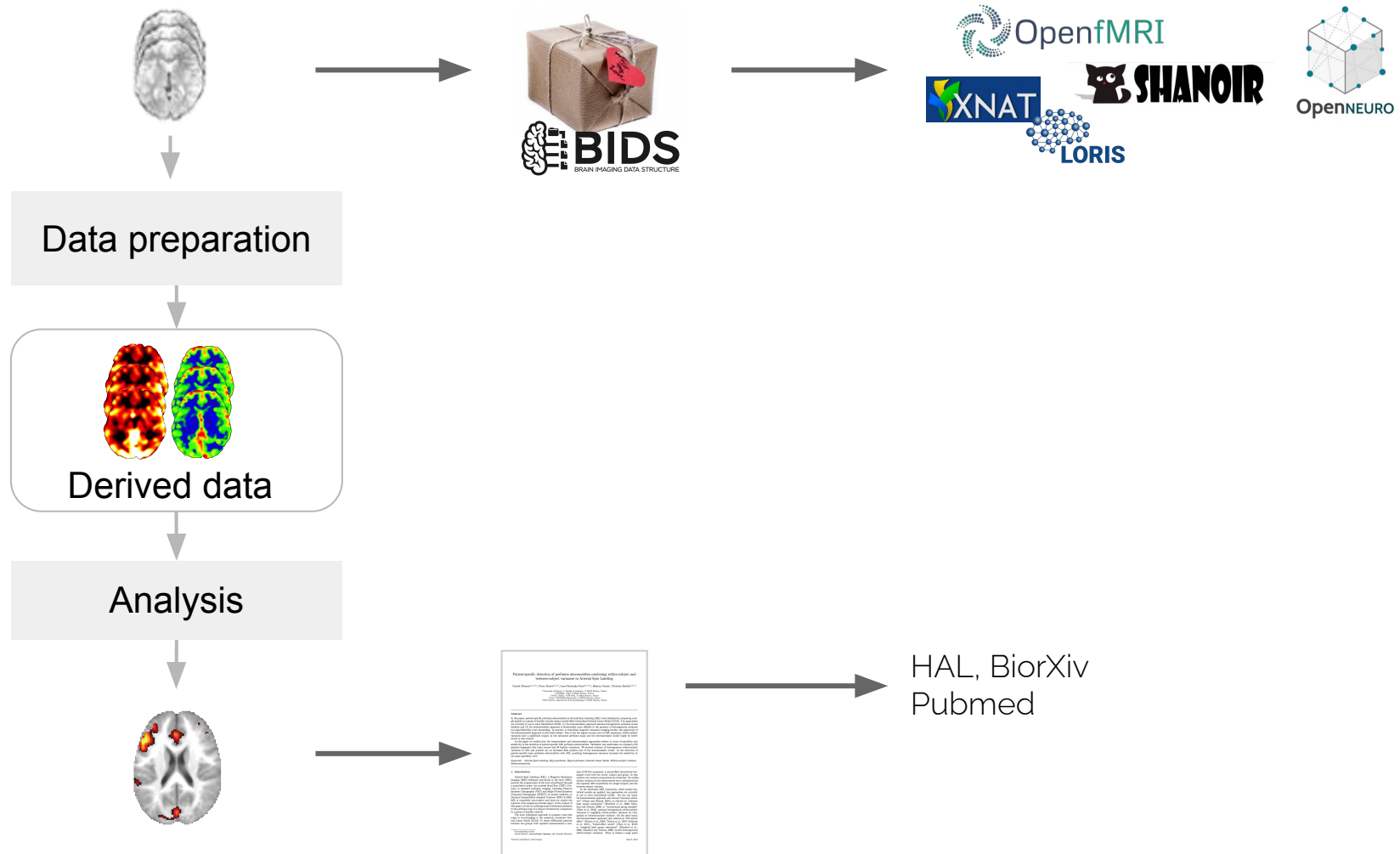


```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   └── func/  
│       ├── sub-01_task-rest_bold.nii.gz  
│       └── sub-01_task-rest_bold.json
```

