



**HAL**  
open science

## Towards reproducible neuroimaging: Solutions for sharing and re-using brain imaging data

Camille Maumet

### ► To cite this version:

Camille Maumet. Towards reproducible neuroimaging: Solutions for sharing and re-using brain imaging data. Neurosciences. University of Rennes, 2024. tel-04458543

**HAL Id: tel-04458543**

**<https://inria.hal.science/tel-04458543v1>**

Submitted on 14 Feb 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



Mémoire pour l'obtention de  
**L'HABILITATION À DIRIGER LES RECHERCHES**  
de l'Université de Rennes

Présenté par  
Camille MAUMET

**Towards reproducible neuroimaging:  
Solutions for sharing and re-using brain  
imaging data**

Présentée le 8 février 2024

Devant un jury composé de :

<i>Rapporteuse</i> :	Carole Lartizien	- Directrice de Recherche, CNRS, France
<i>Rapporteuse</i> :	Jeanette Mumford	- Research Scientist, Stanford Uni., USA
<i>Rapporteur</i> :	Adam Thomas	- Team Lead, NIMH, USA
<i>Examinatrice</i> :	Sarah Cohen-Boulakia	- Professeure, Uni. Paris Saclay, France
<i>Examinatrice</i> :	Anne-Laure Boulesteix	- Professeure, LMU Munich, Germany
<i>Examinatrice</i> :	Elisa Fromont	- Professeure, Uni. Rennes, France



# Foreword

This document presents an overview of my research activities in neuroimaging reproducibility for the past 10 years (since my PhD defense in May 2013).

I have been a postdoctoral fellow at the University of Warwick and Oxford University in the UK (2013-2017). Then, I have been a research scientist at Inria in the VisAGes (and then Empenn) team since 2017. After contributing as a PhD student to the field of neuroimaging statistics for clinical datasets, as a postdoctoral fellow I decided to make neuroinformatics the focus of my research. This change was driven by my growing interest in reproducible neuroimaging. Over the years I have also increasingly contributed to Open Science.

I have co-supervised 1 PhD student: Xavier Rolland (with Dr. Christian Barillot and Dr. Pierre Maurel), who successfully defended his thesis in May 2022. I have co-advised 4 international PhD students on projects that were part of their thesis: Gregory Kiar (McGill Uni., Canada, advised by Dr. Tristan Glatard and Prof. Alan Evans), Freya Acar (Uni. of Ghent, Belgium, advised by Dr. Beatrijs Moerkerke), Alexander Bowring (University of Oxford, UK, advised by Prof. Thomas Nichols) and Ruth Pauli (Uni. of Birmingham, UK, advised by Prof. Thomas Nichols in her pre-lab year). I have co-advised 3 post-doctoral fellows Alexander Bowring (Uni. Oxford, UK, advised by Prof. Thomas Nichols), Sofia Strubbia (Uni. of Nantes, advised by Dr. Alban Gaignard) and Aya Kababra (Lebanese Uni., Lebanon, advised by Dr. Mahmoud Hassan). I have supervised 5 research engineers: Aurélien Cornet (co-advised with Olivier Dameron, Alban Gaignard and Anne Siegel), Rémi Adon, Thomas Betton, Hermann Courteille, and Cyril Regan.

I am currently co-directing the PhD fellowship of Elodie Germani (with Prof. Elisa Fromont), the postdoctoral fellowship of Jérémy Lefort-Besnard (with Prof. Thomas Nichols), the M1 internship of Emma Redor (with Dr. Fanny Dégeilh) as well as the engineering work of Boris Clenet.



# Acknowledgments

I would like to thank the committee members for their time assessing my manuscript as well as for sharing their insights during the defense. I am incredibly thankful to the many people who contributed to this work. First of all, I would like to thank the students, PhD candidates, engineers and postdoctoral fellows I had the pleasure to work with. My thanks also go to the researchers I co-supervised along with. As I am writing this habilitation document, my thoughts are with Christian Barillot and with Olivier Commowick who left us prematurely and who are immensely missed in the Empenn team.

Throughout those 10 past years, I had the chance to contribute and get support from the Open Science community and this has been an important driver of my work. I would like to thank in particular the members of the INCF NIDASH group and of the OHBM Open Science Special Interest Group as well as the members of the IRISA / Inria Rennes gender equality committee. I would like to thank all the members of the research teams I had the chance to evolve in for welcoming me: VisAGeS, the Institute of Digital Healthcare, Nisox and Empenn. And finally I want to express my gratitude to my family and in particular Cyril, Thibault and Olivia and my parents and siblings for their continued support.

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Brain imaging . . . . .	5
1.2	On the meanings of reproducibility . . . . .	6
1.3	When reproducibility fails . . . . .	7
1.4	Overview of the contributions . . . . .	9
<b>2</b>	<b>Opening up neuroimaging data</b>	<b>11</b>
2.1	Community standardization efforts . . . . .	14
2.2	Sharing task-fMRI results . . . . .	18
2.3	Data sharing and privacy . . . . .	22
<b>3</b>	<b>On the vibration of effects in neuroimaging</b>	<b>25</b>
3.1	Same pipeline, different software: different results . . . . .	30
3.2	Same data, different experts: different results . . . . .	33
3.3	BIDS-Prov: a provenance framework for BIDS . . . . .	37
<b>4</b>	<b>Reusing neuroimaging data</b>	<b>39</b>
4.1	Validity of meta-analysis methods for fMRI . . . . .	42
4.2	fMRI mega-analyses with different preprocessing . . . . .	47
4.3	Leveraging large datasets for new analyses . . . . .	50
<b>5</b>	<b>Open Science</b>	<b>53</b>
<b>6</b>	<b>Conclusions and perspectives</b>	<b>59</b>
	<b>My publications</b>	<b>63</b>
	<b>Bibliography</b>	<b>83</b>
	<b>List of Acronyms</b>	<b>85</b>
	<b>Credits</b>	<b>87</b>



# Chapter 1

## Introduction

About 10 years ago, a series of publications pointed to the difficulty of reproducing published findings through a re-analysis (Button et al. 2013b; Open Science Collaboration 2015). Evidence of this phenomenon – later referred to as the *reproducibility crisis* – was identified in many scientific fields including (to name only a few): Cancer research (Begley and Ellis 2012), Psychology (Open Science Collaboration 2015) or Neuroscience (Button et al. 2013b). This lack of reproducibility of research results has been a wake-up call for scientific communities to reconsider the way research is done and to propose more robust approaches going forward.

My research is at the crossroads between reproducibility and brain imaging, a field known as *Neuroimaging reproducibility*. In the following sections, I provide a brief overview of the type of analyses I have been focusing on within brain imaging (see Section 1.1). Then, I present the research topic of reproducibility in more details with examples stemming from neuroimaging (see Section 1.2). Finally I present an overview of my contributions that are then described in details in the remainder of the manuscript (see Section 1.4).

### 1.1 Brain imaging

Brain imaging - also known as *neuroimaging* - provides the ability to observe the living brain. In my research, I have almost always worked with data from *Proton Magnetic Resonance Imaging (MRI)*. This imaging technique uses the magnetic properties of hydrogen and is best suited to image soft tissue (such as muscles, fat or blood vessels). In its simplest form (no injection) MRI is fully non invasive and is therefore a method of choice to measure signals inside the human body. In particular, MRI can be used to image the brain with high resolution and full coverage.

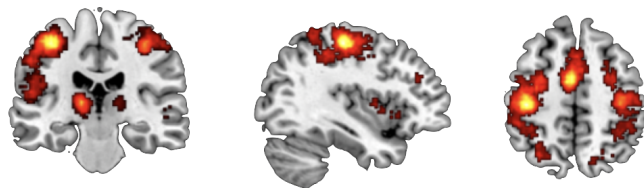


Figure 1: Activations for a finger-tapping experiment in task-fMRI. Activated areas (red-to-yellow clusters) are overlaid onto an anatomical image of a typical brain. Credits: Image generated by Neurosynth (Yarkoni et al. 2011) and displayed with mricron.

Within brain imaging with MRI, I focused on one specific sub-type of images, known as *task-based functional Magnetic Resonance Imaging (task-fMRI)* (also called BOLD for *Blood Oxygen Level Dependent fMRI*). Briefly, the goal of task-fMRI studies is to identify brain regions activated for a task of interest. As an illustration, Fig. 1 presents activations identified using task-fMRI during a finger-tapping experiment. Analysis of task-fMRI data proceeds by performing a statistic test in each point in the brain (known as a *voxel*) to determine whether there is a significant activation. Results of task-fMRI studies are therefore 3D images in which each voxel is identified as active or inactive (i.e. no evidence for activation).

Task-fMRI is a particularly interesting modality to study as a methodologist (like myself) because it is widely-used in the neuroimaging community. The principles for task-fMRI were introduced in the early 1990's and since then many studies have been using this imaging technique.

Neuroimaging is an interdisciplinary research field at the crossing of Cognitive neuroscience, Psychology, Physics, Statistics and Computer Science, and the community has looked at the problem of irreproducibility through different lenses.

## 1.2 On the meanings of reproducibility

The concept of *reproducibility* is core to the scientific process. Not only do scientists develop their research by building upon existing findings (from the scientific literature) but reproducing results (either one's own or those of fellow scientists) is an important part of the scientific process by which proposed results are vetted by the community.

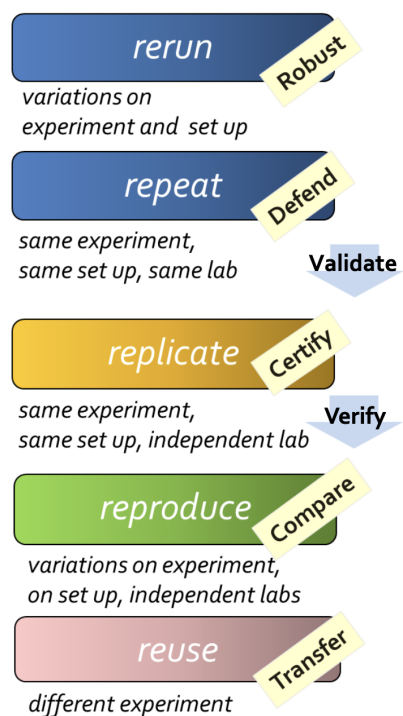


Figure 2: The many R's of reproducibility Credits: (Goble 2016)

In her keynote at the Alan Turing Institute Symposium on Reproducibility for Data-Intensive

Research in April 2016 (Goble 2016), Prof. Carole Goble disentangled the different type of reproducibility through 5 “R” verbs: Rerun, Repeat, Replicate, Reproduce and Reuse (see Fig. 2). Those span the full research cycle: from testing of the robustness of one’s results before publication (*Rerun*) to a fully independent reproduction (*Reproduce*) and then onto transfer (*Reuse*). Confusingly the definitions of *Replicate/Reproduce* are reversed in some communities (see (Barba 2018) for a more complete account of this issue).

### 1.3 When reproducibility fails

The way scientists understand and perceive the *Reproducibility crisis* can be very diverse across fields. Indeed a lack of reproducibility can be the consequence of very different issues.

On one hand, one might not be able to replicate (i.e. *same experiment, same set up, independent lab*) a published finding because the shared description of the methods was not detailed enough or because one of the tools used (typically software) is no longer available. This case, in which a scientist aims at reproducing the results of a published study under the very same conditions (covering *Repeat* and *Replicate* from above) is often referred to as *computational reproducibility*. Here, the goal is to obtain an exact (bit-wise) reproduction of existing results using the very same tool/methods and data. In practice, computational reproducibility is only feasible with access to the original code and data. The difficulty encountered to ensure long-term availability of such artefacts makes it extremely difficult to comply with computational reproducibility in the long run (see the “10 Year reproducibility challenge” (Perkel 2020) for a discussion on this topic).

At the other end of the spectrum, the end-goal of experimental sciences, such as neuroimaging, is to uncover findings that hold in a wide multitude of contexts (covering *Reproduce* from above). For example, findings should hold with different samples of participants, or different acquisition instruments (such as different MRI scanners across hospitals) or using a methodology that is similar but not strictly equal. The ability for a finding to hold in different contexts (and in particular in different samples) refers to the property of *generalizability*. Some authors have argued that the lack of generalizability was one of the root-causes of the reproducibility crisis (see (Yarkoni 2019)).

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Figure 3: An overview of reproducible research according to The Turing Way. Credits: CC-BY reproducible matrix ©The Turing Way Community

The Turing way – a collaborative open science handbook and community (Turing Way et al. 2019) – highlights four different components of reproducible science (see Fig. 3): Reproducible (same-data, same-analysis, i.e. equivalent to *computational reproducibility* above), Replicable (different-data, different-analysis), Robust (same-data, different analysis) and Generalizable

(different-data, different-analysis).

In the remainder of this manuscript we will use the term reproducibility in its most general form (similarly to Prof. Carole Goble above), i.e. a finding is reproducible if it can be computationally reproduced but also reproduced in different samples of the target population and using similar but different methods. In the following sections we discuss different causes that can induce irreproducibility in more details.

## **Lack of generalizability and fairness across the target population**

Irreproducible results can be the consequence of a lack of generalizability of the findings. This problem is particularly acute in brain imaging studies. Due to the high cost of running a brain imaging study, often only a small number of participants can be recruited and sometimes those participants come from the same location (e.g. in a close neighborhood around the MRI facility). If this set of participants is not representative of the target population then the results can become irreproducible when tested on new samples.

For example, in X-ray imaging, a recent study by Larrazabal and colleagues (Larrazabal et al. 2020) illustrated how a lack of representativity in the training data can lead to biased classifiers. They showed how a classifier built to detect lung opacity would be more likely to fail on women's data when trained only on data from men (and conversely). In recent years, awareness of these issues in the scientific community has grown and a number of initiatives have been proposed. For instance, The Psychological Science Accelerator (PSA, <https://psysciacc.org/>) focuses on running distributed studies in different parts of the globe to test the generalizability of psychological constructs. Beyond differences across genders, the question of the applicability of research findings to underrepresented groups is even more complex. Medical decision must provide equal outcome for all subgroups and the topic of measuring the fairness and building fair solutions in medical imaging is currently gaining strong interest (Ganz et al. 2021; Ricci Lara et al. 2022).

## **False positive findings**

Irreproducible findings can also be observed in the presence of a false positive, i.e. when the original finding turns out being artefactual. In theory, the methodology used should provide guarantees that the rate of false positive findings will remain low. For instance, in a classical statistical setting with a typical *alpha*-level of 0.05, we expect on average a result to be falsely identified as positive 5% of the time (in the absence of true underlying evidence). But in real life, this rate might be inflated due to questionable research practices or might not be informative enough due to low statistical power.

**Inflated false positive rates: Questionable Research Practices** Practitioners face a highly complex analytic landscape in which they have to choose from, in order to perform their study. An example of a typical analytical choice is the selection of participants to be excluded from the analysis. While this practice is important to ensure that all data included have a sufficient level of quality, if not defined a priori, it can bias the final results. For instance, sequentially running the analysis as participants are included can cause inflated false positive rate if the recruitment stopping criterion depends on the results (Simmons et al. 2011). This phenomenon known as

*analytical flexibility* or as the *researcher's degree of freedom* refers to the possibility to select amongst a collection of possible analytical choices. If more than one test is performed (i.e. when more than one analytical choice is explored) then, in theory, a correction for multiple comparisons should be applied in order to keep the guarantee provided by the  $\alpha$ -level across the family of tests. This is often impractical and effectively leads to a case of selection bias. While this selective reporting is in a few instances a case of malpractice (sometimes referred to as *p-hacking*) – in which different analytical choices are purposefully tested in order to obtain the desired results – the problem is most likely to arise from a mixture of practical reasons (e.g. brain imaging studies have a very complex setup and often the definition of the pipeline is incremental) and a lack of statistical training. This issue is known beyond neuroimaging (Simmons et al. 2011) and different approaches have been proposed as a solution. In particular, some authors have proposed to select the analytical pipeline a priori, for example, by pre-registering the analysis before the data is acquired (Chambers et al. 2015).

**Inflated false positive rates: Low statistical power** One of the first and more pregnant answers to the reproducibility crisis in the field of neuroimaging was to question the sample size of neuroimaging studies. The impact of low samples on reproducibility was clearly demonstrated by Butter and colleagues in a paper entitled *Power failure: why small sample size undermines the reliability of neuroscience* that demonstrated how small sample size and small effect size led to low power in neuroimaging and to small positive predictive values (Button et al. 2013a). Positive predictive value corresponds to the proportion of detected true findings that are actually true, i.e. with a low positive predictive value, a high proportion of the scientific literature is in fact incorrect. This might be counter-intuitive as the  $\alpha$ -level provides a guarantee on the number of false positives. But this guarantee is only for null data (i.e. the proportion of false positive findings when there is no effect to be found) and the literature is mainly made up of positive results as negative results are much more difficult to publish (a phenomenon known as the *file drawer*).

**Inflated false positive rates: other reasons** Finally, inflated false positive rates can also be observed when the methods used to analyze the data is inappropriate. A seminal paper was published in 2016 (Eklund et al. 2016a) which led to many discussions and debates and even to questioning the reliability of functional neuroimaging in the news. While some of the news coverage was somewhat of an exaggeration, the discovery of inflated false positive rate was real. In this paper, the authors demonstrated that a parameter of the clusterwise inference (a specific type of inference that is done by considering cluster sizes) if not properly set could lead to inflated false positive rates as big as 50%. While the conclusion on the range of valid values for this parameter were known well before that paper (Flandin and Friston 2019), the default value in one of the major software packages was not optimal. This effectively calls for more testing of neuroimaging software packages, possibly through benchmarks and the development of best practices.

## 1.4 Overview of the contributions

I have contributed to the field of reproducible neuroimaging with research in informatics, statistics and data science. I strive for the outputs of my research to be directly useful to the neuroimaging community. This can first be seen in my continuous efforts, throughout my career,



to implement my research into existing tools directly available to practitioners. But beyond, it has shaped the directions of my scientific activities both in terms of the questions I chose to tackle and in terms of the type of solutions I have proposed.

Within the field of neuroimaging reproducibility, I have focused on three main aspects that are covered in the three next chapters of this manuscript (see Fig. 4).

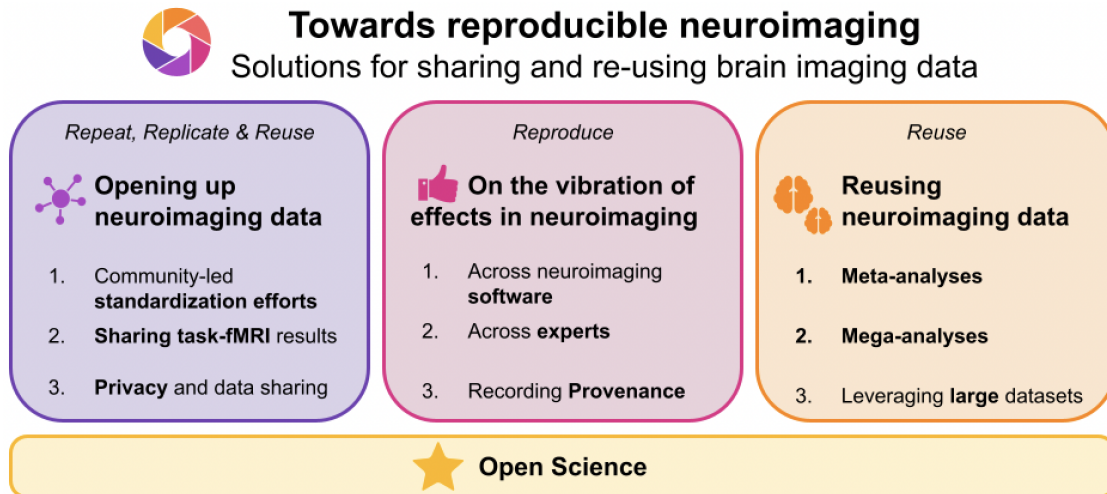


Figure 4: Overview of my research contributions

In my first research axis, I contributed to the impetus towards data sharing focusing on the research question: **How can we make neuroimaging data open?** Here, I focused on knowledge representation but also on infrastructure building and, more recently, on the legal aspects of sharing brain imaging datasets. This will be the focus of Chapter 2.

A key challenge to studying neuroimaging reproducibility is to understand the root-causes of irreproducibility. This is a first necessary step before we can formulate solutions. In my second research axis, I contributed to our common understanding of **What is the impact of different post-processing pipelines on neuroimaging results?** Here, I was part of the community effort that showed how the variability induced by different processing pipelines can lead to a vibration of effects in brain imaging. This will be the focus of Chapter 3.

In my third research axis, I focused on the reuse of neuroimaging datasets. As data sharing has taken a larger place in neuroimaging, there are increased opportunities to build new research on pre-existing data. But moving from the status-quo in which a dataset was created for each new research question opens up new challenges. I contributed to study: **What are the limitations to the reuse of derived neuroimaging datasets?** Here, I derived statistical approaches to combine pre-existing datasets in new studies or to inform new studies based on pre-existing data. This will be the focus of Chapter 4.

Finally, I am also an advocate for a more **Open Science** and have participated actively in international and national communities including: Brainhack, the International Neuroinformatics Coordinating Facility (INCF), the Organization for Human Brain Mapping (OHBM) Open Science Special Interest Group and the French Committee for Open Science (“Comité pour la science ouverte” (CoSO)). This led me to contribute to a number of efforts to make Open Science more widespread in neuroimaging and beyond. This will be the focus of Chapter 5.

# Chapter 2

## Opening up neuroimaging data

One of the first and main answers of the neuroimaging community to the reproducibility crisis was to push for data sharing. The hope was that by pooling data, neuroimagers would achieve larger sample sizes and therefore higher statistical power, leading to less false positive findings (see chapter 1, section 1.3).

### From sharing one study to building community resources

Data sharing in the neuroimaging community has been marked by the difficulties encountered by the fMRIDC (Van Horn and Gazzaniga 2013) – a pioneer effort in the early 2000’s that focused on sharing data accompanying neuroimaging publications. While agreements had been made with journals to make data sharing the rule, the fMRIDC faced a high-level of reluctance from the community and eventually had to be taken down. Community members specifically complained about the time-consuming process of submitting datasets to the fMRIDC but also more generally outlined cultural barriers such as their reluctance to share data in the fear of being scooped or proved wrong. It is only about 10 years later that data sharing in neuroimaging was relaunched and effectively adopted by a growing part of the community.

The first set of efforts, in the vein of the original fMRIDC, focused on sharing datasets that were linked to a research publication. The OpenfMRI (Poldrack et al. 2013a) (now OpenNeuro (Markiewicz et al. 2021)) was specifically dedicated to store and share task-fMRI data.

A bit later, researchers started to take this process more upstream and gathered in consortia to collect similar data and achieve larger sample sizes at the onset of a study. One of the best examples was the *1000 Functional Connectome Project*, led by Biswal, in which researchers across the globe joined forces to create a resting state fMRI dataset with an unprecedented size of 1000 participants. This collaborative effort proved the usability of resting state fMRI data (a specific type of fMRI in which the participant is asked to rest and researchers study correlations across the brain), which is now a flourishing sub-field of neuroimaging (Biswal et al. 2010) (see also B. Biswal’s OHBM 2023 keynote<sup>1</sup>). This success led to the creation of the *International Neuroimaging*

---

<sup>1</sup>[www.ohbmbrianmappingblog.com/1/post/2023/07/a-conversation-with-bharat-biswal-ohbm-2023-keynote-interview-pt-8.html](http://www.ohbmbrianmappingblog.com/1/post/2023/07/a-conversation-with-bharat-biswal-ohbm-2023-keynote-interview-pt-8.html)

*Data-sharing Initiative (INDI)*, a home for multiple consortia that have created and shared open datasets including: The Consortium for Reliability and Reproducibility (CoRR) (Zuo et al. 2014), ABIDE (Di Martino et al. 2014), NKI-Rockland Sample (Nooner et al. 2012) and more.

After this first era, in which scientists self-organized, funders started to get involved in data sharing in the hope that this could counter the reproducibility crisis. As an example, Human Connectome Project: Young Adult (HCP Young Adult) (Van Essen et al. 2013) (funded by NIH's Blueprint for Neuroscience Research) collected and shared data from over 1000 participants. More recent efforts include the UK Biobank (Miller et al. 2016), a prospective data bank in the UK that targets the imaging of 100 000 participants and the Adolescent Brain Cognitive Development (ABCD) (Casey et al. 2018).

Throughout the years, this has led to the creation of resources with an increasing number of subjects available. Those very large datasets that include a higher number of subjects typically cover a limited subset of brain functions. More recently, efforts have been put into filling the gap by acquiring datasets with a fewer number of participants but a very high number of experimental protocols, an approach known as *deep sampling*. Deep sampling datasets include the my connectome project (Poldrack et al. 2015), the courtois Neuromod project, the individual brain charting (Pinho et al. 2018), the CNeuromod project ([www.cneuromod.ca](http://www.cneuromod.ca)) or the Natural Scene Dataset (Allen et al. 2022).

## Data sharing as a way to improve reproducibility

By putting data in common, neuroimagers have achieved larger sample sizes and therefore higher statistical power leading to less false positive findings. Beyond false positive findings, data sharing is also helpful for other aspects of reproducibility. First, it is a stepping stone of computational reproducibility and second, it provides opportunities to test the generalizability of findings across different populations and experimental environments.

## My contributions

The past 10 years have seen the emergence of data sharing (and pooling) in many scientific communities beyond neuroimaging. This notably led to the proposal of best practices for data sharing and in particular to the Findable, Accessible, Interoperable and Reusable (FAIR) principles (Wilkinson et al. 2016). In brief the FAIR principles can be described as: the ability to search for and discover datasets (Findable), the possibility to download the data (Accessible) in a format that is commonly used (Interoperable) and which provides a level of details that is sufficient to enable reuse (Reusable). In practice, the emergence of FAIR has paved the way for the development of community standards and repositories.

## **Community-driven standardization efforts in brain imaging**

In neuroimaging, studies typically involve multiple participants that each have multiple images (anatomical, functional, diffusion, etc.). When the first data sharing effort arose, there was no standardized format or data structure to store information related to a study. Each tool or database had its own version of how to store information about groups of participants and their data.

In a first set of contributions (see Section 2.1), I took part in two community-driven standardization efforts to facilitate data sharing in neuroimaging.

### **Sharing results of task-fMRI studies**

While most of the discussion above has focused on the sharing of raw datasets, sharing derived datasets is also of growing interest. This not only opens up the possibility to perform meta-analysis quantitatively and test convergence of results from the literature but is also a more efficient approach to data reuse. In fact most of the large resources provide a preprocessed version of their data to avoid costly recomputations.

In a second set of contributions (see Section 2.2), I introduced NIDM-Results as a standard to share task-fMRI results. I also contributed to the development of the NeuroVault infrastructure dedicated to the sharing of those results.

### **Providing guidelines for researchers to share their data while complying with European regulations**

A lot of the data sharing efforts in neuroimaging have been driven by efforts stemming from the United States of America. Differences in regulations between Europe and the US – in particular the fact that, under the European General Data Protection Rule (GDPR), a brain image is not anonymous – require special attention so that European researchers can also share their data.

In a third contribution (see Section 2.3), I participated to the *Open Brain Consent*, a collaborative effort to develop template documents to enable data sharing in Europe while preserving the privacy of the participants.

## 2.1 Community standardization efforts

*This work was initiated as part of my postdoctoral fellowship at the University of Warwick and Oxford University, UK. During that period, I was a member of the NeuroImaging Data SHaring task force (NIDASH). NIDASH was an international collaboration, under the aegis of the International Neuroinformatics Coordinating Facility (INCF)<sup>2</sup>, which brought together researchers from 10+ laboratories in the US, Canada, UK and Germany. This working group led the development of the Neuroimaging Data Model (NIDM) standardization effort and was at the inception of the creation of the Brain Imaging Data Structure (BIDS). Since then, I have continuously contributed to the field of neuroimaging standard and in 2023, I joined the BIDS steering group.*

**Context** When the focus of the neuroimaging community turned to data sharing in the hope to achieve higher sample sizes (and hence higher statistical power), it created an acute need for common representations (or standards) to exchange data. Since then (and to date), brain images in MRI had typically been stored using one of the following two formats:

- **DICOM:** a comprehensive data format that includes descriptive metadata about the imaging acquisition and the participants. DICOM is used by MRI scanners in hospitals. A 3D brain image is made up of many DICOM images.
- **NifTI:** a lighter format that mainly includes the image and information about its position in space. NifTI is used by most post-processing software. A 3D brain image typically corresponds to 1 NifTI image.

DICOM has been ubiquitous in hospitals and is the format of choice used by MRI manufacturers. NifTI has gained a lot of traction in post-processing, especially in a research setting (which is our focus here).

But while NifTI perfectly addressed the question of sharing MRI images one-by-one, it was not developed to deal with multi-subject studies in which multiple images (e.g. anatomical, functional, diffusion) are associated with a participant and sets of participants are organized into groups (e.g. controls, patients). In addition, even though the NifTI format was widely used by post-processing neuroimaging tools, the DICOM to NifTI conversion included an information loss that could be problematic for post-processing (e.g. slice numbering was often lost al-

though it is required information to perform slice timing).

This status-quo effectively led many labs to specify their own data structures, which sometimes even varied across projects within a lab. This tremendously limited the ability to reuse datasets and to share datasets with other researchers.

**Contribution** Over the years, I contributed to two standardization efforts: the Neuroimaging Data Model (NIDM) and the Brain Imaging Data Structure (BIDS).

NIDM was developed under an international group - called NIDASH for Neuroimaging Data Sharing - of the International Neuroinformatics Coordinating Facility (INCF). NIDM is based on semantic web technologies and the vision was to develop a collection of specification documents as extensions to the W3C PROV standard that would be suited to the domain of human brain mapping (Keator et al. 2013). NIDM used provenance information as a mean to link components from different stages of the scientific research process, from dataset descriptors and computational workflows, to derived data and publication (see Fig. 5).

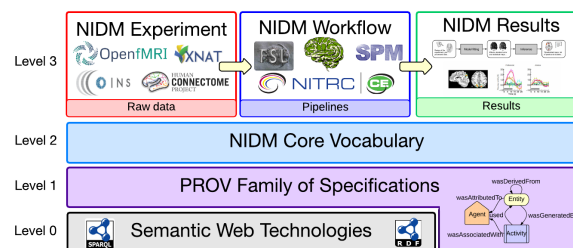


Figure 5: The Neuroimaging Data Model (NIDM) Layer Cake (inspired by the Semantic Web Layer Cake). Credits: Nolan Nichols.

<sup>2</sup><https://www.incf.org/>

As part of NIDM, we developed NIDM-Results to share results of task-fMRI studies (See section 2.2 for more on this effort).

BIDS uses a set of conventions for naming and organizing files in a tree as well as schemas for accompanying JSON files that are used to include additional metadata (Gorgolewski et al. 2016).

In its first iteration BIDS focused on modeling MRI datasets with multi-subjects and 3 types of imaging data: anatomical, functional and diffusion MRI. The main guiding principle to specify mandatory metadata was to include the minimal amount of information required to allow for downstream processing of the data. A schematic overview of the proposed BIDS organisation is presented in Fig. 6.

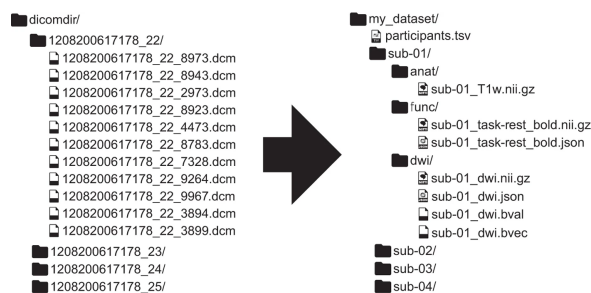


Figure 6: The Brain Imaging Data Structure (BIDS). Credits: CC-BY Fig. 1 from(Gorgolewski et al. 2016)

## Results

The NIDM-Results data model was the first model released by the NIDM group in 2016 (Maumet et al. 2016). We worked closely with the developers of the three main task-fMRI software packages and proposed reference implementations for two of them (see 2.2 for a full description). Beyond this, NIDM prefigured a number of efforts that are still ongoing, such as the DANDI archive - an online archive for publishing and sharing neurophysiology data - (dandiarchive.org).

BIDS has received strong support from the neuroimaging community. Throughout the years, many of the scientists who had initiated NIDM also contributed to BIDS. As an illustration, NIDM worked hand-in-hand with BIDS to create the NIDM-Terms ontology that includes all the terms and definitions from the BIDS vocabulary (Queder et al. 2023)). In another work, we focused on making both standards

converge on the same data format through the use of JSON-LD 1.1 which is both JSON - as in BIDS - and can store semantic web - as in NIDM - (Maumet et al. 2019). This work was then used for the BIDS-Prov specification (see 3.3 for a full description) The adoption of BIDS by the neuroimaging community has been very quick, not only to share data but also, importantly, to store and process data internally within labs. A recent estimate showed that more than 1 million BIDS datasets have been created <sup>3</sup>.

In 2018, BIDS started to extend beyond MRI data, first focusing on Magnetoencephalography (Niso et al. 2018). And then, reflexions started within BIDS to extend BIDS to derived datasets with BIDS-Derivatives.

This led to the formalization of a process to extend BIDS: the *BIDS extension proposal (BEP)*. A BEP is proposed and led by 1 or 2 community members who gather a small group of 5-10 scientists who collaborate on writing a first version of the proposed update. The draft is then open to contributions to the wider BIDS community. Once all comments have been addressed the BEP can be merged into the BIDS specification and effectively becomes part of the standard.

The BEP process was quickly adopted and BIDS was extended to electroencephalography (Pernet et al. 2019), intracranial electroencephalography (Holdgraf et al. 2019), positron emission tomography (Norgaard et al. 2022) as well as other MRI modalities including Arterial Spin Labeling (Clement et al. 2022). Twenty-five BEPs are currently under development (See [bids.neuroimaging.io/get\\_involved.html](https://bids.neuroimaging.io/get_involved.html) for a full list).

Many of the historical/main software packages in neuroimaging have updated their codebase to become BIDS-compliant, including: SPM and FieldTrip. There are also containerized applications (known as BIDS-apps (Gorgolewski et al. 2017)) based on those software, often developed by or with the software developers of the main packages.

In 2019, BIDS evolved to have a governance structure that includes an (elected) steering group that foresee all activities regarding BIDS. There is also a group of maintainers from the community.

<sup>3</sup>[twitter.net/chrisgorgo/status/1668701996000370689](https://twitter.net/chrisgorgo/status/1668701996000370689)



### Take-Home Message

- The neuroimaging community has been rich in collaborative standardization efforts.
- BIDS has been well-adopted by the brain imaging community and quickly extended from MRI to other imaging types.
- NIDM prefigured a number of applications as well as evolutions of BIDS (cf. NIDM-Results in 2.2 and BIDS-Prov in 3.3).
- BIDS now has its own governance structure.

### Publications

Gorgolewski, Krzysztof J., Tibor Auer, Vince D. Calhoun, Cameron R. Craddock, Samir Das, Eugene P. Duff, Guillaume Flandin, Satrajit S. Ghosh, Tristan Glatard, Yaroslav O. Halchenko, Daniel A. Handwerker, Michael Hanke, David Keator, Xiangrui Li, Zachary Michael, **Camille Maumet**, Nolan B. Nichols, Thomas E. Nichols, John Pellman, Jean-Baptiste Poline, Ariel Rokem, Gunnar Schaefer, Vanessa Sochat, William Triplett, Jessica A. Turner, Gaël Varoquaux, and Russell A. Poldrack (2016). “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments”. In: *Scientific Data* 3, p. 160044. DOI: 10.1038/sdata.2016.44.

**Maumet, Camille**, Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, Tristan Glatard, Krzysztof J Gorgolewski, Karl G Helmer, Mark Jenkinson, David B Keator, B. Nolan Nichols, Jean-Baptiste Poline, Richard Reynolds, Vanessa Sochat, Jessica Turner, and Thomas E. Nichols (2016). “Sharing brain mapping statistical results with the neuroimaging data model”. In: *Scientific Data* 3. DOI: 10.1038/sdata.2016.102.

Poldrack, Russell A., Christopher J. Markiewicz, Stefan Appelhoff, Yoni K. Ashar, Tibor Auer, Sylvain Baillet, Shashank Bansal, Leandro Beltrachini, Christian G. Benar, Giacomo Bertazzoli, Suyash Bhogawar, Ross W. Blair, Marta Bortoletto, Mathieu Boudreau, Teon L. Brooks, Vince D. Calhoun, Filippo Maria Castelli, Patricia Clement, Alexander L Cohen, Julien Cohen-Adad, Sasha D’Ambrosio, Gilles de Hollander, María de la iglesia-Vayá, Alejandro de la Vega, Arnaud Delorme, Orrin Devinsky, Dejan Draschkow, Eugene Paul Duff, Elizabeth DuPre, Eric Earl, Oscar Esteban, Franklin W. Feingold, Guillaume Flandin, anthony galassi anthony, Giuseppe Gallitto, Melanie Ganz, Rémi Gau, James Gholam, Satrajit S. Ghosh, Alessio Giacomel, Ashley G Gillman, Pdraig Gleeson, Alexandre Gramfort, Samuel Guay, Giacomo Guidali, Yaroslav O. Halchenko, Daniel A. Handwerker, Nell Hardcastle, Peer Herholz, Dora Hermes, Christopher J. Honey, Robert B. Innis, Horea-Ioan Ioanas, Andrew Jahn, Agah Karakuzu, David B. Keator, Gregory Kiar, Balint Kincses, Angela R. Laird, Jonathan C. Lau, Alberto Lazari, Jon Haitz Legarreta, Adam Li, Xiangrui Li, Bradley C. Love, Hanzhang Lu, **Camille Maumet**, Giacomo Mazzamuto, Steven L. Meisler, Mark Mikkelsen, Henk Mutsaerts, Thomas E. Nichols, Aki Nikolaidis, Gustav Nilsonne, Guiomar Niso, Martin Norgaard, Thomas W Okell, Robert Oostenveld, Eduard Ort, Patrick J. Park, Mateusz Pawlik, Cyril R. Pernet, Franco Pestilli, Jan Petr, Christophe Phillips, Jean-Baptiste Poline, Luca Pollonini, Pradeep Reddy Raamana, Petra Ritter, Gaia Rizzo, Kay A.

Robbins, Alexander P. Rockhill, Christine Rogers, Ariel Rokem, Chris Rorden, Alexandre Routier, Jose Manuel Saborit-Torres, Taylor Salo, Michael Schirner, Robert E. Smith, Tamas Spisak, Julia Sprenger, Nicole C. Swann, Martin Szinte, Sylvain Takerkart, Bertrand Thirion, Adam G. Thomas, Sajjad Torabian, Gael Varoquaux, Bradley Voytek, Julius Welzel, Martin Wilson, Tal Yarkoni, and Krzysztof J. Gorgolewski (2023). *The Past, Present, and Future of the Brain Imaging Data Structure (BIDS)*. Publisher: arXiv Version Number: 1. DOI: 10.48550/ARXIV.2309.05768. URL: <https://arxiv.org/abs/2309.05768>.



## 2.2 Sharing task-fMRI results

I spearheaded the development of *NIDM-Results* during my postdoctoral fellowship as part of the *NIDASH* task force (see section 2.1 for more details on *NIDASH*). Ruth Pauli and Alexander Bowring jointly led the data paper while they were respectively postgraduate and PhD students co-supervised between Prof. Thomas Nichols and myself at the Uni. of Warwick (UK). To a smaller extent, I was also involved in the development of the *NeuroVault* database, a collaborative effort led by Krzysztof Gorgolewski from the Max Planck Institute for Human Cognitive and Brain Sciences (Germany) and then Stanford Uni. (USA).

**Context** Since its discovery in the early 1990's, task-fMRI has been widely used leading to rich scientific production. As an illustration, a search of "task-fMRI" on PubMed reveals nearly 450 000 publications<sup>4</sup> (Note: we limited the time-frame up to 2016, i.e. the year in which our contributions were published). Yet task-fMRI studies - as many neuroimaging studies - have long suffered from low sample sizes which dampened the reliability of single experiments. This made the task-fMRI literature a perfect candidate for meta-analyses in order to extract convergent patterns across studies.

However, in a typical task-fMRI study, only a tiny fraction of the data and metadata produced was finally conveyed to the community (Poline et al. 2012). Indeed, task-fMRI studies were reported with 2 main components (see Fig. 7): 1- static figures showing the location of the activations 2- a table listing the coordinates and associated statistic values of the main local maxima (also known as *peaks*) in each cluster.

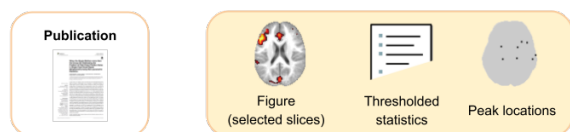


Figure 7: Typical reporting of an task-fMRI study: 1. graphical representation, 2. Table with statistical values (max 3 per cluster) and peak locations

This limited reporting effectively restricted the possibilities to perform meta-analyses and dedicated methodologies (known as *coordinate-based meta-analyses*) had to be developed. Those coordinate-based meta-analyses were known to be sub-optimal compared to image-based meta-analyses in which a test is done at each voxel (see Chapter 4 section 4.1 for more on image-based meta-analyses). In fact,

good practices from Statistics would require the results of all statistic test to be reported regardless of whether or not they reached significance.

Here, we proposed practical solutions to make it easier for neuroimagers to provide a more complete reporting in support of their task-fMRI studies.

**Contribution** We proposed three main contributions. First, we introduced *NIDM-Results*, a specification providing a machine-readable description of neuroimaging statistical results along with key image data summarizing the experiment (Maumet et al. 2016). Second, we shared a dataset and its accompanying data paper (Pauli et al. 2016) with processed task-fMRI results analyzed with AFNI (Cox 1996), FSL (Jenkinson et al. 2012) and SPM (Penny et al. 2011), three of the most-widely used task-fMRI software. Third, we introduced *NeuroVault*: a web based repository that allows researchers to store, share, visualize, and decode statistical maps of the human brain (Gorgolewski et al. 2015a).