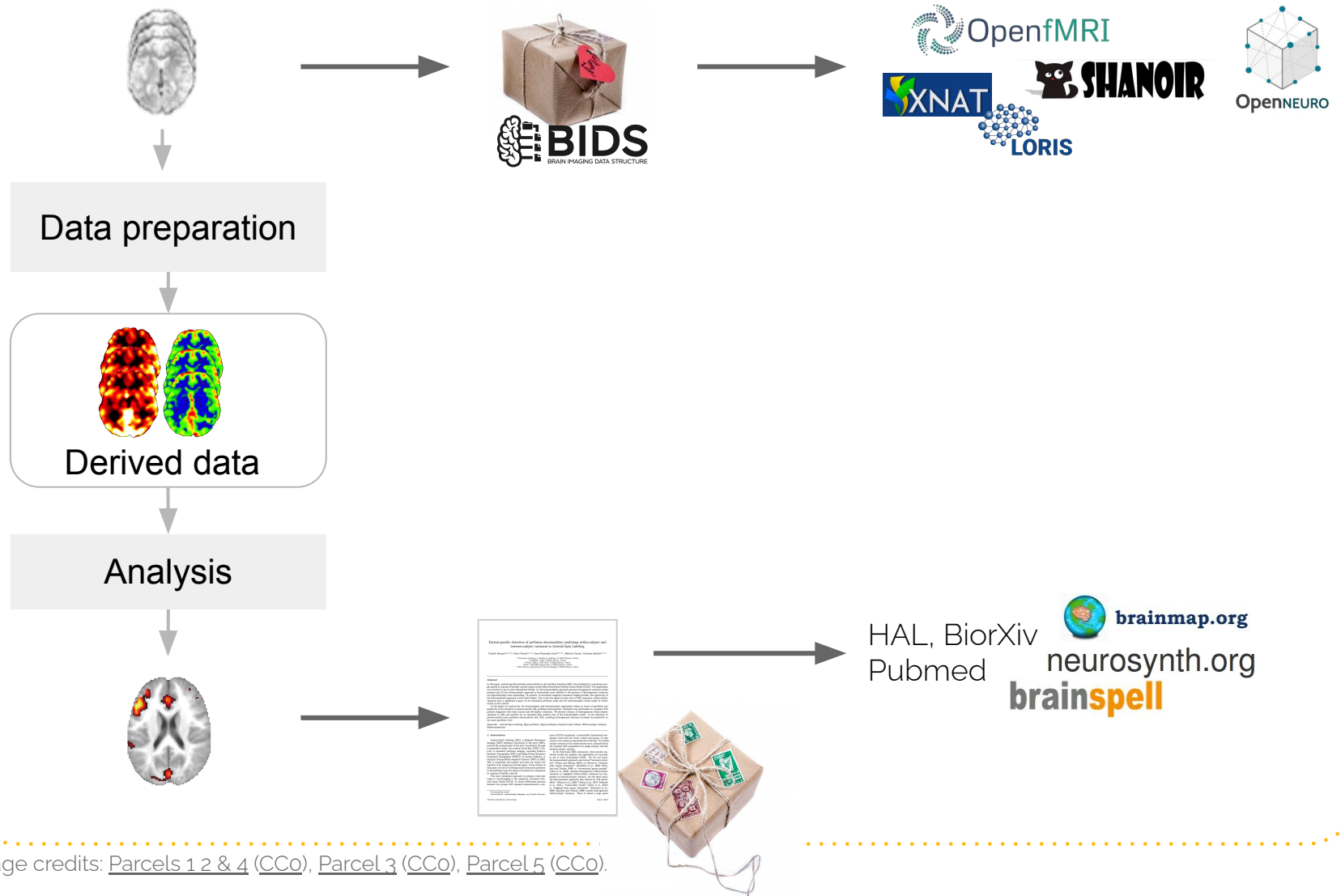
Sharing more research outputs



Sharing more research outputs

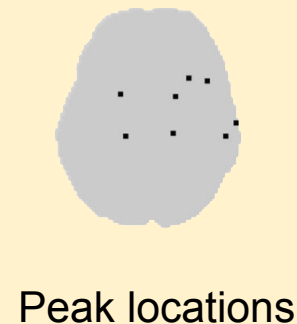
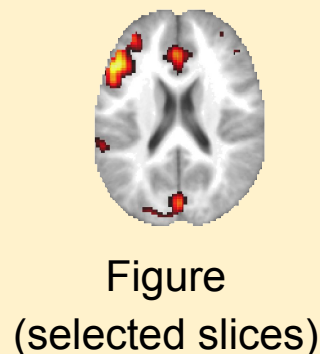