**A specification for task-fMRI results** The goal of *NIDM-Results* was to provide a unified representation of mass-univariate analyses that would make it possible to perform both coordinate-based and image-based meta-analyses. A mass-univariate analysis proceeds by performing a family of statistic tests in parallel. This type of analysis is typically employed in task-fMRI analyses in which a test is performed independently at each location (or voxel) in the brain.

Within *NIDASH*, we took special care to build a harmonized representation across neuroimaging software. An example of such endeavour is the clarification of the different reference spaces used. While

<sup>4</sup><https://pubmed.ncbi.nlm.nih.gov/?term=fmri&filter=years.1990-2016>

most studies only mentioned standardization into the "MNI space" - in reference to the *Montreal Neurological Institute* - this template was in fact available in various versions and different neuroimaging packages had made different choices. Table 1 provides


a summary of the different MNI spaces (see the full discussions on Github <sup>5</sup> for more context). These descriptions were then used as the basis of the *Coordinate System* in BIDS.

Name	Short description	Used by
Colin27	"stereotaxic average of 27 T1-weighted MRI scans of the same individual"	SPM96
IcbmMni152 Linear	"average of 152 T1-weighted MRI scans, linearly transformed to Talairach space".	SPM99 to SPM8
IcbmMni152 NonLinear2009c Asymmetric	"average of 152 T1-weighted MRI scans, non-linearly transformed to MNI152 linear space".	DARTEL toolbox in SPM12b
IcbmMni152 NonLinear 6thGeneration	Reference space defined by the "average of 152 T1-weighted MRI scans, linearly and non-linearly (6 iterations) transformed to form a symmetric model in Talairach space"	FSL
Ixi549	average of the "549 [...] subjects from the IXI dataset" linearly transformed to ICBM MNI 452	SPM12b

Table 1: Expliciting the MNI spaces

**A dataset to explore the task-fMRI result space** We generated a set of task-fMRI result data with AFNI, FSL and SPM. This was useful both to better understand the task-fMRI result space as we were developing NIDM-Results and to test and build proof of concepts of NIDM-Results exporters. This dataset eventually became the *fmri comparison software dataset* and was realased on NITRC (Kennedy et al. 2016) and shared with the community.

**A repository for task-fMRI results** Finally, the goal of the NeuroVault repository was to create a "home" for brain statistic maps in order to make full statistic images available in support of task-fMRI publications. Those images can be uploaded either directly or using NIDM-Results.

 **Results** The NIDM-Results standard was defined by a W3C-style specification, pub-

licly available at <http://nidm.nidash.org/specs/nidm-results.html> and by an ontology (owl) file available at <http://bioportal.bioontology.org/ontologies/NIDM-RESULTS>. It is comprised of a controlled vocabulary, as well as instructions of how to use PROV to represent mass-univariate neuroimaging results.

The model provides terms to describe key elements of neuroimaging methods using a common framework across neuroimaging software packages. For example, as illustrated in Fig. 8, error models are described in terms of assumed variance (homoscedastic, heteroscedastic), assumed covariance structure (independent, spatially correlated, etc.) and how these structures vary in space (defined independently at each voxel, globally throughout the brain or spatially regularized).

A *NIDM-Results pack* is a zip archive that includes:

<sup>5</sup><https://github.com/incf-nidash/nidm-specs/issues/52>

a NIDM graph (containing all the metadata), a contrast map, a standard error map, a statistic map and a representation of the design matrix used in the statistical test. The NIDM graph includes information about the inferences, for instance the threshold used (and the type of threshold) as well as the specific template that was used for spatial registration into a template space.

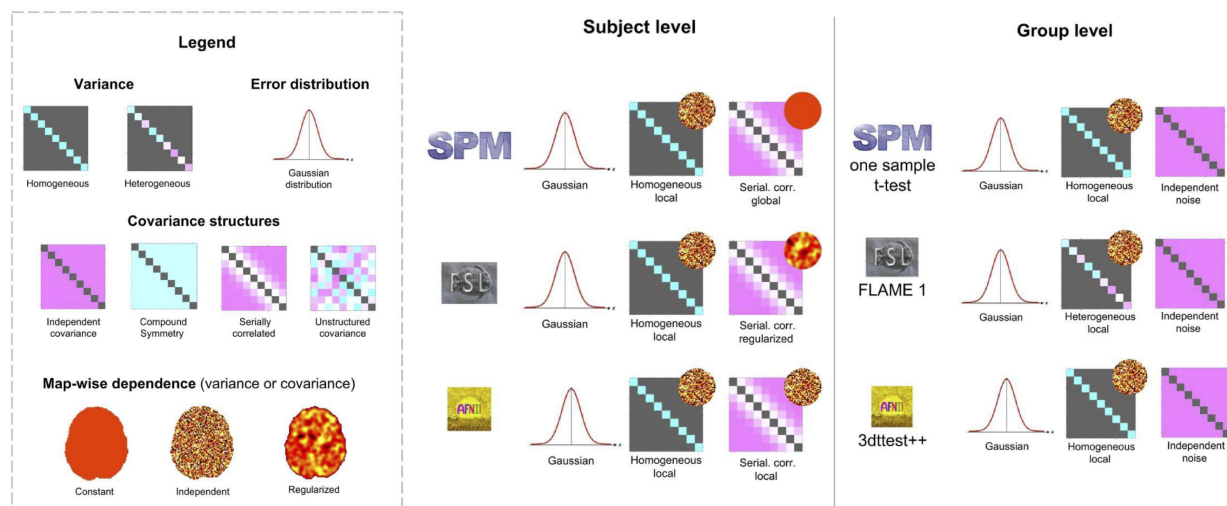


Figure 8: The different noise models in task-fMRI as specified in NIDM-Results

We focused on metadata that could be automatically extracted but also included pieces of information that were deemed crucial for meta-analysis. This was the case for the number of subjects (an information that is not automatically available by default).

Tools are available to export NIDM-Result graphs and associated files from the widely used SPM and FSL software packages, and the NeuroVault repository can import NIDM-Results archives.

A set of results of mass-univariate task-fMRI analyses was shared as the *task-fMRI comparison software library* using the most common software packages: AFNI, FSL, and SPM [which between them cover 80% of published task-fMRI analyses (Carp 2012)], utilizing publicly available data from OpenfMRI (Poldrack et al. 2013b). The analyses (‘variants’) proposed cover

the most common options available in each software package at each analysis stage, from different Hemodynamic Response Function (HRF) to varying group-level tests. The variants are arranged so that readers can compare the closest equivalent across software packages.

NeuroVault has been well adopted by the neuroimaging community (in particular for task-fMRI statistic maps). As of July 5th, 2023 the NeuroVault repository contains 470 000+ images<sup>6</sup> grouped in about 10 000 collections<sup>7</sup>. By providing a very large and diverse dataset of task-fMRI statistic maps, the NeuroVault database has opened up new opportunities for research. In particular it has been used in multiple scientific papers to train machine learning models (see for example (Menuet et al. 2022)).

<sup>6</sup><https://neurovault.org/api/images/>

<sup>7</sup><https://neurovault.org/api/collections/>

### Take-Home Message

- NIDM-Results provides a software-agnostic representation of task-fMRI results.
- We shared a dataset with 31 variants of the results of a mass univariate task-fMRI analysis using 3 of the most widely-used task-fMRI software packages: AFNI, FSL, and SPM.
- NeuroVault is an online database for task-fMRI statistic maps with 470 000+ of images.

### Publications

Gorgolewski, Krzysztof J., Gael Varoquaux, Gabriel Rivera, Yannick Schwarz, Satrajit S. Ghosh, **Camille Maumet**, Vanessa V. Sochat, Thomas E. Nichols, Russell A. Poldrack, Jean-Baptiste Poline, Tal Yarkoni, and Daniel S. Margulies (2015). “NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain”. In: *Frontiers in Neuroinformatics* 9.8. DOI: 10.3389/fninf.2015.00008.

**Maumet, Camille**, Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, Tristan Glatard, Krzysztof J Gorgolewski, Karl G Helmer, Mark Jenkinson, David B Keator, B. Nolan Nichols, Jean-Baptiste Poline, Richard Reynolds, Vanessa Sochat, Jessica Turner, and Thomas E. Nichols (2016). “Sharing brain mapping statistical results with the neuroimaging data model”. In: *Scientific Data* 3. DOI: 10.1038/sdata.2016.102.


Pauli, Ruth, Alexander Bowring, Richard Reynolds, Gang Chen, Thomas E. Nichols, and **Camille Maumet** (2016). “Exploring fMRI Results Space: 31 Variants of an fMRI Analysis in AFNI, FSL, and SPM”. In: *Frontiers in Neuroinformatics* 10. DOI: 10.3389/fninf.2016.00024.

### Talks *Follow the link to access the slides on HAL*

“Tools and standards to make neuroimaging derived data reusable” (2018). In: *Neuroinformatics, Montreal (Canada)*. URL: <https://inria.hal.science/inserm-01886089>.

“NIDM-Results: a model to share brain mapping statistical results” (2016). In: *CBU Open Science Workshop, University of Cambridge (UK)*. URL: <https://inria.hal.science/inserm-01887752>.

“Neuroinformatics techniques for provenance & data sharing” (2014). In: *GlaxoSmithKline-Neurophysics Workshop on Skeptical Neuroimaging, London (UK)*. URL: <https://inria.hal.science/inserm-01887730>.

 **Open artefacts** The task-fMRI Results Comparison Library is available on NITRC at <https://www.nitrc.org/projects/frcl>.

## 2.3 Data sharing and privacy

*This work was done as part of an international collaboration with 35+ researchers within the working group Multi-site data integration of the European COST Action Glimr that was co-led by Dr. Cyril Pernet from Uni. of Edinburgh (UK) and then Dr. Stephan Heunis from Research Center Jülich (Germany) and myself. The Open Brain Consent project pre-dated this collaboration and was driven by Dr. Cyril Pernet, Dr. Yaroslav Halchenko from Dartmouth College (USA), Dr. Peer Herholz from McGill Uni. (Canada) and Dr. Stephan Heunis. They gathered multiple teams of researchers throughout Europe to propose translations that would be compliant with local regulations. With Dr. Elise Bannier and Dr. Anne Hespel, both from Rennes hospital, we contributed the French translation*



### Context

A lot of the data sharing initiatives that have spread in the neuroimaging community come from the United States of America (USA). This is for instance the case of many platforms that are used for neuroimaging data sharing such as Openneuro (Markiewicz et al. 2021) or Neurovault (Gorgolewski et al. 2015a), but also of widely-used open datasets such as The Human Connectome Project (Van Essen et al. 2013). While these efforts have been instrumental in triggering neuroimaging data sharing, the fact that they are centralized in a single country (the USA) is a weakness. First, because there is growing evidence of the risks of creating biased models when training is done on datasets that are too homogeneous (see for instance (Larrazabal et al. 2020) for an illustration of gender bias in x-ray imaging). Second, because USA-based infrastructure might not be accessible to all scientists due to local regulations. This is for instance the case of European researchers.

A number of landmark efforts have emerged from Europe such as the UK Biobank – A biobank, including neuroimaging data that targets 100 000 participants for the imaging part (Miller et al. 2016). Yet, there is limited support and options available to the individual researchers in Europe who would like to share their neuroimaging data. The specificity of European regulation makes it impossible for EU-based researchers to use USA-based infrastructures to share brain data (see for example OpenNeuro’s note on GDPR in the FAQ<sup>8</sup>).

The GDPR requires stricter controls when sharing

individual human data. More specifically, there is a crucial difference between what is defined as anonymous data in the USA versus in Europe. The USA definition focuses on primary identifiable information (i.e. name, date of birth, address and face) while the European definition is much broader alluding to correlation, individualization and inference. So what does this mean for brain imaging data?

The property of individualization by itself is particularly stringent. The CNIL – i.e. the French Data Protection Agency – defines individualization as the possibility to isolate an individual in the dataset<sup>9</sup>. This means that as soon as a dataset is a unique signature of a participant it cannot be considered as anonymous under European laws. Beyond facial features that can be used to directly identify an individual, multiple publications have shown that brain images hold unique properties for each participant, a characteristic known as fingerprinting (see for instance in functional connectome (Finn et al. 2015)). European laws have been built with stronger guarantees for the privacy of the participants but they do enable a lot of possibilities for researchers. The issue is not so much the law itself but its practical application, a problem that is compounded by the various interpretations across countries and even across institutions.


Here, we proposed a simple approach to enable GDPR-compliant data sharing. In practice, neuroimaging data can be shared provided that 1/ participants are informed and their consent is collected and 2/ researchers accessing the data agree to com-

<sup>8</sup><https://docs.openneuro.org/faq>

<sup>9</sup><https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>




ply with a number of terms (e.g. no attempt to de-anonymize data).

 **Contribution** The goal of this project called *The Open Brain Consent GDPR-edition* was to share two template files. First, a template consent form that includes a paragraph explaining how data will be reused so that participants can elect to agree or not. Second, a template data user agreement to be signed by researchers who would like to reuse the data.

The first version was drafted in English and finalized during a workshop of the European COST Action Glimr in November 2019 at the COST Association in Brussels. This event gathered researchers from various fields in neuroimaging (cognitive neuroscience, clinical studies, neuroinformatics and data management) as well as an expert in law. As a re-

sult of this meeting the English version was shared with the Glimr network (and beyond) to create translations in as many European languages as possible. Special care had to be taken in order to make the translations compatible with local regulations. To this aim, each translation team involved a local data protection officer. With Dr. Elise Bannier and Dr. Anne Hespel - who is Data Protection Officer for Rennes hospital - we were in charge of the French translation.

 **Results** The Open Brain Consent initiative led to the creation of two template documents that are openly shared at <https://open-brain-consent.readthedocs.io> (Open Brain Consent working group 2021). As of August 29, 2023, both documents are available in 13 languages. In Rennes, the Open Brain Consent GDPR consent form is at the Neuroinfo<sup>10</sup> platform.


### Take-Home Message

- Open Brain Consent provides templates of GDPR-compliant consent form and data user agreement for neuroimaging data sharing.
- Those documents are available in 13 European languages.

### Publication

Open Brain Consent working group (2021). “The Open Brain Consent : Informing research participants and obtaining consent to share brain imaging data”. In: *Human Brain Mapping* 42.7, pp. 1945–1951. DOI: 10.1002/hbm.25351.

Members of the Open Brain Consent working group (in alphabetic order): *Elise Bannier, Gareth Barker, Valentina Borghesani, Nils Broeckx, Patricia Clement, Kyrre Emblem, Satrajit Ghosh, Enrico Glerean, Krzysztof Gorgolewski, Marko Havu, Yaroslav Halchenko, Peer Herholz, Anne Hespel, Stephan Heunis, Yue Hu, Chuanpeng Hu, Dorien Huijser, Vayá María Iglesia, Radim Jancalek, Vasileios Katsaros, Marie-Luise Kieseler, **Camille Maumet**, Clara Moreau, Henk-jan Mutsaerts, Robert Oostenveld, Esin Ozturk-isik, Nicolas Pascual Leone Espinosa, John Pellman, Cyril Pernet, Francesca Benedetta Pizzini, Amira Šerifović Trbalić, Paule-joanne Toussaint, Matteo Visconti Di Oleggio Castello, Fengjuan Wang, Cheng Wang, Hua Zhu.*

 **Talks** Follow the link to access the slides on HAL

<sup>10</sup><https://neuroinfo.org>

“Panel discussion ”Information, consent to data processing and the use of AI” (2023). In: *Colloquium: Data and artificial intelligence in healthcare, Faculty of Law and Political Science, Rennes (France)*. URL: <https://inria.hal.science/hal-04323409>.

“EU Cost Action GliMR - European multi-site data integration & large dataset creation for glioma diagnostics” (2021). In: *EORTC imaging group meeting, Online*. URL: <https://inria.hal.science/inserm-03193804>.

**Related work** In a follow-up work of the European COST Action Glimr, spearheaded by Jérémy Lefort-Besnard as part of his postdoctoral fellowship, we are currently reviewing infrastructure available for neuroimaging data sharing in Europe.

# Chapter 3

## On the vibration of effects in neuroimaging

In their daily practice, researchers face a complex analysis landscape in which they have to choose between different tools, different algorithms, and different parameters, available in different environments in order to analyze their data. In the past few years, increasing evidence has shown that the exact choice of analysis steps in the post-processing of the data (often referred to as a "workflow" or "pipeline") can have a non-negligible impact on research findings, effectively giving rise to a *vibration of effects* (i.e. a multiplicity of possible results for a given study).

### Neuroimaging pipelines

The issue of the vibration of effects is compounded when the number of steps included in the analysis is big. This is the case for neuroimaging studies in which the analytical pipeline is deep, typically made up of 8-15 steps from pre-processing to the final statistical analysis.

An example for task-fMRI pre-processing is provided in Fig. 9. In brief, both an anatomical and a functional image are needed. The two images can initially be processed in parallel. On one hand, the anatomical image is treated to remove heterogeneity in intensities (that are observed due to inhomogeneity of the magnetic field) then, if multiple anatomical images are available, they are merged, and finally the skull is removed from the image to keep only brain tissues. The anatomical image is then registered into a standardized space (so that later conclusions can be drawn at the group level), segmented into grey and white matter and then the surface of the cortex is extracted. On the other hand, the functional image is segmented to get a brain mask, then realigned in space (to correct for any motion of the participant) and in time (to compensate for the different acquisition time of different slices in the brain). Finally the anatomical and functional images are aligned to each other and the functional images are smoothed with a Gaussian kernel.

After pre-processing, the task-fMRI image enters a multi-level statistical analysis. At the first-level, subject-level statistics modelling and estimation are computed. At the second level, group statistics modelling and estimation are computed to infer results at the level of a population. Finally there is a phase of thresholding using a correction for multiple comparisons.

At each level (subject, group), the statistics are computed using a General Linear Model (GLM)



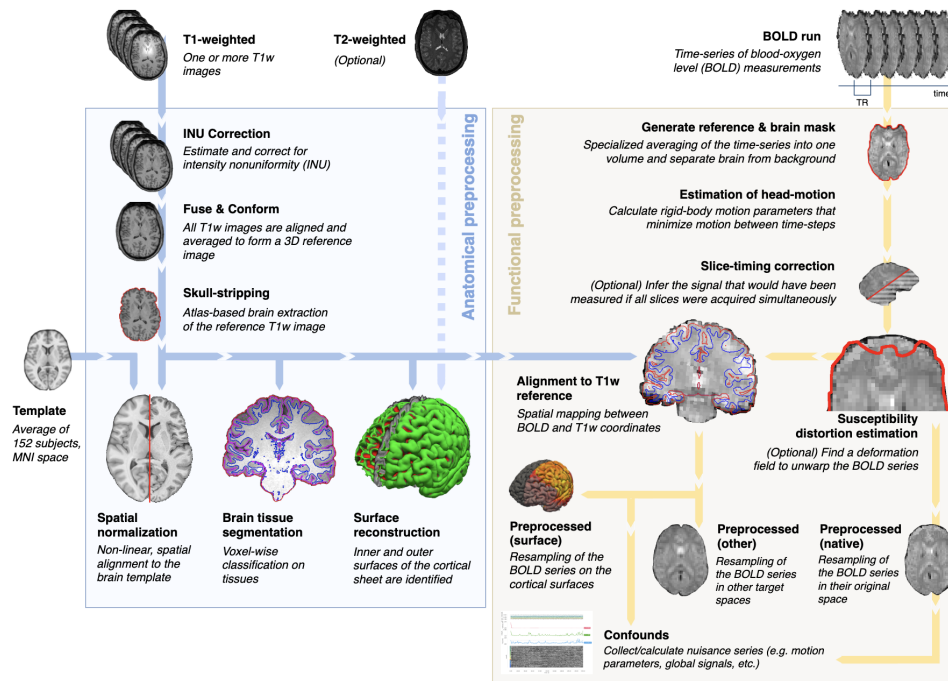


Figure 9: An example of task-fMRI preprocessing pipeline: fMRIPrep. Credits: CC-BY Fig. 1 from (Esteban et al. 2019)

at each voxel in the brain, an approach known as *mass-univariate*. Using the GLM, the task-fMRI signal is modelled as a linear combination of a set of pre-defined regressors representing the experimentally controlled factors as well as possible confounds.

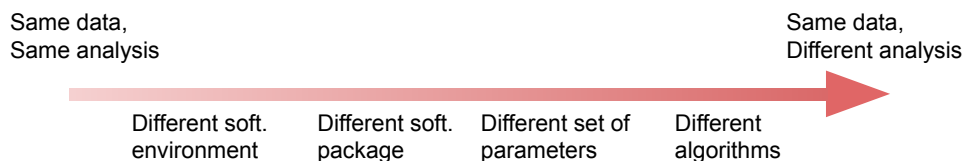
The output of the overall process is a 3D thresholded brain map that indicates regions of significant activation. In addition, the (unthresholded) statistic map includes the output values for all tests (one per brain location) regardless of whether they outlined a significant activation.

## How analytical variability impacts reproducibility

The variability observed in the results of a study when varying the analytical pipeline is sometimes referred to as *analytical variability*. Analytical variability can be induced by different levels of variations in the analytical pipeline including: different software environments, different software packages, different sets of parameters, different algorithms for each step in the pipeline. In addition, scientists may vary the order of some of the steps, exclude and/or include additional steps. While there are good practices to guide practitioners in those choices, there typically remains a large space of possible pipelines (Carp 2012).

In practice, analytical variability can arise due to a variety of reasons: from unwanted variability that is the consequence of a lack of robustness of the study, to variability of interest that informs researchers on their scientific question (see Fig. 10).

### A. Different components of analytical reproducibility



### B. What is analytical reproducibility effectively investigating?



Figure 10: Different components of analytical variability and how this variability can be explained as a continuum between unwanted variability related to a lack of robustness of the study up to variability of interest that informs scientist on their research question.

## A lack of robustness of the tools

Reaching different scientific conclusions when varying the operating system (while running the very same analysis on the same data) is likely to be an indication of the lack of robustness of the results obtained. This was well-illustrated in a study by Glatard and colleague (Glatard et al. 2015) who looked at neuroimaging reproducibility across different operating systems. The authors observed that results varied depending on whether they were obtained using a CentOS or a Fedora Linux operating system. After further investigation, they identified that differences in system libraries handling floating point values were at the root of the divergent results and that errors accumulated along the pipeline. While the researchers could relate the observed differences to the use of different low-level libraries, in practice the fact that using a different operating system leads to different scientific conclusions points at a lack of robustness of the original results.

As there is often no “ground truth” that could be used to act as a common benchmark, testing biomedical software can be especially challenging (Lundgren and Kanewala 2016), and neuroimaging is no exception. This issue is known as the oracle problem (Barr et al. 2015) in software engineering. As a way to test the robustness of pipelines, some authors analyzed how generating small variations in the input data induced changes in the results (Kiar et al. 2020), an approach sometimes referred to as *sensitivity analysis*. If a small perturbation in the input data induces a high level of variations in the output, then the method is not robust.

### **On the trap of finding the “best” pipeline...**

At the other end of the spectrum, researchers may reach different conclusions when using different algorithms or different statistical models, effectively implying different assumptions on their problem. Using a number of alternative approaches to test the robustness of the results or to exclude pipelines that are not suited is typical in fields in which the analysis pipeline can be characterized by many levels of variations. In principle, the goal is to exclude pipelines that are not (or less) suited to answer the problem in question.

Each method comes with a set of assumptions. For example, a registration algorithm might only be suitable for individual’s brains that are anatomically close to a standardized adult brain. Sometimes methods will be robust to assumption violations - in the previous example, the same registration algorithm might also provide good results when applied on infant brains even though it was not intended for this usage. Other times, assumption violations will lead to the failure of the method, which can manifest as inflated false positive rates (invalidity) (Eklund et al. 2018). When processing real data, researchers strive to find a method for which the data does not break assumptions, or a method that is robust to assumption violations (Mumford and Nichols 2009b) – a subtle balance to be found. When it comes to scientific pipelines, this problem is exasperated, as each individual step is characterized by its own assumptions and robustness. This effectively creates a pyramid of assumptions that can be very difficult to assess.

But, by analyzing their data multiple times, practitioners can fall into the issue of *p-hacking*. This issue is known beyond neuroimaging (Simmons et al. 2011) and different approaches have been proposed as a solution. Some authors have proposed to select the analytical pipeline a priori, for example, by pre-registering the analysis before data are acquired (Chambers et al. 2015). Other authors have proposed to combine results from multiple pipelines using specification curves from instance (Bloom et al. 2022; Burns et al. 2019; Simonsohn et al. 2020); or multiverse analysis (Steegen et al. 2016). Finally other authors, prefer to focus their efforts on reducing the analytical space by proposing a single best practice pipeline. This is the approach proposed by the fMRIPrep software (Esteban et al. 2019)). Other studies have focused on identifying optimum pipelines with respect to a predefined criterion (e.g. predicting brain age (Dafflon et al. 2022)) or to a specific metric (e.g. reproducibility (Strother et al. 2004)).

In conclusion, understanding why analytical variability arises is crucial in order to propose solutions to these sources of irreproducibility. This is challenging in the context of methodological variations because choosing a pipeline is a key part of the scientific process.

## **My contributions**

### **Variability across software packages**

While there is typically no ground truth that would make it possible to easily compare the performance of two competing pipelines, one approach to identify possible discrepancies is to run comparable pipelines and to check whether the results are consistent. Often the same conceptual steps are available in multiple implementations in different software packages. This is similar to

the approach used by *N-version testing* in software engineering (see (Barr et al. 2015) for a review on software testing in the absence of ground truth). N-version tests run by comparing the outputs of different implementations of the same piece of software. A lack of consistency in the results indicates a failure of (at least) one of the implementations.

In a first set of contributions (see Section 3.1), we tested the output of three software packages (AFNI, FSL and SPM) – selected for their large coverage of the task-fMRI literature – on three neuroimaging studies. We proceeded by reproducing 3 publications and by generating carefully-aligned pipelines in the alternative two research software.

### **Variability across experts**

In a second set of contributions (see Section 3.2), as part of a large collaborative effort, we used a similar approach to test whether the post-processing choices made by different experts can lead to different conclusions (using the same data). This approach is also known as a *Many analysts, one dataset* study.

### **Recording neuroimaging provenance**

In a third contribution (see Section 3.3), in support of our future understanding of analytical variability, we propose BIDS-Prov: a representation for neuroimaging provenance. In order to build solutions to better understand (and eventually restrain) analytical variability in neuroimaging, having a complete record of the steps that were applied to a dataset (including all parameters) is an important first step.

### 3.1 Same pipeline, different software: different results

*This work was led by Alexander Bowring as part of his PhD and postdoctoral fellowship at Oxford University. On these two projects, he was co-supervised by Prof. Thomas Nichols from Oxford Uni. and myself. The senior authorship is shared between Prof. Nichols and myself on the two resulting publications.*

**Context** Software is central to modern scientific research and with the development of data science and its subfields (such as bioinformatics or neuroinformatics) the different tools and approaches available to study a dataset have multiplied. Those software have been very valuable to practitioners and brought the capacity to process more data in a shorter amount of time. But overall, they also provide a large number of possible analysis paths to address a scientific question. Practitioners face a highly-complex analysis landscape in which they have to choose between different tools, algorithms, parameters, all of which are available in different environments (different operating systems, different low-level libraries, etc.). While all options are not equally appropriate for a given research problem, there is most of the time a wide range of acceptable options (see (Carp 2012)) for an illustration). Here, we focused on evaluating the impact of using competing neuroimaging research software on the same task-fMRI dataset.

**Data** We selected 3 task-fMRI studies from OpenfMRI (Poldrack et al. 2013b): ds000001 version 2.0.4 (Schonberg et al. 2012), ds000109 version 2.0.2 (Moran et al. 2012), and ds000120 version 1.0.0 (Padmanabhan et al. 2011). These datasets were chosen following an extensive selection procedure in which we sought to find studies with simple analysis pipelines and clearly reported regions of brain activation that would be easily comparable to our own results. Exclusion criteria included the use of custom software, activations defined using small volume correction, and application of more intricate methods such as region of interest and robust regression analysis, which we believed could be impractical to implement across all analysis software.

**Contribution** In (Bowring et al. 2019), our goal was to understand how the choice of a software

package impacts the analysis results. We re-analyzed each of the 3 studies using three widely-used task-fMRI software packages (AFNI, FSL, and SPM). We obtained all information on how to process, analyze, and model each dataset from the publications. In a follow-up work (Bowring et al. 2021), we revisited our findings, seeking to identify the stages of the pipeline where the greatest variation between analysis software was induced.

We made 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.

**Results** Qualitatively we found similarities between packages both in terms of final activated clusters (Fig. 11) and statistic maps (Fig. 12).

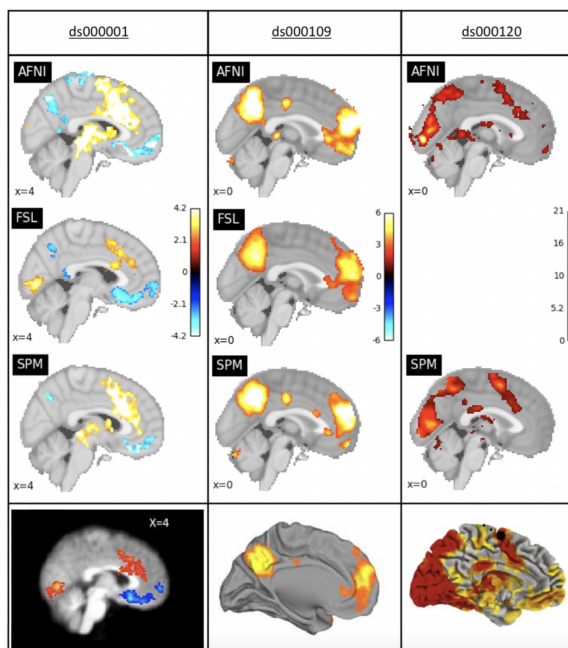


Figure 11: Activation clusters for each reproduction (1 study per column, 1 software per row). The last row includes a copy of the original results. Credit: CC-BY Fig. 1 from (Bowring et al. 2019)

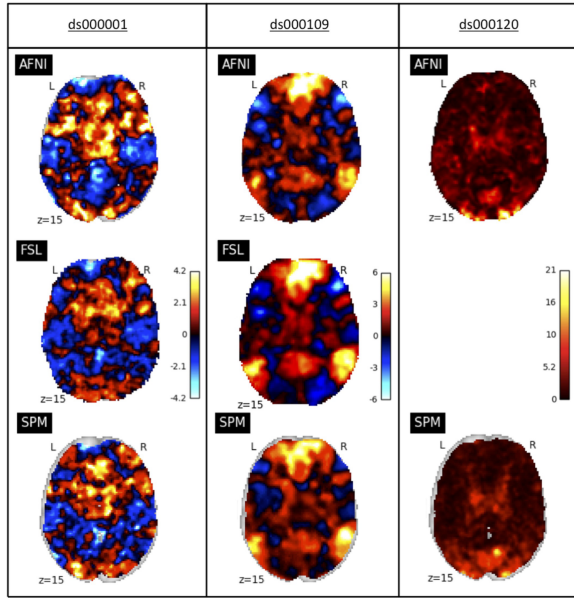


Figure 12: Statistic maps for each reproduction (1 study per column, 1 software per row) Credit: CC-BY Fig. 2 from (Bowring et al. 2019)

However, we also discovered marked differences, such as Dice similarity coefficients ranging from 0.000 to 0.769. Overall the degree of variability increased along the processing pipelines.

In a follow-up work, we carried out further analyses on the three datasets in order to better understand the sources of the observed variability. We employed a common processing strategy across parts of the analysis workflow (in particular fMRIPrep (Esteban et al. 2019) for preprocessing) and then used specific procedures from the three software packages (AFNI, FSL and SPM) across the remaining steps of the pipeline (see Fig. ??). Using our expertise, we

identified 5 possible sources of software variability: 1- Preprocessing, 2- 1<sup>st</sup>-level signal model (including differences in HRF, orthogonalisation of regressors), 3- 1<sup>st</sup>-level drift model (different approaches to remove slow-signal drifts), 4- 1<sup>st</sup>-level noise model (including differences in assumptions to estimate noise parameters across the brain) and 5- 2<sup>nd</sup>-level model and inference (including different assumptions for group modeling: heteroscedasticity versus homoscedasticity). Of note, the analyses performed in each of the three software had already been carefully aligned using all the available options in the first set of results. But a perfect correspondence was impossible and the goal here was to model the remaining variations which illustrate differences in the underlying assumptions that are often encoded deeply into the software package.

We used quantitative methods to compare the statistical maps and isolate the main stages of the workflow where the three packages diverge. Across all datasets, we found that variation between the packages' results is largely attributable to a handful of individual analysis stages, and that these sources of variability were heterogeneous across the datasets (e.g. choice of first-level signal model had the most impact for the ds000001 dataset, while first-level noise model was more influential for ds000109 dataset). We also observed areas of the analysis workflow where changing the software package causes minimal differences in the final results, finding that the group-level results were largely unaffected by which software package is used to model the low-frequency task-fMRI drifts.

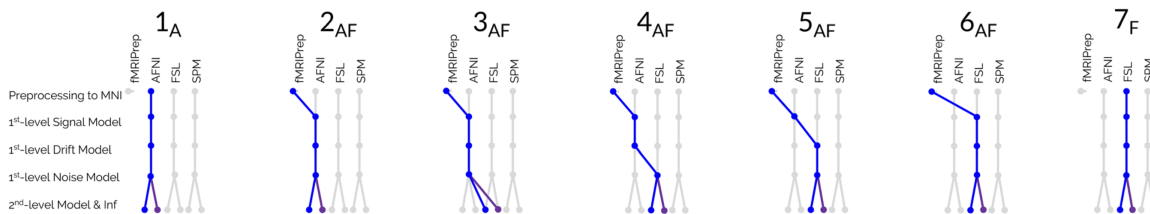


Figure 13: Step-by-step re-analyses from an AFNI analysis ( $1_A$ ) to an FSL analysis ( $7_F$ ). For example:  $3_{AF}$  is a hybrid re-analysis using the common preprocessing strategy (with fMRIPrep), AFNI for 1<sup>st</sup>-level model, drift and noise modelling and FSL for 2<sup>nd</sup>-level model and inference. Credit: CC-BY Fig. 1 from (Bowring et al. 2021)



### Take-Home Message

- task-fMRI results can vary across neuroimaging software packages even when considering aligned pipelines. Observed variations were stronger for thresholded maps.
- Across the three reproductions performed, there was no consensus as to which steps were the most impactful.

### Publications

Bowring, Alexander, **Camille Maumet\***, and Thomas E. Nichols\* (2019). “Exploring the Impact of Analysis Software on Task fMRI Results”. In: *Human Brain Mapping* 40.11, pp. 3362–3384. DOI: 10.1002/hbm.24603.

Bowring, Alexander, Thomas Nichols\*, and **Camille Maumet\*** (2021). “Isolating the Sources of Pipeline-Variability in Group-Level Task-fMRI results”. In: *Human Brain Mapping*. DOI: 10.1002/hbm.25713.

### Open artefacts

All scripts and results are available through our Open Science Framework Projects at <https://osf.io/U2Q4Y/> and <https://osf.io/axy3w/> (for each publication). The resulting statistic maps are available on NeuroVault: <https://neurovault.org/collections/4110/> and <https://neurovault.org/collections/8381/> (ds0000001), <https://neurovault.org/collections/4099/> and <https://neurovault.org/collections/7113/> (ds000109), <https://neurovault.org/collections/4100/> and <https://neurovault.org/collections/9324/> (ds000120). All analysis scripts are available through Zenodo at <https://doi.org/10.5281/zenodo.1413540> and <http://doi.org/10.5281/zenodo.5070414>.

### Related work

**Exploring analytical variability across tools in EEG:** In a collaboration with Dr. Mahmoud Hassan, specialized in EEG analysis, we proposed to use the same type of approach but applied to EEG preprocessing. The work was initiated by Nina Forde during her internship at Inria and then continued by Dr. Aya Kabbara, a postdoctoral fellow at Lebanese University, Lebanon. Both of them were co-supervised between Dr. Hassan and myself and the senior authorship is shared on the corresponding publication.

In (Kabbara et al. 2023) we studied software-induced variability in EEG. To this aim, we reproduced a recent study by Williams and colleagues (Williams et al. 2021) that was characterized by the use of a dataset with a large sample size (N=500) and was publicly available. We showed a good degree of convergence in terms of the general profile of ERP waveforms, peak latencies and effect size estimates related to specific signal features. However, considerable variability was also observed in the magnitude of the absolute voltage observed with each software package.


Kabbara, Aya, Nina Forde, **Camille Maumet\***, and Mahmoud Hassan\* (2023). “Successful reproduction of a large EEG study across software packages”. In: *Neuroimage: Reports* 3.2, p. 100169. DOI: 10.1016/j.ynirp.2023.100169.

---


\* Both senior authors contributed equally.


## 3.2 Same data, different experts: different results

This work was done as part of an international collaboration with 200+ researchers spearheaded by Rotem Botvinik-Nezer and directed by Dr. Tom Schonberg from Tel Aviv University (Israel) in collaboration with Prof. Russel Poldrack from Stanford University (USA). We participated with Prof. Thomas Nichols and Dr. Alexander Bowring from Oxford University (UK) as one of the 70 recruited teams.

 **Context** Using the same data to answer the same scientific question, researchers may reach contradictory conclusions depending on the analytical pipeline they choose. For many years this problem had remained rampant in experimental sciences but recently many studies demonstrated this issue including in: climate science (Knight et al. 2007), bioinformatics (Cashman et al. 2018), psychology (Silberzahn et al. 2018), social sciences (Breznau et al. 2021). Overall this phenomenon has reduced confidence in research findings and is effectively an important driver of the reproducibility crisis (Hoffmann et al. 2021).

Here, this phenomenon was tested in a task-fMRI setting: 70 teams were recruited to analyze the same dataset in order to answer the same set of nine yes/no research questions. This approach, known as "many analyst analysis" was initiated in the field of psychology (Silberzahn et al. 2018) and is increasingly used as a method to uncover analytical variability (Aczel et al. 2021)

 **Data** The MRI dataset is publicly available on OpenNeuro with DOI: 10.18112/openneuro.ds001734 under CC0 (Public Domain) and is described in details in a data paper (Botvinik-Nezer et al. 2019). The dataset comprised data from N=108 participants (a total of N=119 healthy participants completed the experiment, 11 were excluded based on too many failed trials or memory errors when processing the data). All participants gave written informed consent. The study was approved by the institutional review board at the Sheba Tel Hashomer Medical Center and the ethics committee at Tel Aviv University.

 **Contribution** A task-fMRI dataset was acquired in order to test a set of hypotheses extracted from two prior (high-impact) publications that investigated a mixed gambles task (De Martino et al.

2010; Tom et al. 2007). The central question behind this experiment was to better understand the neural underpinning of losses and gains in the brain and specifically whether loss and gain share the same system (responding negatively to losses and positively to gains) or if two separate systems are used. Special care was taken in order to ensure that statistical power would be sufficient, in particular by tailoring the design of the experiment (task onset times).

Recruitment was performed through social media (Twitter, Facebook) and in-person at the 2018 annual meeting of the Society for Neuroeconomics. Teams were tasked to independently analyze the data using their favourite pipeline to answer 9 yes/no research questions. All team members signed a non-disclosure agreement in which they agreed not to discuss their results with anyone until the end of the study.

The 9 yes/no research questions all referred to the presence of activated voxels in a given brain region for a given aspect of the task under consideration. As an illustrative example, the first research question was: *Is there a parametric effect of gain: Positive effect in ventromedial prefrontal cortex (PFC) - for the equal indifference group.* In brief "parametric effect of gain" means that the level of activation is proportional to the amount of money won, the *ventromedial prefrontal cortex* is a brain region and the *equal indifference group* corresponds to a subset of the participants for which the design of the experiment involved potential losses that were half of the potential gains (as opposed to *equal range group* for which the range of gains was equal to the range of losses).

For each of the 9 yes/no research questions (see full list in the original publication (Botvinik-Nezer et al. 2020)) each team had to provide a yes/no answer and a level of confidence (from 1 (not at all) to 10 (extremely)). They were also requested to share their results as statistic and activation maps. To de-



scribe their methodology each team had to provide a complete checklist following the COBIDAS guidelines (Nichols et al. 2017b).

In our team, we chose to run a pipeline based on SPM. We pre-registered our analysis on the OSF: [osf.io/dq9eg/](https://osf.io/dq9eg/).

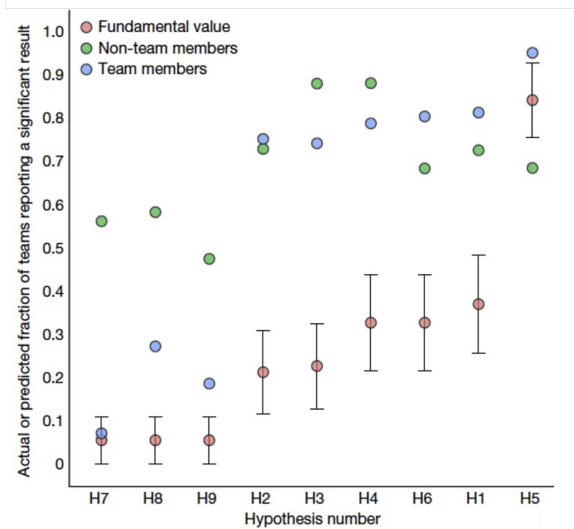


Figure 14: The 9 research questions answered by the 70 teams. Pink dots represent the proportion of teams that replied *yes* (0 = all *No*; 1 = all *Yes*). (Note: green and blue dots are related to the market analysis not described in this manuscript) Credit: Reproduced with permissions from Fig. 1 in (Botvinik-Nezer et al. 2020)

### Results

An interesting note is that, as a result of this experiment, each of the 70 teams had chosen a different pipeline highlighting the diversity of the approaches chosen by experts in task-fMRI.