Sharing more research outputs



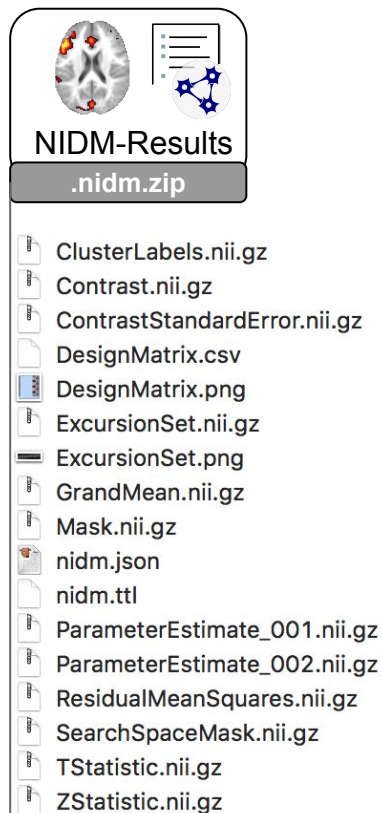
fMRI Results

Publication



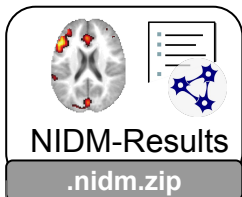
- ✘ Incomplete statistical results
- ✘ Ambiguous/incomplete methods
- ✘ Metadata is not searchable

NIDM-Results pack



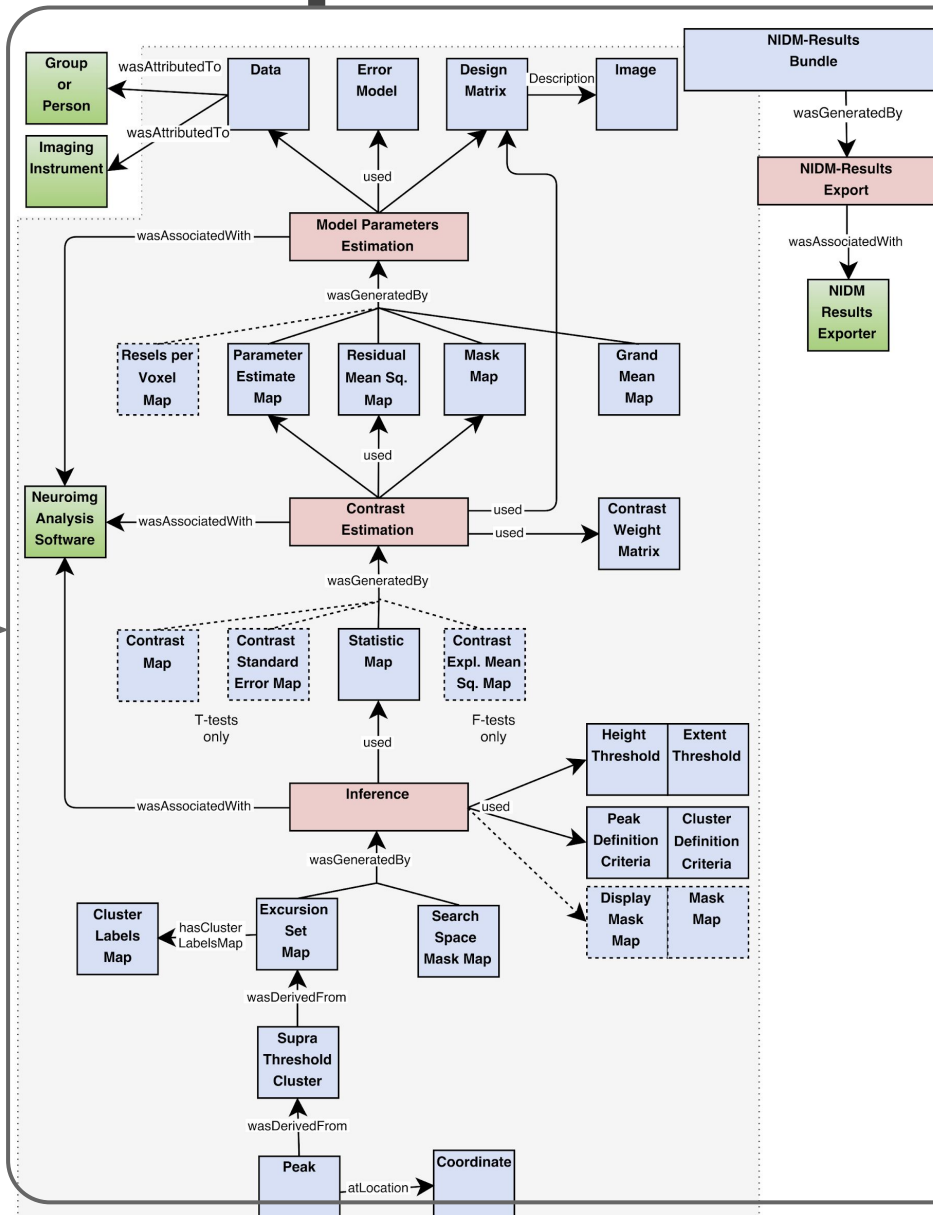
[Maumet et al.,
Sci. Data 2016]