The proportion of *yes* (versus *no*) answers across

teams for each of the hypothesis are presented in Fig 14. Overall: 3 research questions (H7, H8, H9) received a *No* answer from over 90% of the teams, 1 research question (H5) received a *Yes* answer from over 80% of the teams. The remaining 5 research questions (H2, H3, H4, H6 and H1) received between 20% and 40% of *Yes* leaving an overall conflicted picture.

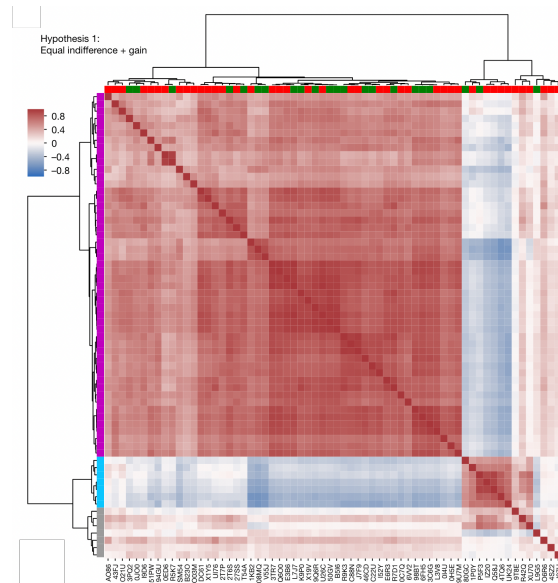


Figure 15: Voxelwise correlation of the statistic maps obtained by the 70 teams for the first research question (H1). Credit: Reproduced with permissions from Fig. 2 in (Botvinik-Nezer et al. 2020)

When looking at the statistic maps, the picture is less conflicted, showing a large proportion of the teams presenting highly correlated maps and this even for the research question that led to the highest level of divergence (i.e. H1) (see Fig. 15).

### Take-Home Message

- In a *many analysts, one dataset* approach, the so-called *NARPS* project demonstrated that the same task-fMRI dataset could lead to different conclusions depending on how it was processed.
- While *yes/no* answers could vary drastically across teams (up to a nearly 40% / 60% split), statistic maps showed a more consistent picture.

 Publication

Botvinik-Nezer, Rotem, Felix Holzmeister, Colin F Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A Mumford, R. Alison Adcock, Paolo Avesani, Blazej M Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari, Nadège Bault, Derek Beaton, Julia Beitner, Roland G Benoit, Ruud M.W.J. Berkers, Jamil P Bhanji, Bharat B Biswal, Sebastian Bobadilla-Suarez, Tiago Bortolini, Katherine L Bottenhorn, Alexander Bowring, Senne Braem, Hayley R Brooks, Emily G Brudner, Cristian B Calderon, Julia A Camilleri, Jaime J Castellon, Luca Cecchetti, Edna C Cieslik, Zachary J Cole, Olivier Collignon, Robert W Cox, William A Cunningham, Stefan Czoschke, Kamalaker Dadi, Charles P Davis, Alberto De Luca, Mauricio R Delgado, Lysia Demetriou, Jeffrey B Dennison, Xin Di, Erin W Dickie, Ekaterina Dobryakova, Claire L Donnat, Juergen Dukart, Niall W Duncan, Joke Durnez, Amr Eed, Simon B. Eickhoff, Andrew Erhart, Laura Fontanesi, G. Matthew Fricke, Shiguang Fu, Adriana Galván, Remi Gau, Sarah Genon, Tristan Glatard, Enrico Glerean, Jelle J Goeman, Sergej A.E. Golowin, Carlos González-García, Krzysztof J Gorgolewski, Cheryl L Grady, Mikella A Green, João F Guassi Moreira, Olivia Guest, Shabnam Hakimi, J. Paul Hamilton, Roeland Hancock, Giacomo Handjaras, Bronson B Harry, Colin Hawco, Peer Herholz, Gabrielle Herman, Stephan Heunis, Felix Hoffstaedter, Jeremy Hogeveen, Susan Holmes, Chuan-Peng Hu, Scott A Huettel, Matthew E Hughes, Vittorio Iacovella, Alexandru D Iordan, Peder M Isager, Ayse I Isik, Andrew Jahn, Matthew R Johnson, Tom Johnstone, Michael J.E. Joseph, Anthony C Juliano, Joseph W Kable, Michalis Kassinopoulos, Cemal Koba, Xiang-Zhen Kong, Timothy R Koscik, Nuri Erkut Kucukboyaci, Brice A Kuhl, Sebastian Kupek, Angela R Laird, Claus Lamm, Robert Langner, Nina Lauharatanahirun, Hongmi Lee, Sangil Lee, Alexander Leemans, Andrea Leo, Elise Lesage, Flora Li, Monica Y.C. Li, Phui Cheng Lim, Evan N Lintz, Schuyler W Liphardt, Annabel B Losecaat Vermeer, Bradley C Love, Michael L Mack, Norberto Malpica, Theo Marins, **Camille Maumet**, Kelsey Mcdonald, Joseph T Mcguire, Helena Melero, Adriana S Méndez Leal, Benjamin Meyer, Kristin N Meyer, Glad Mihai, Georgios D. Mitsis, Jorge Moll, Dylan M Nielson, Gustav Nilsson, Michael P Notter, Emanuele Olivetti, Adrian I Onicas, Paolo Papale, Kaustubh R Patil, Jonathan E Peelle, Alexandre Pérez, Doris Pischedda, Jean-Baptiste Poline, Yanina Prystauka, Shruti Ray, Patricia A Reuter-Lorenz, Richard C Reynolds, Emiliano Ricciardi, Jenny R Rieck, Anais M Rodriguez-Thompson, Anthony Romy, Taylor Salo, Gregory R Samanez-Larkin, Emilio Sanz-Morales, Margaret L Schlichting, Douglas H Schultz, Qiang Shen, Margaret A Sheridan, Jennifer A Silvers, Kenny Skagerlund, Alec Smith, David V Smith, Peter Sokol-Hessner, Simon R Steinkamp, Sarah M Tashjian, Bertrand Thirion, John N Thorp, Gustav Tinghög, Loreen Tisdall, Steven H Tompson, Claudio Toro-Serey, Juan Jesus Torre Tresols, Leonardo Tozzi, Vuong Truong, Luca Turella, Anna E van 't Veer, Tom Verguts, Jean M Vettel, Sagana Vijayarajah, Khoi Vo, Matthew B Wall, Wouter D Weeda, Susanne Weis, David J White, David Wisniewski, Alba Xifra-Porxas, Emily A Yearling, Sangsuk Yoon, Rui Yuan, Kenneth S.L. Yuen, Lei Zhang, Xu Zhang, Joshua E Zosky, Thomas E. Nichols, Russell A Poldrack, and Tom Schonberg (2020). "Variability in the analysis of a single neuroimaging dataset by many teams". In: *Nature* 582.7810, pp. 84–88. DOI: 10.1038/s41586-020-2314-9. URL: <https://www.biorxiv.org/content/10.1101/843193>.

**❄️ Open artefacts**

Each team shared their results on NeuroVault (See <https://neurovault.org/collections/?q=narps>). In addition a detailed description of the selected pipeline was shared in compliance with the COBIDAS guidelines (Nichols et al. 2017b) in this sheet. The pre-registration of the analysis for our team is available on the OSF at: [osf.io/dq9eg/](https://osf.io/dq9eg/).

**Related work**

**NARPS Open pipelines** In a follow up project (currently ongoing), we propose to make public reproductions of the code of the NARPS teams. This project named ‘NARPS open pipelines’ is available [https://github.com/Inria-Empenn/narps\\_open\\_pipelines](https://github.com/Inria-Empenn/narps_open_pipelines) and archived on Software Heritage [swh:1:dir:b1885ec4e4ec39b18076d5009793fdb6ef45365](https://swh:1:dir:b1885ec4e4ec39b18076d5009793fdb6ef45365). This work was initiated by Elodie Germani as part of her Master 2 internship and is now led by Boris Clenet (research engineer) and funded by the Inria exploratory action GRASP. Jérémy Lefort-Besnard (postdoc) has contributed as part of the MIND project funded by the Brittany Region. All of them have worked under my supervision. This project has also received contributions at multiple hackathons, see full list at: [https://github.com/Inria-Empenn/narps\\_open\\_pipelines#credits](https://github.com/Inria-Empenn/narps_open_pipelines#credits).

### 3.3 BIDS-Prov: a provenance framework for BIDS

The development of the BIDS-Prov specification was done as part of an international collaboration with 15+ contributors based in the USA, Canada, Switzerland, UK and France (cf. full list here) co-led between Prof. Satrajit Ghosh from MIT (USA) and myself. Four research engineers worked under my supervision at INRIA to develop the set of examples: Rémi Adon, Thomas Betton, Hermann Courteille and Cyril Regan.

**Context** The NIDM standardization effort (see 2.1 for more on NIDM) was built on the vision that a dataset should be shared with its history. For example, it should be possible to link back activation clusters in task-fMRI to the original task and anatomical images that were used to compute them. This is important first to contextualize the findings but also useful to understand dynamics of data reuse in the community.

While the original NIDM-Workflows effort had been dormant for a few years, the emergence of BIDS (see Section 2.1 for more details) as well as a growing realization that analytical variability can have an impact on neuroimaging findings (see Sections 3.1 and 3.2), created the perfect opportunity to revive provenance in neuroimaging.

Here, we proposed the BIDS-Prov specification to record neuroimaging provenance.

**Contribution** Provenance refers to a detailed description of the sequence of steps that led to the creation of a scientific artefact. Beyond the field of brain imaging, there is a growing interest in many scientific communities to store provenance, in the perspective of improving data sharing through the FAIRification of datasets (see Chapter 2 for more on the FAIR principles). The W3C has proposed a flexible domain-agnostic representation for provenance based on semantic web technologies. In brief the W3C described 3 types of objects:

*An entity* is a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary.

*An activity* is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities.

*An agent* is something that bears some form of responsibility for an activity taking place, the existence of an entity, or for another agent's activity.

In addition, each object has a set of predefined relations with other objects (see Fig. 16).

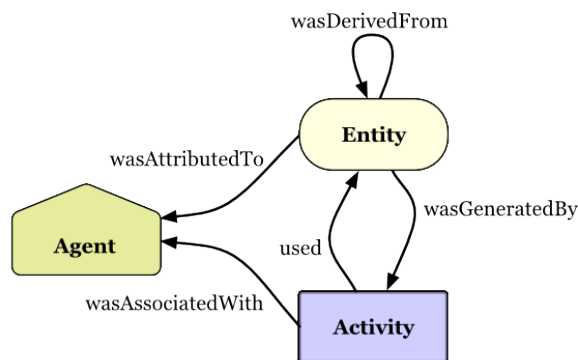


Figure 16: Overview of the W3C PROV model

With BIDS-Prov, we introduced provenance-encoding based on PROV for BIDS. The main challenge was to introduce the use of semantic web technologies – that are better suited to encode graphs such as the data flows described by provenance – while complying with the general principles of BIDS.

**A change of paradigm for BIDS** While metadata in BIDS were encoded in JSON files, our goal in BIDS-Prov was to use semantic web technologies to allow for more flexible graph-like representation. This seemed useful for multiple reasons:

BIDS was initially created to represent and share raw data. In this context, identifying mandatory parameters – i.e. parameters that if missing would impede the analysis of the data – was feasible with expert knowledge of the domain. As an illustration for task-fMRI it was decided to include in BIDS the timing of the task (see section 2.2 for more details on task-fMRI) as further analyses of activation patterns would be impossible without those.

For derivative data (i.e. dataset obtained after some form of processing), identifying which meta-data are crucial for re-usability is extremely complex (if not impossible). This was well-illustrated by the studies that showed that analytical variability can have impact on scientific findings (Botvinik-Nezer et al. 2020; Bowring et al. 2019) or that invalidate the use of certain sets of parameters (see for example (Eklund et al. 2016b) on the value of cluster-forming thresholds for cluster-wise inference).

In BIDS-Prov, contrary to BIDS derivatives, the goal was to encode as many parameters as possible, to make later exploration possible. In this context, using an extensive graph structure to closely match the steps used by a pipeline seemed ideal.

**An extensible graph structure** Another advantage of semantic web graphs is their ability to be merged automatically based on identifiers. We used SHA checksum so that files with identical contents could be represented and identified across different graphs. This may prove particularly useful in a context of an increased use of shared datasets.



### Results

The BIDS-Prov specification is available at: <https://bids.neuroimaging.io/bep028>. Following the BEP guidelines, this specification was sent to BIDS steering committee (in view of opening for community feedback) in September 2023.

A total of 50 examples were built to illustrate BIDS-Prov on AFNI, FSL and SPM: [https://github.com/bids-standard/BEP028\\_BIDSprov](https://github.com/bids-standard/BEP028_BIDSprov).

### Take-Home Message

- BIDS-Prov can be used to model provenance in neuroimaging, compatible with the widely-used *Brain Imaging Data Structure* and with the W3C PROV.
- A set of 50 examples are available at [github.com/bids-standard/BEP028\\_BIDSprov](https://github.com/bids-standard/BEP028_BIDSprov).
- BIDS-Prov was submitted to the BIDS steering committee.



### Publications

Adon, Rémi, Stefan Appelhoff, Tibor Auer, Laurent Guillo, Yaroslav O Halchenko, David Keator, Christopher J Markiewicz, Thomas E. Nichols, Jean-Baptiste Poline, Satrajit Ghosh, and **Camille Maumet** (2021). “BIDS-prov: a provenance framework for BIDS”. In: *Annual Meeting of the Organization for Human Brain Mapping (OHBM)*. Online, South Korea, pp. 1–3. URL: <https://www.hal.inserm.fr/inserm-03478998/document>.



### Talks

Follow the link to access the slides on HAL

“BIDS-Prov: Recording neuroimaging provenance” (2023). In: *Brain Imaging Data Structure (BIDS) derivatives meeting, Copenhagen (Denmark)*. URL: <https://inria.hal.science/hal-04137329>.  
 “Reviewing neuroimaging flexibility: Components and records of provenance” (2022). In: *OHBM 2022 - 28th Annual Meeting of the Organization of Human Brain Mapping*. URL: <https://inria.hal.science/hal-03705748>.

# Chapter 4

## Reusing neuroimaging data

With more and more data being shared across the scientific community, there is a paradigm shift towards reusing existing datasets rather than producing new data for each new scientific question. This practice is probably best illustrated by the growing use of community resources. As an example, datasets from the International Neuroimaging Data-sharing Initiative (INDI) were used in over 900 publications through the period 2010-2016 covering a wide range of scientific domains including: biomedical science, computer science and mathematics (Milham et al. 2018).

### Reusing data as the new normal

The experience of researchers who decided to reuse data rather than acquire their own has not always been smooth. In an editorial, back in 2016, they were even accused of being so-called *data parasites* (Longo and Drazen 2016). A nickname that led to the (satirical) creation of the Research Parasite Award in 2017 (*Research Parasite Award* 2023), a prize that is still ongoing and recognize original secondary use of scientific data (i.e. reuse of data for a different purpose than what they were initially collected for).

While the term "data parasite" was rather strong, the editorial outlined a number of issues that were real. Their first claim was about the expertise that is required in order to acquire and create an original dataset. This echoed a lack of standard and best practices for data sharing that was particularly pronounced a few years back. For example, back in 2016, there was no standard way to store information about an MRI study, leading to a myriad of setups for data storage, rendering impossible the application of any automatic pipeline. Since then, the emergence of best practices, such as the FAIR principles (Wilkinson et al. 2016), as well as the development of standards in the field of brain imaging, such as BIDS (See Section 2.1 for more on BIDS) have greatly improved the process of data sharing.

In addition, the reluctance of part of the research community to see data being reused could (and still can) be related to a lack of recognition for the efforts needed to perform data acquisition and curation. While this problem is still unsolved and the question of credits for those who create the data remains open (White et al. 2022), a number of initiatives have helped. For instance, data papers (Gorgolewski et al. 2013) – scientific publications that are shared with the community to introduce a dataset as an original contribution – are a first way to gain credit when sharing

data. Finally, in a more global dynamic towards open science, sharing data can be beneficial to the career of the individual sharing the data by providing more visibility and opportunities for collaborations (McKiernan et al. 2016). For example, articles with shared data in an open repository have been shown to attract more citations (Piwowar and Vision 2013)

## Data reuse as a way to improve reproducibility

Data reuse can be useful to tackle different sources of irreproducibility, in that it is the logical follow-up to data sharing and therefore allows for the same types of improvements: increasing sample sizes, building more representative datasets, and providing opportunities to test for generalizability (see Chapter 2 for more on data sharing). In addition, reusing and combining existing data, especially in the form of meta-analyses, can be a mean to summarize the current body of knowledge and better identify which results hold in multiple contexts (and are hence generalizable) and which results are divergent from the rest of the literature.

Reuse is not only beneficial to reproducibility but also more sustainable – not only in terms of computational time but also time required by experts and by participants – as it reduces the need to re-acquire data to answer new research questions (Rae et al. 2022). Some authors have also argued that sharing and reusing data is more ethical (Brakewood and Poldrack 2013) as it is the best use of the data that required time from research participants as well as (often publicly-funded) scientific resources. Yet a number of impediments are still in the way of efficient data reuse.

## My contributions

### Meta-analyses

While the sharing and secondary reuse of raw data is relatively new, there is one practice of data reuse that has been going on many years: *meta-analysis*. The goal of a meta-analysis is to combine results from previously published studies to analyze convergence (or lack thereof) across published results.

In the task-fMRI literature, coordinate-based meta-analyses have traditionally been favoured (due to limited availability of data supporting the results, see section 2.2 for more info). Nowadays, the increased number of statistic maps shared with accompanying publications make it finally feasible to perform image-based meta-analyses.

In a first contribution, we tested the validity of meta-analysis methods on task-fMRI data (see Section 4.1) and proposed best practices for neuroimaging reporting to facilitate future meta-analyses.

### Mega-analyses

Mega-analyses are a specific type of meta-analyses in which individual derived datasets from previous studies are combined in a new statistical analysis. Mega-analyses can either be used in



a similar way as a meta-analysis – i.e. mainly to identify convergence or divergence across the literature – or lead to entirely new analyses. This is possible because the individual data are shared and therefore provide more flexibility as to which scientific questions can be explored. Mega-analyses are still seldomly found in the neuroimaging literature but with increasing individual result data being shared, we can imagine that those could flourish in future years.

In a second contribution, we investigated the validity of statistical analyses comparing two datasets derived from different pipelines in a mega-analysis 4.2).

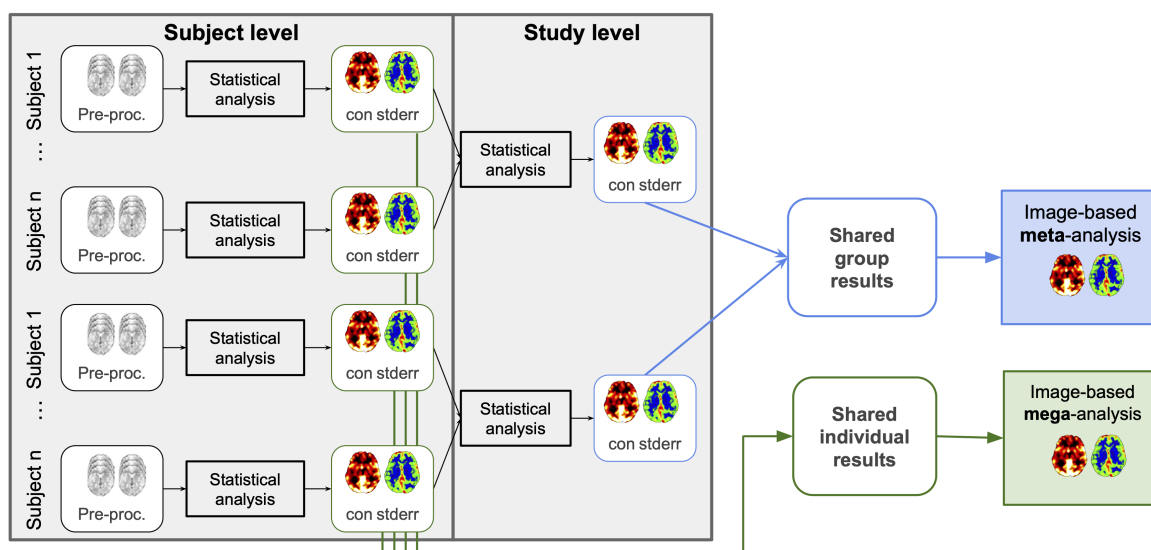


Figure 17: Meta-analyses and mega-analyses.

### Leveraging large data

Finally, data reuse paves the way for new types of analyses in which large amounts of data can be used to inform smaller-scale studies. This is particularly important as large datasets, although invaluable, cannot cover all scientific questions that the research community might want to investigate. For instance, the UK Biobank provides task-fMRI data for an unprecedented number of participants (target N=100 000) but only for a limited number of cognitive tasks. Although large datasets may not be used to answer all research questions, there is an opportunity to leverage them in an agnostic manner and then fine-tune towards a variety of problems.

In a third contribution, we showed how leveraging a large derived dataset can lead to more generalizable representations and improve decoding in task-fMRI (see Section 4.3).



## 4.1 Validity of meta-analysis methods for fMRI

I initiated this work as part of my postdoctoral fellowship at the University of Warwick and Oxford University (UK). This was then continued as a collaboration with Prof. Thomas Nichols from Oxford University (UK).

**Context** Reporting of task-fMRI studies is complex and historically the reporting of task-fMRI studies has proceeded by sharing only the location of a limited number of local maxima (or "peaks"). An approach that is still widespread in the literature (Acar et al. 2022).

While this approach led to the development of so called *Coordinate-Based Meta-Analysis* approaches, the last ten years have seen increased interest in sharing statistic maps in support of task-fMRI studies (Gorgolewski et al. 2015b) (see section 2.2 for more on task-fMRI results and the *NeuroVault* database) and contrast maps (Chen et al. 2017).

This open the door for *Image-Based Meta-Analyses* (IBMA) that are more optimal as they can combine sub-threshold statistic values (see Fig. 18) (Salimi-Khorshidi et al. 2009).

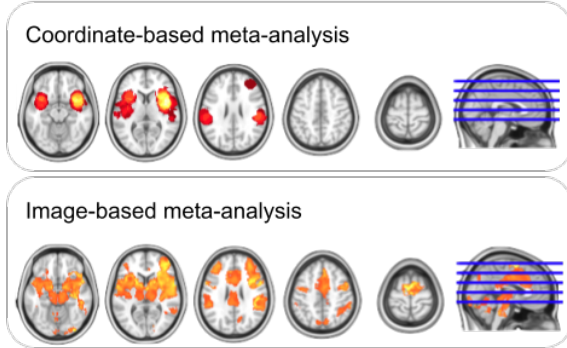


Figure 18: A coordinate-based and an image-based meta-analysis of the same study of pain (data courtesy of the Tracey group at Uni. of Oxford).

Yet, traditional meta-analyses might not be well-suited for task-fMRI data. Here, we assessed the validity of various meta-analyses methods in the presence of violation of their underlying assumptions (as may typically be seen in task-fMRI).

**Contribution** A number of IBMA approaches have been proposed combining: standardised statistics (Z's), just effect estimates (E's) or both effect estimates and their standard errors (E+SE's). While

using E+SE's and estimating between-study variance should be optimal, the methods are not guaranteed to work for small number of studies. Also, it is common to only see standardised estimates being shared, reducing the possible meta-analytic approaches. Finally, because the BOLD signal is non-quantitative, special care has to be taken in order to ensure that E's are expressed in the same units (Nichols 2012; Pernet 2014). We will use the following notations: for study  $i = 1, \dots, k$  we have contrast estimate  $\hat{\beta}_i$ , its contrast variance estimate  $s_i^2$  (i.e. squared standard error), its equivalent Z-statistic map  $Z_i$  and its sample size  $n_i$ .

### Combining contrast estimates and their standard error

The recommended approach is to fit contrast estimates and their standard error with a hierarchical GLM (Cummings 2004), creating a third-level (level 1: subject; level 2: study; level 3: meta-analysis). The general formulation for the study-level data is a meta-regression:

$$\hat{\beta} = X\gamma + \epsilon \quad (1)$$

where  $\hat{\beta} = [\hat{\beta}_1 \dots \hat{\beta}_k]^T$  is the vector of contrast estimates,  $X$  is the  $k \times p$  study-level matrix. While the design matrix  $X$  can take any form, typically a one-sample model is used with  $p = 1$  and  $X$  comprised of a column of ones.  $\gamma = [\gamma_1 \dots \gamma_p]^T$  is a vector of meta-analytic parameters to estimate and  $\epsilon \sim \mathcal{N}(0, W)$  is the residual error term.

In the most general case of a random-effects meta-analysis, i.e. assuming non-zero between-study variance,  $W$  is a diagonal matrix with entries  $\tau^2 + \sigma_i^2/n_i$  where  $\tau^2$  denotes the between-study variance and  $\sigma_i^2/n_i$  denotes the variance of the contrast estimate for study  $i$ . Optimal estimation of Eq. (1) is obtained by weighted least squares giving:

$$\hat{\gamma} = (X^T W^{-1} X)^{-1} X^T W^{-1} \hat{\beta} \quad (2)$$

$$\text{Var}(\hat{\gamma}) = (X^T W^{-1} X)^{-1} \quad (3)$$

However in practice, the weight matrix  $W$  is unknown and has to be estimated from the data. Given  $\hat{W}$  a consistent estimate of  $W$ , the feasible generalized least squares (FGLS) estimator is:

$$\begin{aligned}\hat{\gamma} &= (X^T \hat{W}^{-1} X)^{-1} X^T \hat{W}^{-1} \hat{\beta} & (4) \\ \text{Var}(\hat{\gamma}) &= (X^T \hat{W}^{-1} X)^{-1} & (5)\end{aligned}$$

Although FGLS is asymptotically efficient, it has no exact null distribution for finite samples (Greene 2012). We refer to this standard best-practice as *Mixed-effects (MFX) GLM*.

In a fixed-effects meta-analysis, we assume that there is no or negligible between-study variance. In this case  $W$  is a diagonal matrix with entries  $\sigma_i^2/n_i$ , and we obtain an FGLS estimate by replacing  $\sigma_i^2/n_i$  with  $s_i^2$ . Inference uses a Student distribution with  $(\sum_{i=1}^k (n_i - 1)) - p$  degrees of freedom. This approach will be referred to as *Fixed-effects (FFX) GLM*.

**Combining contrast estimates** If the  $s_i^2$  are unavailable, the contrast estimates  $\hat{\beta}_i$  can be combined by assuming that the within-study contrast variance is negligible in comparison to the between-study variance ( $\sigma_i^2/n_i \ll \tau^2$ ) or that  $\sigma_i^2/n_i$  is constant over studies at each voxel. Then  $W$  has a constant diagonal equal to  $\sigma_C^2$ , where  $\sigma_C^2$  is equal to the sum of the within and between-study variances, though note that we do not separately estimate  $\tau^2$  or  $\sigma^2$  (Mumford and Nichols 2009a). Under these assumptions, Eq. (1) can be solved by ordinary least squares giving:

$$\begin{aligned}\hat{\gamma} &= (X^T X)^{-1} X^T \hat{\beta} & (6) \\ \text{Var}(\hat{\gamma}) &= (X^T X)^{-1} \sigma_C^2 & (7)\end{aligned}$$

Given  $\hat{\sigma}_C^2$  the unbiased sample variance, we obtain the statistics presented in Table 2 for one sample tests. This approach will be referred to as *Random-effects (RFX) GLM* in the following. Inference can be carried out by comparing the RFX GLM statistic to a Student distribution with  $k - 1$  degrees of freedom, this result is not asymptotic, i.e. should work in small samples (Greene 2012).

**Combining standardised statistics** When only test statistic images are available there are sev-

eral alternative approaches available. *Fisher's* meta-analysis provide a statistic to combine the associated p-values (Fisher 1932). *Stouffer's* approach directly combines the standardized statistic (Stouffer et al. 1949). It has also been proposed to weight each study's  $Z_i$  by the square root of its sample size (Zaykin 2011), in an approach we'll refer to as *Weighted Z*. These three methods are all fixed effects methods and do not account for between-study variance; further, they only apply for a one-sample model (no meta-regression). The corresponding statistics and nominal null distributions are presented in Table 2. Non-parametric inference (Holmes et al. 1996; Nichols and Holmes 2002) can also be obtained by sign flipping on the  $Z_i$ 's. This approach will be referred to as *Z permutation*.

**Approximations** In practice, all the methods are either approximate in small samples or rely on assumptions that might not be verified in the context of neuroimaging meta-analyses. Methods based on FGLS (MFX GLM and FFX GLM) have approximate parametric null distributions. The nominal distributions of RFX GLM is under the (unrealistic) assumption of homogeneous standard errors over studies; even if all studies are 'clean' and conducted at the same center, variation in sample size will induce differences in  $s_i^2$ 's. The fixed-effects methods (Fisher, Stouffer, weighted Z and FFX GLM) assume homogeneity across studies, i.e. zero between-study variance. All contrast methods (MFX GLM, RFX GLM, Contrast permutation and FFX GLM) require the contrasts to be expressed in the same units.

**Simulations** We investigated the validity of each estimator (see Table. 2)) with Monte Carlo simulations under null hypothesis  $H_0$  with:

- $\kappa \in 5, 10, 25, 50$  studies;  $n = 20$  subjects; also  $\kappa = 25, n = 100$ .
- $\tau^2 = 0$  (homogeneity) or  $\tau^2 = 1$ ;
- $\sigma_i^2 = n \times 0.25, 0.5, 1, 2, 4$  (homoscedasticity) or varying between 1 and  $\alpha \in 2, 4, 8, 16$ ,
- $10^6$  realisations.

	Meta-analytic statistic	Nominal $H_0$ distrib.	Inputs	Assumptions
MFX GLM	$\sum_i \kappa_i \hat{\beta}_i / \sqrt{\sum_i \kappa_i}$	$\mathcal{T}_{k-1}$	$\hat{\beta}_i, s_i^2$	IGE, large sample.
RFX GLM	$\frac{1}{k} \sum_i \hat{\beta}_i / \sqrt{\widehat{\sigma}_C^2/k}$	$\mathcal{T}_{k-1}$	$\hat{\beta}_i$	IGE; $\tau^2 + \sigma_i^2 = \sigma_C^2 \forall i$
Contrast Perm.	$\frac{1}{k} \sum_i \hat{\beta}_i / \sqrt{\widehat{\sigma}_C^2/k}$	Empirical	$\hat{\beta}_i$	ISE.
FFX GLM	$\sum_i \hat{\beta}_i / s_i^2 / \sqrt{\sum_i 1/s_i^2}$	$\mathcal{T}_{(\sum_{i=1}^k n_i - 1) - 1}$	$\hat{\beta}_i, s_i^2$	$\tau^2 = 0$ , large sample.
Fisher	$-2 \sum_i \ln P_i$	$\chi_{(2k)}^2$	$Z_i$	$\tau^2 = 0$
Stouffer	$\frac{1}{k} \sum_i Z_i / \sqrt{1/k}$	$\mathcal{N}(0, 1)$	$Z_i$	$\tau^2 = 0$
Weighted Z	$\sum_i \sqrt{n_i} Z_i / \sqrt{\sum_i n_i}$	$\mathcal{N}(0, 1)$	$Z_i, n$	$\tau^2 = 0$
Z Perm.	$\frac{1}{k} \sum_i Z_i / \sqrt{1/k}$	Empirical	$Z_i$	ISE.

Table 2: Meta-analysis statistics and their sampling distributions under  $H_0$ . IGE=Independent Gaussian Errors, ISE=Independent Symetric Errors. Note:  $\kappa_i = 1/(\hat{\tau}^2 + s_i^2)$ ,  $P_i = \Phi(-Z_i)$ ,  $\widehat{\sigma}_C^2$  is the unbiased sample variance.

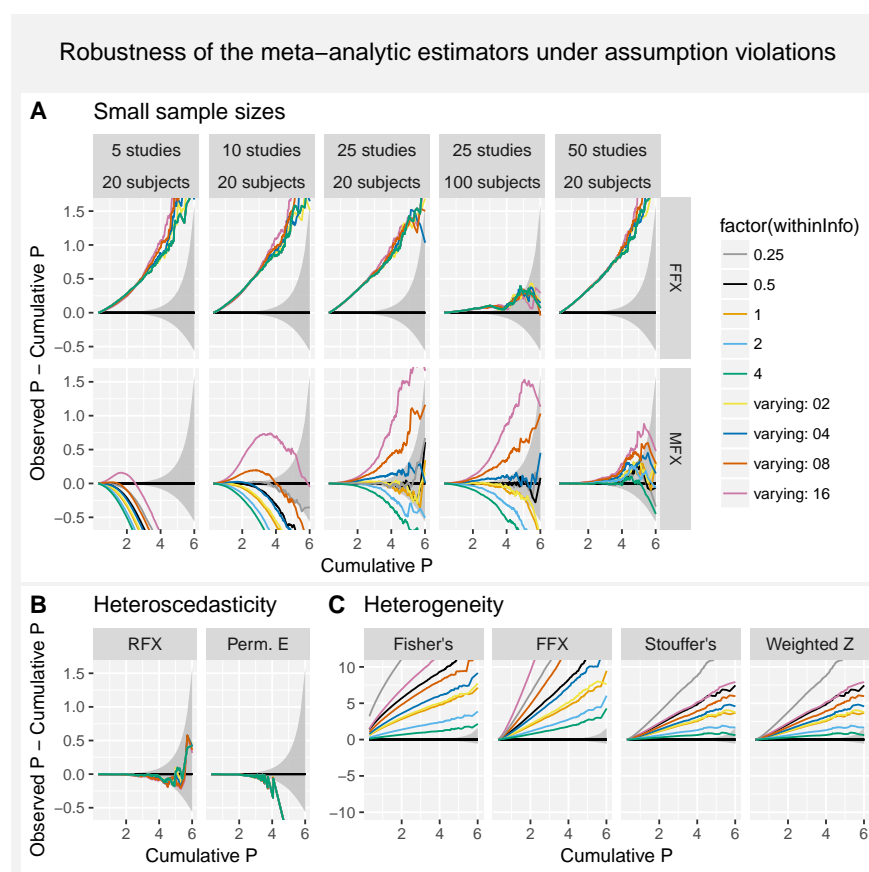


Figure 19: Deviation from theoretical P-values under: small sample sizes (A), heteroscedasticity (B) and heterogeneity (C). P-values are displayed using a negative  $\log_{10}$  scale.

Notations:  $\tau^2$ : between-study variance,  $\sigma_i^2$ : ith study's variance,  $\sigma_C^2$ : usual one-sample variance. IGE=Independent Gaussian Errors, ISE=Independent Symmetric Errors. Note:  $P_i = \Phi(Z_i)$

**Results** Fig. 19A presents the one-sample simulation results in small samples, i.e. under small number of studies or small number of subjects. We focus here on methods for which validity is only guaranteed in large samples: FFX GLM and MFX GLM, under ideal conditions otherwise (i.e.  $\tau^2=0$  for FFX GLM and  $\tau^2=1$  for MFX GLM). When the number of subjects is small, FFX GLM is invalid for all within-study variances investigated, regardless of the number of studies included in the meta-analysis. On the other hand MFX GLM is conservative for small number of studies and constant within-study variance. More surprisingly, while MFX GLM is valid for constant within-study variances it is invalid in the presence of large variations in the within-study

variances, regardless of the number of subjects included in each study.

Fig. 19B presents the one-sample simulation results under heteroscedasticity ( $\sigma_i^2$  varying across studies). We focus here on methods for which validity is only guaranteed under homoscedasticity: RFX and Contrast permutation, in a sample of 25 studies with 20 subjects each under ideal conditions otherwise (i.e.  $\tau^2=1$ ). RFX GLM and contrast permutation are robust to heteroscedasticity for all settings studied. RFX GLM is closer to nominal. For small P-values, Contrast Permutation is conservative as expected due to the discrete nature of its distribution.

Fig. 19C presents the one-sample simulation results under heterogeneity ( $\tau^2 > 0$ ). We focus here on methods for which validity is only guaranteed under homogeneity: Fisher, Stouffer, Weighted Z and FFX GLM, with a sample of  $k = 25$  studies with  $n = 20$  subjects each. All fixed-effects methods are invalid under heterogeneity.

### Take-Home Message

- When data are available, Image-Based is preferred to Coordinate-Based meta-analysis.
- In practice, it is difficult to use the gold standard Mixed-Effects GLM.
- Random-effect meta-analysis method is the most robust to small sample sizes, heteroscedasticity and can handle heterogeneity.

### Publications

**Maumet, Camille** and Thomas E. Nichols (2023). "Validity and Accuracy of Neuroimaging Image-Based Meta-Analysis with Z-statistics". Preprint in preparation.

### Related work

#### Review and recommendations for task-fMRI reporting in view of future meta-analysis

In a project led by Freya Acar as part of her PhD fellowship at the Uni. of Ghent (Belgium) in co-direction between Dr. Ruth Seurinck and Dr. Beatrijs Moerkerke, we reviewed existing practices to report task-fMRI studies and summarized best practices to enable future image-based meta-analyses.

Acar, Freya, **Camille Maumet**, Talia Heuten, Maya Vervoort, Han Bossier, Ruth Seurinck, and Beatrijs Moerkerke (2023). "Improving the Eligibility of Task-Based fMRI Studies for Meta-Analysis: A Review and Reporting Recommendations". In: *Neuroinformatics*. DOI: 10.1007/s12021-023-09643-5.

**Same data meta-analysis** In an ongoing project, led by Dr. Jérémy Lefort-Besnard (part of his postdoctoral fellowship under my supervision) in collaboration with Prof. Thomas Nichols, we study different methods for same data meta-analysis. Those are specific types of meta-analysis that are suited to combine results of previous studies obtained on the same dataset. Our objective with this work is to propose statistical approaches to integrate results over different pipelines, an approach known as multiverse analysis.

Lefort-Besnard, Jeremy, Thomas E. Nichols\*, and Camille Maumet\* (2023). “Systematic review and evaluation of meta-analysis methods for same data meta-analyses”. In: *Cogbases 2023 Workshop*. Paris, France. URL: <https://inserm.hal.science/inserm-04222730/document>.

## 4.2 fMRI mega-analyses with different preprocessing

A work led by Xavier Rolland as part of his PhD in co-direction between Dr. Christian Barillot, Dr. Pierre Maurel and myself. This work was then extended by Elodie Germani as part of her PhD in co-direction between Prof. Elisa Fromont and myself, this project was co-supervised with Prof. Pierre Maurel.

**Context** Beyond meta-analyses that focus on combining the results of previously published studies (see Section 4.1), the goal of mega-analyses (also known as Individual Participant Data (IPD) meta-analyses) is to perform a new analysis by reusing individual data from previous studies.

In the field of task-fMRI this can be made possible by the sharing of statistic maps at the participant level. When raw data are available, another approach for mega-analyses is to re-process all data through the same preprocessing and subject-level analysis pipelines. Yet, with the emergence of data sharing and the increase in sample sizes there would be great value in directly combining the subject-level statistic map provided by each study. In fact this would be similar to how typical coordinate-based meta-analyses are performed, where studies are analyzed together (regardless of the original preprocessing and subject-level analysis pipeline) but would provide more flexibility as to which research questions can be investigated.

In this work, we evaluated the statistical validity of (two-sample) analyses that combine together task-fMRI statistic maps that were obtained with different pipelines. Our experiments were performed under the null hypothesis and any excess detections are therefore demonstrating invalidity.

**Data** We used data from the Human Connectome Project: Young Adult (HCP Young Adult). Written informed consent was obtained from participants and the original study was approved by the Washington University Institutional Review Board. We agreed to the Open Access Data Use Terms<sup>1</sup> before getting access to the data.

**Contribution** Our contributions are two-folds. First, in (Rolland et al. 2022), we tested the va-

lidity of between-group analyses using subject data processed with different pipelines. Second, in (Germani et al. 2023) we made the corresponding dataset available as an open resource for the scientific community. We believe that this dataset will be useful to neuroimagers who would like to get better insights into analytical variability but also more broadly to scientists studying software variability and their impact on scientific findings (see for example (Acher 2022)).

In order to assess the validity of mega-analyses combining data preprocessed with different pipelines, we performed between-group analyses under the null hypothesis (making any detection a false positive) (see Fig. 20).

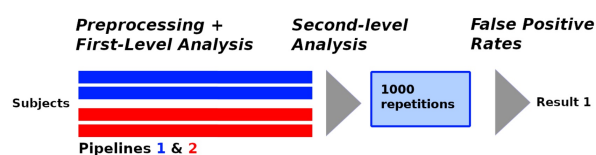


Figure 20: Between-group analyses were performed using individual data processed with two different pipelines. The second-level analysis was repeated 1000 times and false positive rate was estimated across the repetitions. Credits: Fig. 1 from (Rolland et al. 2022) adapted, © 2022 IEEE.

Participants were randomly drawn from the HCP Young Adult and affected to Pipeline 1 or Pipeline 2. The data were preprocessed according to the selected pipeline (see below for the parameters that varied across pipelines). We then performed a two-sample t-test to detect any significant differences in mean across the two groups (preprocessed with Pipeline 1 and 2 respectively). We thresholded the maps with an FWE-corrected threshold at  $P < 0.05$  for both positive and negative effects. The empirical false positive rate was computed by counting the number of observed significant differences across 1 000 repetitions. We compared this empirical false positive rate

<sup>1</sup>[humanconnectome.org/study/hcp-young-adult/document/wu-minn-hcp-consortium-open-access-data-use-terms](https://humanconnectome.org/study/hcp-young-adult/document/wu-minn-hcp-consortium-open-access-data-use-terms)



to the nominal rate (i.e. 5% in each tail and therefore in total 10% chance to get at least one false positive detection in the brain on average).