NIDM-Results pack



- ClusterLabels.nii.gz
- Contrast.nii.gz
- ContrastStandardError.nii.gz
- DesignMatrix.csv
- DesignMatrix.png
- ExcursionSet.nii.gz
- ExcursionSet.png
- GrandMean.nii.gz
- Mask.nii.gz
- nidm.json
- nidm.ttl
- ParameterEstimate_001.nii.gz
- ParameterEstimate_002.nii.gz
- ResidualMeanSquares.nii.gz
- SearchSpaceMask.nii.gz
- TStatistic.nii.gz
- ZStatistic.nii.gz

[Maumet et al.,
Sci. Data 2016]



Sharing more research outputs

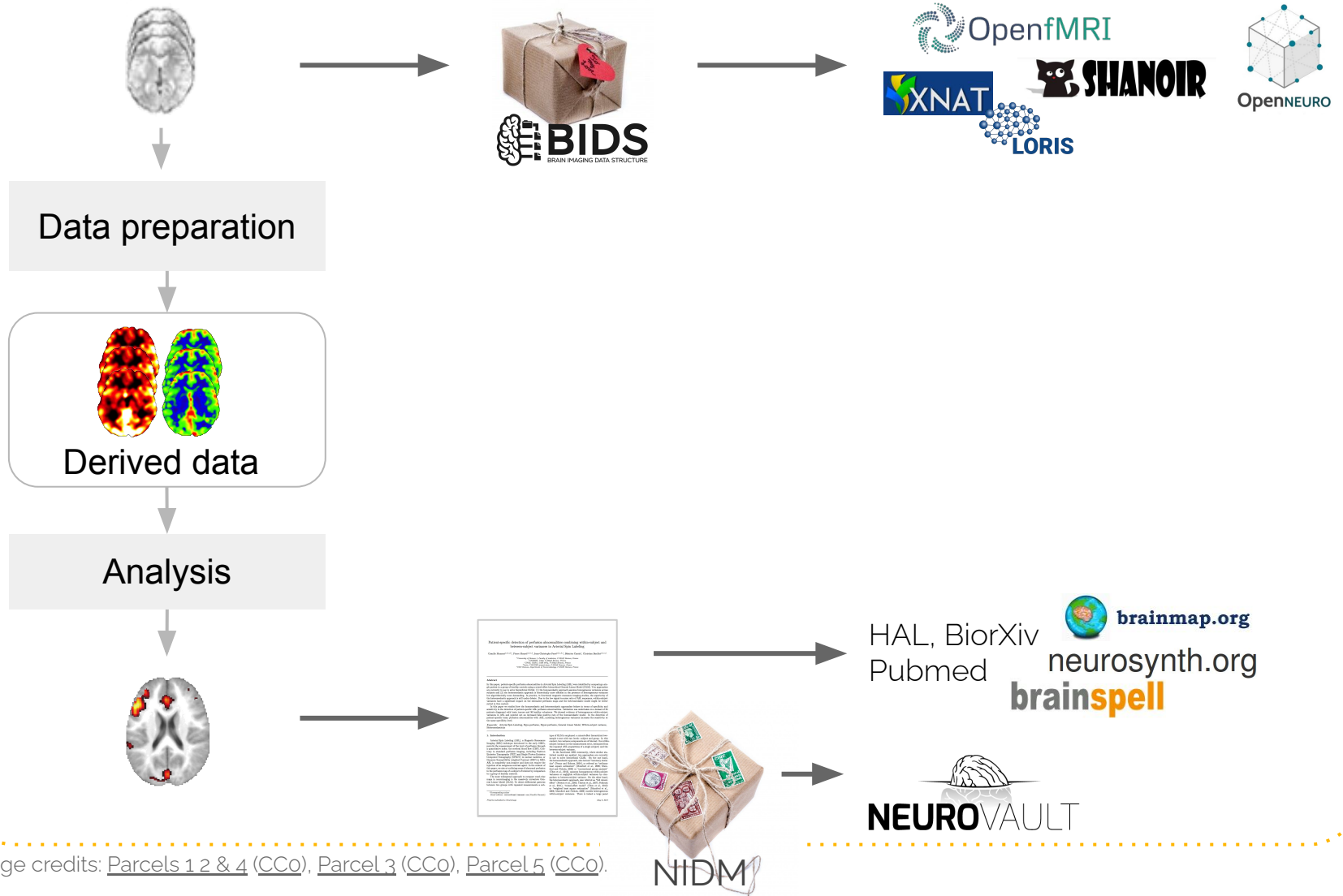
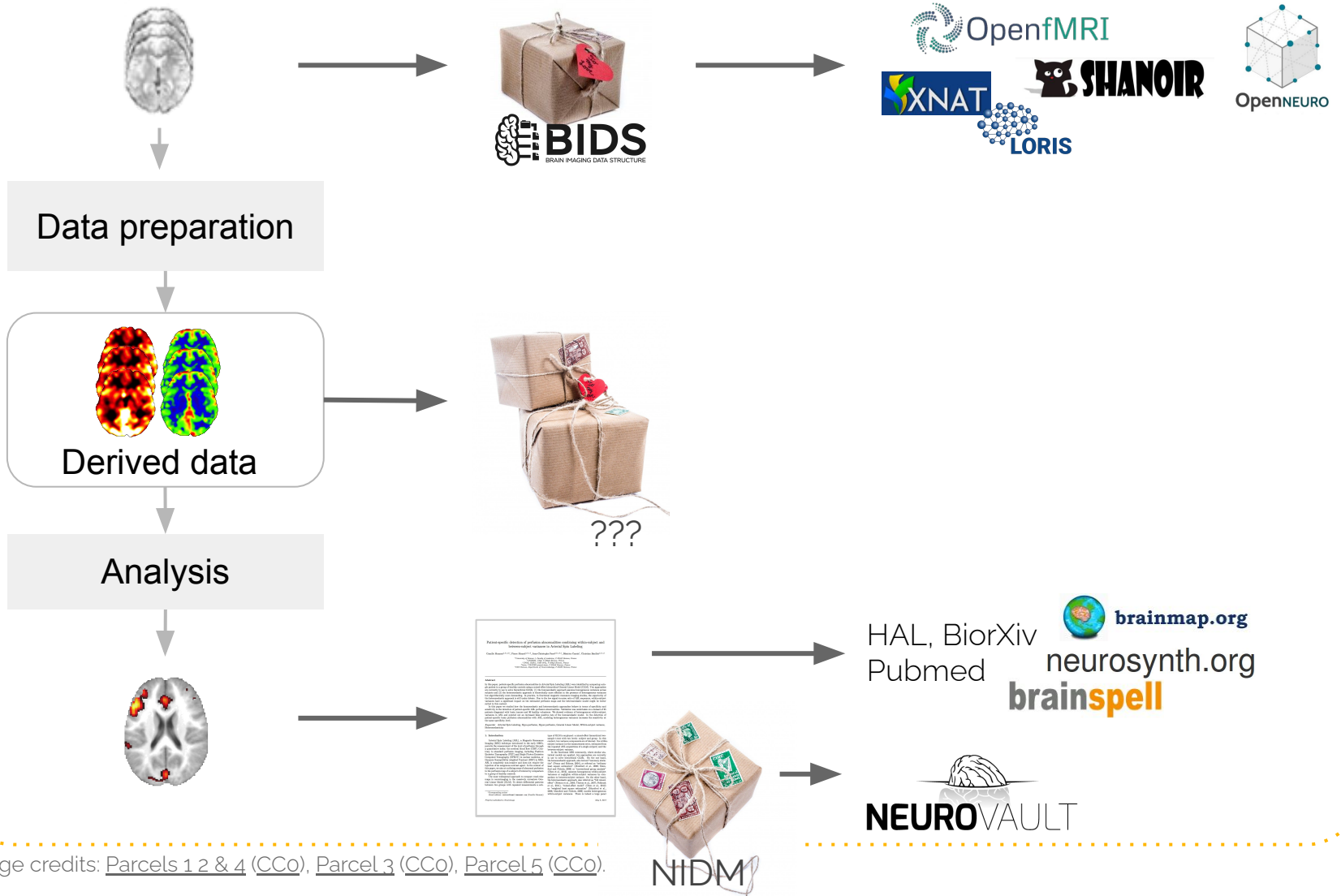