We varied 3 parameters in the preprocessing pipeline: 1/ the smoothing kernel (the full width at half maximum (FWHM) of the Gaussian kernel was 5mm or 8mm); 2/ the number of motion regressors included as nuisance variables in the GLM for the first-level analysis (0, 6 or 24 motion regressors); 3/ the function used to model the brain response to a stimuli: the canonical HRF or an HRF with derivatives. Those options were selected to reflect typical variations observed in the literature (Carp 2012). Apart from these parameters, each pipeline used the default settings. In total, we tested 12 different subject-level pipelines (2 FWHM x 3 numbers of motion regressors x 2 HRF), implemented with SPM (Penny et al. 2011).

**Results** To ease presentation of the results, the pipeline with smoothing kernel 5mm, 24 motion regressors and canonical HRF was used as a reference in the following (referred to as the *default pipeline*). Using the default pipeline in both groups led to conservative results with a false positive rate of 0.040. The empirical false positive rate obtained when comparing each of the 11 pipelines to the default pipeline are presented in table 3.

	Smoothing, 5 mm		Smoothing, 8 mm	
	No der.	Der.	No der.	Der.
0 motion reg.	<b>0.109</b>	0.097	<b>0.237</b>	<b>0.245</b>
6 motion reg.	0.044	0.046	<b>0.139</b>	<b>0.145</b>
24 motion reg.		0.035	<b>0.113</b>	<b>0.131</b>

Table 3: Empirical false positive rates for analyses comparing the 11 alternative pipelines to the default pipeline (smoothing=5mm, canonical HRF. and 24 motion reg.) at FWE- corrected  $p < 0.1$  two-tailed. Bold indicates invalid results (i.e.  $> 0.1$ ). The false positive rate obtained for same pipeline analysis (default pipeline in both groups, corresponding to the grey cell) was equal to 0.040. reg.=regressors, der.=derivatives. Credits: Table. 1 from (Rolland et al. 2022), adapted © 2022 IEEE.

When varying a single factor (see cells with a blue background in Table 3), the temporal derivatives of the HRF were the least impacting of all three factors with a false positive rate of 0.035, and similarly to

0.044 for 6 motion regressors (instead of 24). Smoothing was the most impacting factor with a false positive rate of 0.113. When varying multiple factors, all comparisons with different smoothing kernel (FWHM=8mm) led to invalid results. Other parameters being equal, comparisons with same smoothing always led to smaller false positive rates (e.g. see column 2 versus 4). Similarly, comparisons with different numbers of motion regressors (especially with no motion regressors in the alternative pipeline) gave higher false positive rates than comparisons with the same number of motion regressors (line 1 versus line 3). More generally, multiple varying factors always led to higher false positive rates.

P-P plots are presented in Fig. 21. Positive differences outside the confidence interval indicate invalidity (P-values smaller than expected) whereas negative differences indicate conservativeness.

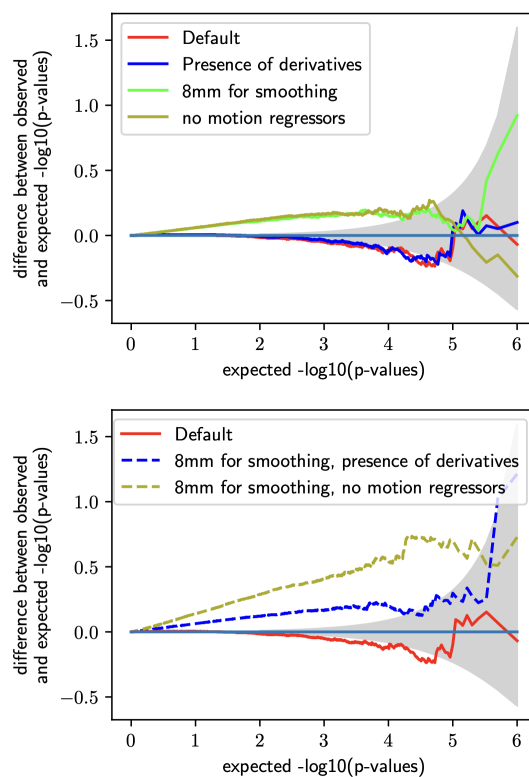


Figure 21: P-P Bland-Altman plots for between-group analyses between the default pipeline and pipelines with one (top) or two (bottom) varying parameters (as stated in the legend). The 95% confidence interval is plotted as a grey shade. The curve for the same pipeline analysis (alternative pipeline equal to default) is plotted in red. Credits: Fig. 2 from (Rolland et al. 2022) © 2022 IEEE.

The P-P plots confirm the behaviour observed on the false positive rates: the choice of smoothing and number of regressors have the strongest impact and varying more parameters leads to stronger differences.

### Take-Home Message

- task-fMRI results from different pipelines cannot be systematically combined directly in a mega-analysis.
- This calls for sharing more information on the pipeline used when sharing data.

### Publications

Germani, Elodie, Elisa Fromont, Pierre Maurel, and **Camille Maumet** (2023). “The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis”. Preprint in preparation.

Rolland, Xavier, Pierre Maurel, and **Camille Maumet** (2022). “Towards efficient fmri data reuse: can we run between-group analyses with datasets processed differently with spm ?” In: *IEEE International Symposium on Biomedical Imaging (ISBI)*. Calcutta, India: IEEE, pp. 1–4. URL: <https://hal.science/hal-03607384>.

### Open artefacts

All the scripts used to perform the study (subject-level processing and analysis, group-level analysis, false positive rate estimation) are available at <https://github.com/Inria-Empenn/pipelines> and archived on Software Heritage: [swh:1:dir:a23afdc15cdf948445ec67ffea14e2c637925048](https://swh1.dir.a23afdc15cdf948445ec67ffea14e2c637925048). The data paper is about to be shared as a preprint. The corresponding dataset will be shared on Public-nEUro.



### 4.3 Leveraging large datasets for new analyses

A work led by Elodie Germani as part of her PhD in co-direction between Prof. Elisa Fromont and myself.

**Context** As more shared data are made available to the research community there is an opportunity to develop models of the brain that rely on larger and crucially more diverse samples. This is essential in order to get more generalizable models that can be applied to different populations but also in different contexts (different acquisition instruments and protocols, different post-processing pipelines, etc.).

In particular the NeuroVault online database, which is dedicated to the collection of statistic maps from task-fMRI studies (see Section 2.2) has been flourishing since its creation about 10 years ago. As of today (October 24, 2023), it includes over 450 000 images<sup>2</sup>. NeuroVault is not only large but also includes data from a wide range of contributors which effectively covers a diversity of practices in the task-fMRI community.

Here, we studied the benefits of using NeuroVault, in a self-taught learning framework to model brain statistic images. We then fine-tuned a classifier on a decoding problem in which the goal was to predict the task (or cognitive state) associated with a given task-fMRI data.

**Contribution** First, we trained an autoencoder to learn how to recompute any statistic maps from the NeuroVault database. This allowed us to obtain a low-level representation of an task-fMRI statistic map (known as *latent space*). Then, the encoding part was used as an initialization and fine-tuned to decode task-fMRI statistic maps (see Fig. 25 for an overview).

**Data** The convolutional autoencoder was built on a subset of the NeuroVault database including 25 000+ statistic maps. Those were selected according to the following criteria: modality was “fMRI-BOLD”; all required metadata were present (“is\_valid” to True), registered in MNI space (“not\_mni” to False); type of map was “T map” or

“Z map”; thresholded statistic maps were excluded. The classification performance was assessed on two datasets:

- Human Connectome Project (Van Essen et al. 2013): 18 000 statistic maps for a total of 787 subjects from HCP Young Adult Project 900 release corresponding to 7 tasks and 23 contrasts.
- Brainpedia (Varoquaux et al. 2018): 6 500 statistic maps from 30 task-fMRI studies corresponding to 36 cognitive processes.

**Results** **Autoencoder** Reconstructed statistic maps using the autoencoder had a correlation greater than 0.85 with the original maps. An example of reconstruction is illustrated in Fig. 22.

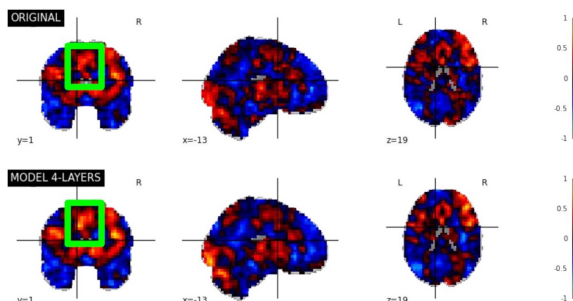


Figure 22: A statistic map reconstructed with the convolutional autoencoder (2<sup>nd</sup> row) compared to the original (1<sup>st</sup> row). Credits: CC-BY Fig. 4 in (Germani et al. 2022).

**Decoding performance** We compared the performance of the classifier pre-trained using the convolutional autoencoder (referred to as *pre-trained CAE* in the following) with a classifier built from scratch (referred to as *Default algorithm*). The classification task was performed in different contexts.

First, we tested the performance obtained for different sample sizes from subsets of the Human Connectome Project (see Fig 24). Over all tested sample sizes (N=50, 100, 200, 400) the pre-trained classifier always outperformed the default classifier. But as

<sup>2</sup><https://neurovault.org/api/images/>

sample sizes increased the difference between the two was less pronounced.

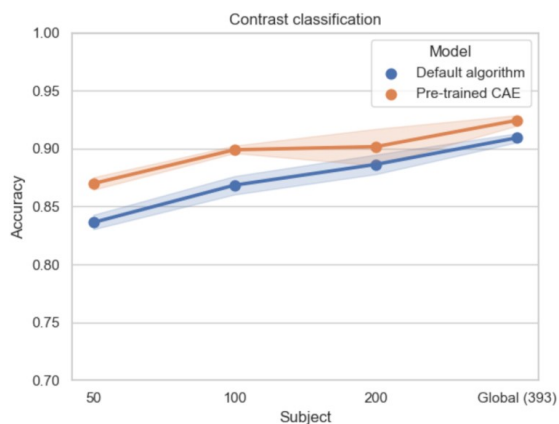


Figure 23: Mean accuracies and standard errors on classification of the contrasts in the Human Connectome Project dataset for the *Default algorithm* (blue) and *pre-trained CAE* (orange). Credits: CC-BY Fig. 5 from (Germani et al. 2022)

Second, we studied the performance on a more

complex task using the Brainpedia dataset. The goal was to predict the cognitive state (in each of the 30 studies) by training on a smaller number of samples (see Fig 24). For both sample sizes the performance of the pre-trained classifier were superior.

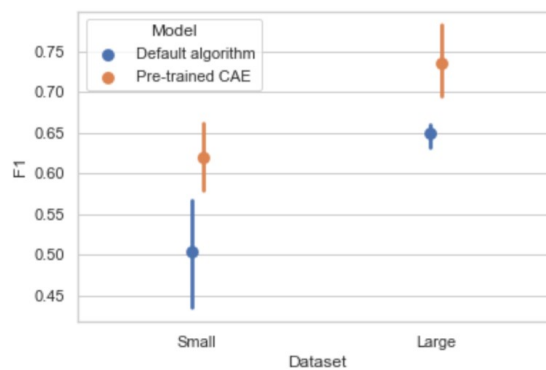


Figure 24: F1-scores and standard errors of the classification of mental concepts on the Brainpedia dataset for the *Default algorithm* (blue) and *pre-trained CAE* (orange). Credits: CC-BY Fig. 8 from (Germani et al. 2022)

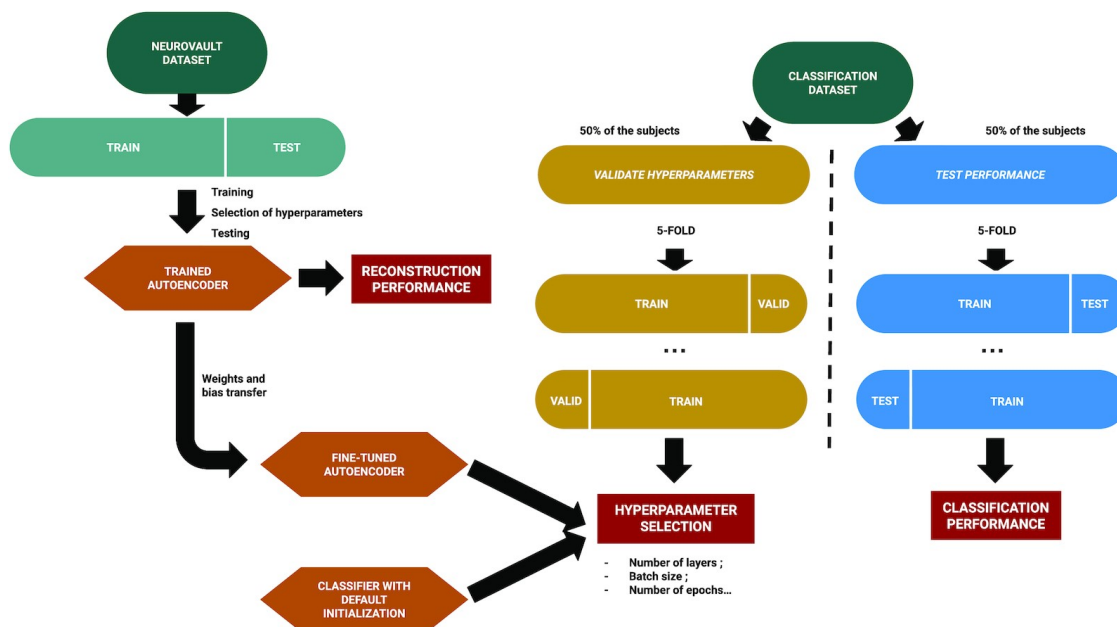



Figure 25: Self-taught learning framework. NeuroVault dataset was used to train a convolutional autoencoder. The encoder was used to initialize a convolutional neural network and to train it to classify new datasets. These classification datasets were split in 2 disjoint datasets: a “validation” set used to optimize hyperparameters and a “test” set to evaluate performance. In each, a 5-fold cross-validation was performed. Credits: Reproduced from CC-BY Fig. 1 in (Germani et al. 2022)

**Take-Home Message**

- We leveraged the NeuroVault database in a self-taught learning framework to improve decoding classification.
- We successfully built an autoencoder to reconstruct any statistic maps from Neurovault.
- We showed better decoding performance of the pre-trained classifier based on the autoencoder. The gap in performance is more marked on complex tasks (e.g. small samples, heterogeneous data, etc.).

 **Publications**

Germani, Elodie, Elisa Fromont\*, and **Camille Maumet\*** (2023). “On the benefits of self-taught learning for brain decoding”. In: *GigaScience* 12, giad029. ISSN: 2047-217X. DOI: 10.1093/gigascience/giad029.

 **Open artefacts**

The code to run the experiments and create the figures and tables is available at [https://gitlab.inria.fr/egermani/self\\_taught\\_decoding](https://gitlab.inria.fr/egermani/self_taught_decoding) and archived on Software Heritage swh:1:dir:caaa671e678e14c874727d972052064a8802bfdc. The derived data produced in this study are available on Zenodo DOI: 10.5281/zenodo.7046055.

---

\* Both senior authors contributed equally.

# Chapter 5

## Open Science

The previous three chapters have provided an overview of my scientific contributions to the field of neuroimaging reproducibility. Throughout my career, I have been lucky to interact closely with the open (neuro)science community and this has shaped not only the way I do research but also the type of research questions I would like to answer. In this last chapter, I focus on my path into, and contributions to Open Science.

### A personal journey into open science

The theme of my postdoctoral fellowship ("neuroinformatics methods for neuroimaging meta-analysis") led me early on to interact closely with international collaborative efforts and to advocate for sharing more than standard research papers in support of scientific findings. At that time, *open science* to me was closely related to the notion of **transparency** and to computational reproducibility, i.e. how we as researchers can share more than just papers so that others in the community can fully understand what we have done and therefore better contextualize our findings.

During that period, I attended my first Brainhack (Paris in 2013) and slowly I started to realize that open science also brings a cultural change as to how research is done (and which type of research we can pursue) towards more **collaborations**. The Brainhacks are hackathons for the neuroscience community. Run over the course of 2 or 3 days, their motto is "*collaboration not competition*". The projects developed during those hackathons are proposed by the participants themselves and the content evolves based on interactions between attendees. In addition to the projects, there is also ample time for discussions amongst participants as well as lightning talks, known as *unconferences*, in which attendees get the opportunity to share an issue they face, an interesting program/tool, an open question etc. I was inspired by the welcoming atmosphere at Brainhack and the opportunity to discuss how we do science as opposed to presenting shiny end-results (as would typically be expected in other international conferences). With time, attending the Brainhacks made me realize that we had an opportunity to do science more collaboratively: exchanging ideas early on (to get feedback, collaborations, avoid duplicate work), finding new collaborators with similar interests, building research projects that we could not achieve alone.

A few years later, I had the chance to meet with an incredible team of researchers as I chaired the

Open Science Special Interest Group of the Organization for Human Brain Mapping (OHBM). This group is in charge of the organization of two main events: the OHBM Brainhack (that happens a few days before the main OHBM conference) and the Open Science Room (a mini-conference run in parallel to OHBM with a program centered on open science). That year was marked by the Covid crisis that forced us to pivot to building online events, just as we had secured a conference room in Montreal. This intense experience made me realize the importance of inclusive events, and the purposefulness required to make it all work. **Inclusivity** in open science is an answer to a very simple question: "*Open to who?*". If I am making all my code publicly available but that code requires an extremely high-powered computer that no one has, is it still open science? While this is an extreme case, the same question applies to the use of proprietary software or more broadly to expensive imaging instruments (such as MRI). The work on low-field MRI led by Prof. Obungoloch from Mbarara University of Science & Technology (Uganda) (Obungoloch et al. 2018), which was featured in the Open Science Room in 2019, is an illustration of imaging research that takes into account the very limited availability of (high-field) MRI scanners in Africa. Inclusivity can be seen as a necessity for social justice (and for everyone to get the opportunity to do science) but it is also essential to produce the best science (see for example (Hofstra et al. 2020; Whitaker and Guest 2020)). Since then, I continuously took part in efforts that focus on bringing more inclusivity in science. Locally, I joined the gender-equality committee of IRISA / Inria Rennes as a member in 2021 and I have been co-chair of this group with Dr. Nicolas Markey since 2021.

The three facets of open science described above are therefore: 1- sharing research artefacts for transparency, 2- building a more collaborative research 3- acting for a more inclusive science. This nicely echoes the definition of Open Science proposed by the UNESCO in its recommendations on Open Science (UNESCO 2021) (see Fig. 26):

[...] open science is defined as an inclusive construct that combines various movements and practices aiming:

- to make multilingual scientific knowledge openly available, accessible and reusable for everyone,
- to increase scientific collaborations and sharing of information for the benefits of science and society, and
- to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.

## Contributions

Since 2013, I have actively taken part in Open Science communities. The first ones included the Brainhacks (Gau et al. 2021) and the INCF<sup>1</sup> NIDASH task force (see 2.1 for more on NIDASH). I have then joined Mozilla Open Leaders<sup>2</sup>, an international online program mixing training and

<sup>1</sup>[www.incf.org/](http://www.incf.org/)

<sup>2</sup>[foundation.mozilla.org/en/initiatives/mozilla-open-leaders/](https://foundation.mozilla.org/en/initiatives/mozilla-open-leaders/)

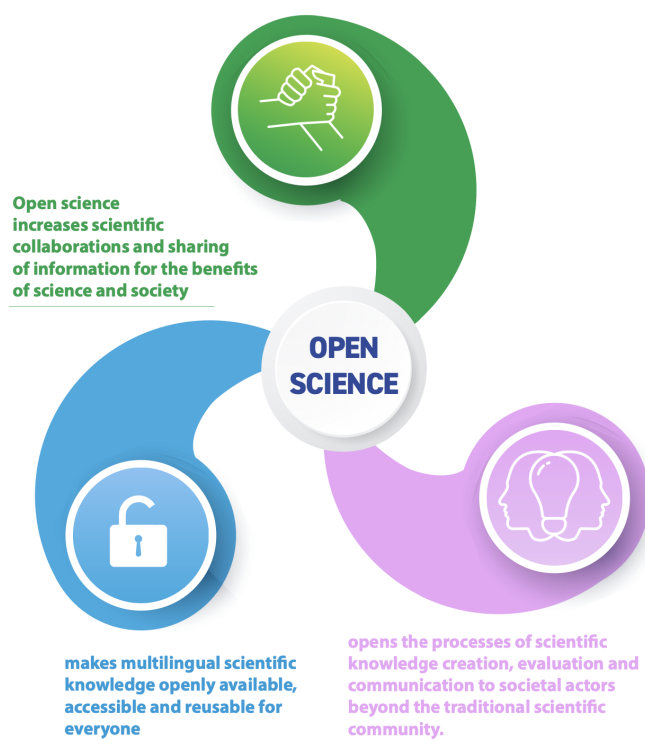


Figure 26: Open Science as defined by UNESCO. Credits: Reproduced from CC BY-SA 3.0 IGO figure, p.8 in (UNESCO 2021)

mentoring. This program has since expanded into domain specific programs (see for example Open Life Science<sup>3</sup>). Since 2018, I have joined the French Open Science Committee (“Comité pour la Science Ouverte” or COSO<sup>4</sup>) in the group on research software, which then transformed into its own college.