Image credits: [Parcels 1 2 & 4](#) (CC0), [Parcel 3](#) (CC0), [Parcel 5](#) (CC0).

Sharing more research outputs



BIDS Prov **Bridging BIDS & NIDM**

BIDS


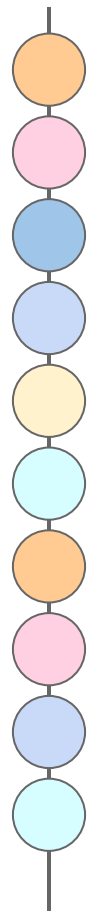


```
my_dataset/  
├── participants.tsv  
├── sub-01/  
│   ├── anat/  
│   │   └── sub-01_T1w.nii.gz  
│   └── func/
```

File naming convention + JSON files

—

60 labs worldwide



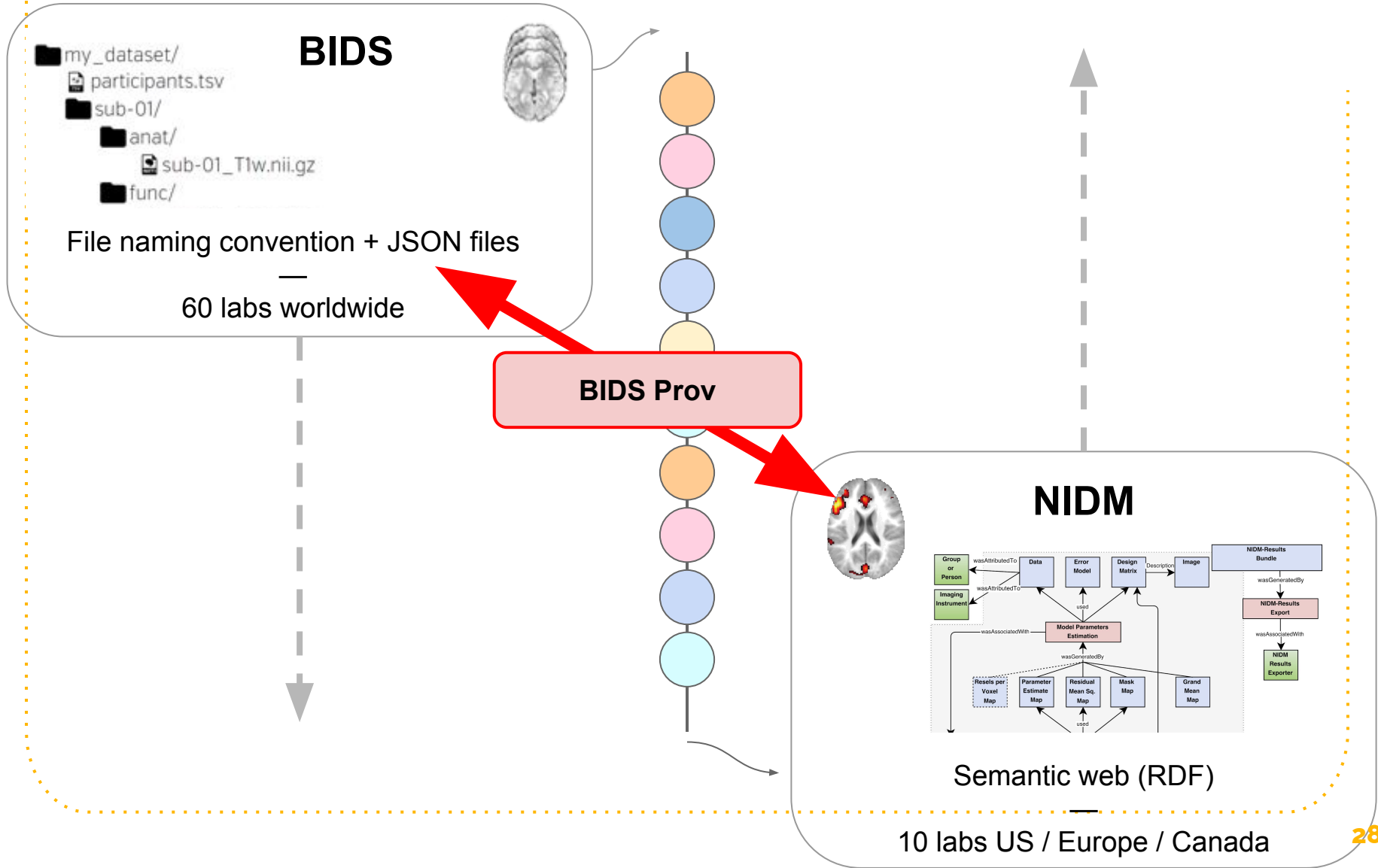
NIDM

Semantic web (RDF)

—

10 labs US / Europe / Canada

BIDS Prov Bridging BIDS & NIDM



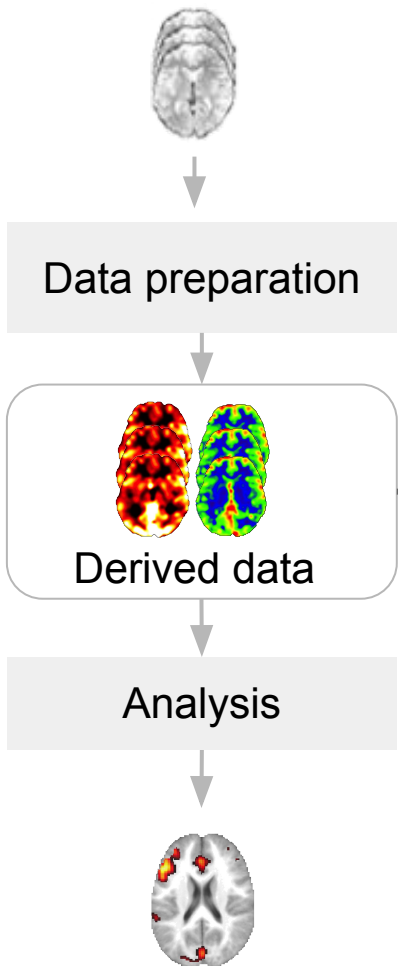


Satra Ghosh



Rémi Adon

BIDS Provenance



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

Bowring, A., Maumet*, C., & Nichols*, T. E. (2019). Exploring the impact of analysis software on task fMRI results. *Human Brain Mapping*, 0(0). <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>

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... 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>

Maumet, C., Auer, T., Bowring, A., Chen, G., Das, S., Flandin, G., Ghosh, S., Glatard, T., Gorgolewski, K. J., Helmer, K. G., Jenkinson, M., Keator, D. B., Nichols, B. N., Poline, J.-B., Reynolds, R., Sochat, V., Turner, J., & Nichols, T. E. (2016). Sharing brain mapping statistical results with the neuroimaging data model. *Scientific Data*, 3, 160102. <https://doi.org/10.1038/sdata.2016.102>

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>

Towards **large-scale** brain imaging studies

Camille Maumet



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

Thank you!