As an open science advocate I have been regularly invited to present about open science (see below). Sometimes these activities have also led us to publish scientific contributions (see below). The first contribution focuses on practical solutions for building online conferences that are more inclusive. The second contribution describes the Brainhack hackathons in more details. The last contribution is a guide to open neuroimaging.

### Publications

Gau, Rémi, Stephanie Noble, Katja Heuer, Katherine Bottenhorn, Isil Bilgin, Yu-Fang Yang, Julia Huntenburg, Johanna M.M. Bayer, Richard A.I. Bethlehem, Shawn Rhoads, Christoph Vogelbacher, Valentina Borghesani, Elizabeth Levitis, Hao-Ting Wang, Sofie van den Bossche, Xenia Kobeleva, Jon Hartz Legarreta, Samuel Guay, Selim Melvin Atay, Gael Varoquaux, Dorien Huijser, Malin Sandström, Peer Herholz, Samuel Nastase, Amanpreet Badhwar, Guillaume Dumas, Simon Schwab, Stefano Moia, Michael Dayan, Yasmine Bassil, Paula Brooks, Matteo Mancini,


<sup>3</sup>[openlifesci.org/publications.html](https://openlifesci.org/publications.html)

<sup>4</sup>[www.ouvri.la.science.fr/home/](https://www.ouvri.la.science.fr/home/)

James Shine, David O’connor, Xihe Xie, Davide Poggiali, Patrick Friedrich, Anibal Heinsfeld, Lydia Riedl, Roberto Toro, César Caballero-Gaudes, Anders Eklund, Kelly Garner, Christopher Nolan, Damion Demeter, Fernando Barrios, Junaid Merchant, Elizabeth Mcdevitt, Robert Oostenveld, R. Cameron Craddock, Ariel Rokem, Andrew Doyle, Satrajit Ghosh, Aki Nikolaidis, Olivia Stanley, Eneko Uruñuela, Nasim Anousheh, Aurina Arnatkeviciute, Guillaume Auzias, Dipankar Bachar, Elise Bannier, Ruggero Basanisi, Arshitha Basavaraj, Marco Bedini, Pierre Bellec, R. Austin Benn, Kathryn Berluti, Steffen Bollmann, Saskia Bollmann, Claire Bradley, Jesse Brown, Augusto Buchweitz, Patrick Callahan, Micaela Chan, Bramsh Chandio, Theresa Cheng, Sidhant Chopra, Ai Wern Chung, Thomas Close, Etienne Combrisson, Giorgia Cona, R. Todd Constable, Claire Cury, Kamalaker Dadi, Pablo Damasceno, Samir Das, Fabrizio de Vico Fallani, Krista Destasio, Erin Dickie, Lena Dorfschmidt, Eugene Duff, Elizabeth Dupre, Sarah Dziura, Nathalia Esper, Oscar Esteban, Shreyas Fadnavis, Guillaume Flandin, Jessica Flannery, John Flournoy, Stephanie Forkel, Alexandre Franco, Saampras Ganesan, Siyuan Gao, José García Alanis, Eleftherios Garyfallidis, Tristan Glatard, Enrico Glerean, Javier Gonzalez-Castillo, Cassandra Gould van Praag, Abigail Greene, Geetika Gupta, Catherine Alice Hahn, Yaroslav Halchenko, Daniel Handwerker, Thomas Hartmann, Valérie Hayot-Sasson, Stephan Heunis, Felix Hoffstaedter, Daniela Hohmann, Corey Horien, Horea-Ioan Ioanas, Alexandru Iordan, Chao Jiang, Michael Joseph, Jason Kai, Agah Karakuzu, David Kennedy, Anisha Keshavan, Ali Khan, Gregory Kiar, P. Christiaan Klink, Vincent Koppelmans, Serge Koudoro, Angela Laird, Georg Langs, Marissa Laws, Roxane Licandro, Sook-Lei Liew, Tomislav Lipic, Krisanne Litinas, Daniel Lurie, Désirée Lussier, Christopher Madan, Lea-Theresa Mais, Sina Mansour L, Jose Pedro Manzano-Patron, Dimitra Maoutsa, Matheus Marcon, Daniel Margulies, Giorgio Marinato, Daniele Marinazzo, Christopher Markiewicz, **Camille Maumet**, Felipe Meneguzzi, David Meunier, Michael Milham, Kathryn Mills, Davide Momi, Clara Moreau, Aysha Motala, Iska Moxon-Emre, Thomas E. Nichols, Dylan Nielson, Gustav Nilsson, Lisa Novello, Caroline O’Brien, Emily Olafson, Lindsay Oliver, John Onofrey, Edwina Orchard, Kendra Oudyk, Patrick Park, Mahboobeh Parsapoor, Lorenzo Pasquini, Scott Peltier, Cyril Pernet, Rudolph Pienaar, Pedro Pinheiro-Chagas, Jean-Baptiste Poline, Anqi Qiu, Tiago Quendera, Laura Rice, Joscelin Rocha-Hidalgo, Saige Rutherford, Mathias Scharinger, Dustin Scheinost, Deena Shariq, Thomas Shaw, Viviana Silless, Molly Simmonite, Nikoloz Sirmipilatz, Hayli Spence, Julia Sprenger, Andrija Stajduhar, Martin Szinte, Sylvain Takerkart, Angela Tam, Link Tejavibulya, Michel Thiebaut de Schotten, Ina Thome, Laura Tomaz da Silva, Nicolas Traut, Lucina Uddin, Antonino Vallesi, John Vanmeter, Nandita Vijayakumar, Matteo Visconti Di Oleggio Castello, Jakub Vohryzek, Jakša Vukojević, Kirstie Jane Whitaker, Lucy Whitmore, Steve Wideman, Suzanne Witt, Hua Xie, Ting Xu, Chao-Gan Yan, Fang-Cheng Yeh, B.T. Thomas Yeo, and Xi-Nian Zuo (2021). “Brainhack: Developing a culture of open, inclusive, community-driven neuroscience”. In: *Neuron* 109.11, pp. 1769–1775. DOI: 10.1016/j.neuron.2021.04.001.

Levitis, Elizabeth, Cassandra Gould van Praag, Remi Gau, Stephan Heunis, Elizabeth Dupre, Greg Kiar, Katherine Bottenhorn, Tristan Glatard, Aki Nikolaidis, Kirstie Jane Whitaker, Matteo Mancini, Guiomar Niso, Soroosh Afyouni, Eva Alonso Ortiz, Stefan Appelhoff, Aurina Arnatkeviciute, Melvin Selim Atay, Tibor Auer, Giulia Baracchini, Johanna Margarete Marianne Bayer, Michael J. S. Beauvais, Janine Diane Bijsterbosch, Isil Poyraz Bilgin, Saskia Bollmann, Steffen Bollmann, Rotem Botvinik-Nezer, Molly Bright, Vince Calhoun, Xiao Chen, Sidhant

- Chopra, Hu Chuan-Peng, Thomas Close, Savannah Cookson, Cameron Craddock, Alejandro de La Vega, Benjamin de Leener, Damion Demeter, Paola Di Maio, Erin Dickie, Simon Eickhoff, Oscar Esteban, Karolina Finc, Matteo Frigo, Saampras Ganesan, Melanie Ganz, Kelly Garner, Eduardo Garza-Villarreal, Gabriel Gonzalez-Escamilla, Rohit Goswami, John David Griffiths, Tijn Grootswagers, Samuel Guay, Olivia Guest, Daniel Handwerker, Peer Herholz, Katja Heuer, Dorien Huijser, Vittorio Iacovella, Michael Joseph, Agah Karakuzu, David Keator, Xenia Kobleva, Manoj Kumar, Angie Laird, Linda Larson-Prior, Alexandra Lautarescu, Alberto Lazari, Jon Hartz Legarreta Gorroño, Xue-Ying Li, Jinglei Lv, Sina Mansour, David Meunier, Dustin Moraczewski, Tulika Nandi, Samuel Nastase, Matthias Nau, Stephanie Noble, Martin Norgaard, Johnes Obungoloch, Robert Oostenveld, Edwina Orchard, Ana Luísa Pinho, Russell Poldrack, Anqi Qiu, Pradeep Reddy Raamana, Ariel Rokem, Saige Rutherford, Malvika Sharan, Thomas Shaw, Warda Syeda, Meghan Testerman, Roberto Toro, Sofie Valk, Sofie van den Bossche, Gael Varoquaux, Frantisek Vasa, Michele Veldsman, Jakub Vohryzek, Adina Svenja Wagner, Reubs Walsh, Tonya White, Fu-Te Wong, Xihe Xie, Chao-Gan Yan, Yu-Fang Yang, Yohan Yee, Gaston Zanitti, Ana van Gulick, Eugene Duff, and **Camille Maumet** (2021). “Centering inclusivity in the design of online conferences - An OHBM - Open Science perspective”. In: *GigaScience* 10.8. DOI: 10.1093/gigascience/giab051.
- Niso, Guiomar, Rotem Botvinik-Nezer, Stefan Appelhoff, Alejandro de La Vega, Oscar Esteban, Joset Etzel, Karolina Finc, Melanie Ganz, Remi Gau, Yaroslav Halchenko, Peer Herholz, Agah Karakuzu, David Keator, Christopher Johnson Markiewicz, **Camille Maumet**, Cyril Pernet, Franco Pestilli, Nazek Queder, Tina Schmitt, Weronika Sójka, Adina Svenja Wagner, Kirstie Jane Whitaker, and Jochem Rieger (2022). “Open and reproducible neuroimaging: from study inception to publication”. In: *NeuroImage*. DOI: 10.1016/j.neuroimage.2022.119623.

 **Talks** Follow the link to access the slides on HAL

- “Open science practices for keeping inventions alive after project ends” (2023). In: *ESMRMB 2023, Basel (Switzerland)*. URL: <https://inria.hal.science/hal-04236827>.
- “Open science: A journey from sharing research artefacts to collaborative research” (2022). In: *Rigor and Reproducibility Seminar Series, University of Florida (USA), Online*. URL: <https://inria.hal.science/hal-03716748>.
- “Effort involved in truly FAIR neuroimaging: Towards community-driven research” (2021). In: *INCF Assembly, Online*. URL: <https://inria.hal.science/inserm-03194778>.
- “Building a more collaborative neuroimaging science” (2020). In: *Think Open Rovereto Workshop, Trento (Italy) Online*. URL: <https://inria.hal.science/inserm-02909432>.
- “Sharing more than research papers for transparent & reusable research” (2020). In: *Workshop “Open and reproducible neuroimaging”, Oldenburg University (Germany)*, pp. 1–39. URL: <https://inria.hal.science/inserm-02986992>.
- “Let’s do open neuroimaging sciences” (2019). In: *GEANT Workshop, Marseille (France)*. URL: <https://inria.hal.science/inserm-02146290>.





# Chapter 6

## Conclusions and perspectives

### Overview of the contributions

In this manuscript, I have proposed an overview of my research in neuroimaging reproducibility for the past 10 years, organized in three main axes.

First, we tackled the questions of how to share neuroimaging data (see Chapter 2). I took part in community-driven standardization efforts in order to build harmonized representations of neuroimaging data along with key metadata. Sharing descriptive metadata is crucial in order to enable reuse of shared data as emphasized by the FAIR principles (Wilkinson et al. 2016). Specifically, we introduced NIDM-Results, a common representation for the results of mass-univariate analyses (such as those used in task-fMRI studies) and the Brain Imaging Data Structure (BIDS) which is now ubiquitous in brain imaging.

Second, we focused on analytical variability – i.e. the changes observed in the result of scientific experiments when post-processing pipelines vary – (see Chapter 3). In the field of brain imaging, until recently, this analytical variability was typically ignored, effectively considering that it was negligible compared to other sources of variability (e.g. intra- and inter-subjects, test-retest, measurement error, etc.). I contributed to demonstrate the impact of analytical variability in neuroimaging. An important driver of this work has been to try and provide a better understanding of why analytical variability arises so that we can, in the future, envision solutions.

Third, we investigated neuroimaging data reuse (see Chapter 4). On the statistical side, we tested the validity of meta-analyses approaches for image-based meta-analyses. Then, we investigated how, in practice, using different pipelines can impact our ability to reuse task-fMRI statistic maps in a new analysis. Finally, we proposed an approach to leverage a very large dataset (the NeuroVault database) to build a lower-dimensional representation of task-fMRI statistic maps. This can be used as an initialization for all sort of classification problems.

Finally, since the start of my postdoctoral fellowship I have continuously contributed to open science communities (see Chapter 5). Starting with the question of sharing more than scientific papers (see Chapter 2), continuing into the realm of collaborative science (starting with the Brainhack hackathons) and finally towards more inclusive neuroimaging.

My research is by essence collaborative and this work was only possible by working closely with other scientists and engineers. I seek to find solutions at the intersection between neuroimaging and other fields. Throughout the years I have collaborated with Statisticians, Physicians,

Neuroscientists, Medical doctors from a variety of disciplines, Psychologists, other Computer Scientists and lately with researchers in Privacy and Law.

## Future work and research directions

In the past few years, an important focus of my research has been to understand analytical variability and to try and propose solutions to keep it under control in neuroimaging studies. This was at the center of Chapter 3 but also an important driver for the contributions of Chapter 4. Now that the brain community has come to realize that a better understanding of the variability induced by alternative analytical paths is crucial, the next step is to derive practical solutions.

My vision for the future is that we will be able to guide practitioners in their exploration of the pipeline space and extend on the traditional neuroimaging analysis approach – which typically reports results with respect to a unique analytical pipeline –. In the literature, researchers sometimes demonstrate the robustness of their results or the impact of a set of parameters (e.g. do the results still hold if the age of the participants is added as a covariate? Or if a specific preprocessing step is removed?). The goal would be to extend this approach to provide practitioners with a (limited) subset of analytical pipelines to be tested in order to investigate both the robustness of their analysis and its variability in the pipeline space (see Fig. 27 top panel).

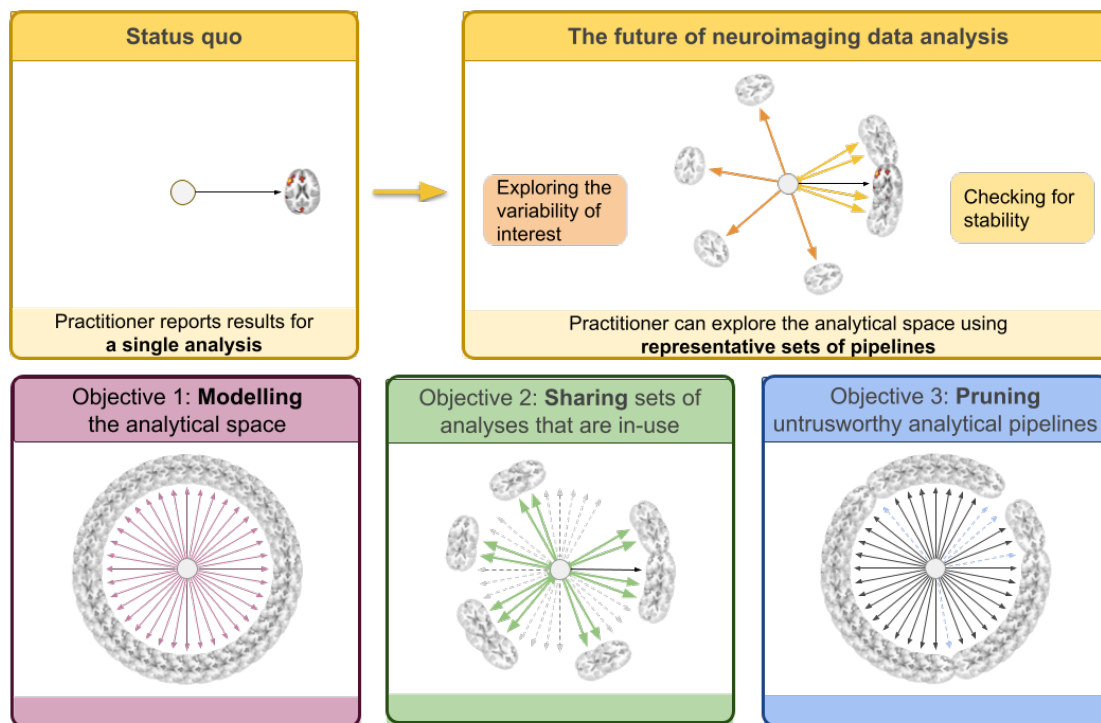


Figure 27: Tackling analytical variability in brain imaging studies

The first studies (including ours) that investigated analytical variability only looked at a tiny proportion of the pipeline space. As a very coarse estimation – considering that an analytical pipeline is made up of 10 steps, with 5 different software packages in 5 versions and 4 algorithms with 10 parameter values at each step – a given brain imaging study has a total of  $10^{30}$  possible

alternative analysis paths. This high order of magnitude makes it impossible to use exhaustive approaches (aka “brute force”) to explore the pipeline space at large.

In order to better understand the analytical space, I propose to tackle 3 objectives: 1/ to model the pipeline space, 2/ to build solutions for sharing pipelines and 3/ to focus on detecting pipeline failures (see Fig. 27 bottom panel).

### **Objective 1: Modelling the pipeline space**

A first objective will be to get a concrete picture of the set of pipelines available to practitioners. This is challenging because typically the description of the analysis is limited to the (free-form) text included in the Methods section of the associated publication. Due to space limitation, those descriptions are often sparse and lack the required level of details needed for computational reproducibility. Recent efforts, such as the COBIDAS recommendations for MRI (Nichols et al. 2017a) and electro/magnetoencephalography (Pernet et al. 2020), have helped by providing guidelines on how to increase the level of details provided in Methods section. In addition, the practice of code sharing is increasing in the community, effectively providing greater transparency as to which analyses are used “in the wild”.

Beyond describing the pipeline associated with a single study, we need a common software-agnostic model representing *\*all\** possible alternatives. I propose to draw inspiration from the study of variants in software engineering and view the multiplicity of analytical pipelines as a software configuration problem. Due to their ubiquitousness, software engineering has been increasingly interested in understanding the behavior of software with a very large number of configurations (also known as software product lines (Apel et al. 2016)). The study of software variability has been focusing on deriving methodologies to “sample, measure and learn” (Pereira et al. 2021) in large configuration spaces: sampling from the configuration space in order to measure and learn characteristics (e.g. compilation time, configuration failure) to get an overview of the overall behavior over the full set of configurations. This idea is at the root of my research project *VICUNA* funded by the French National Research Agency (ANR) young researchers program (JCJC), in which I will co-supervise a PhD student with Prof. Mathieu Acher (expert of software variants) and an engineer as of 2024.

### **Objective 2: Sharing scientific pipelines**

Another important question is to get a better understanding of which pipelines are used by practitioners. This is important first for transparency but also to help understand if specific sub-parts of the pipeline space are more often used by the research community.

In an effort to build a community resource, we are currently developing the *NARPS open pipelines* project (see Section 3.2 for more on NARPS). The goal of this project is to re-implement and share the 70 pipelines proposed by experts in the original NARPS study. This work was initiated by Elodie Germani as part of her Master 2 internship and is now led by Boris Clenet (research engineer) and funded by the Inria exploratory action *GRASP*. Jérémy Lefort-Besnard has also contributed as part of his post-dotoral fellowship funded by the *MIND* project of the Brittany Region. NARPS open pipelines has received contributions at multiple hackathons, see full list at: [https://github.com/Inria-Empenn/narps\\_open\\_pipelines#credits](https://github.com/Inria-Empenn/narps_open_pipelines#credits).

As the community realizes that pipelines are an important roadblock of the scientific process – not only to enable computational reproducibility but also to understand analytical variability – there is an opportunity for pipelines to become a first-class research output. BIDS-Prov by enabling reporting of provenance in neuroimaging (see Section 3.3) was a first step in that direction. In the future, we would like to be able to provide more insights about the pipelines. This is at the root of my participation to the PEPR Santé numérique *ShareFAIR* led by Prof. Sarah Cohen-Boulakia, in which I will co-supervise a PhD student with Dr. Elise Bannier.

### **Objective 3: Providing best practices**

Finally, another research direction is to try and reduce the size of the pipeline space by providing best practices. This could be done for example by identifying analytical pipelines that systematically fail in order to exclude them from the space of possible pipelines. As any piece of software, neuroimaging analytical pipelines will inevitably fail in some instances. In particular, pipeline might systematically (i.e. with all datasets) fail for two main reasons: 1/ Verification failure: when the software behaves differently from what was intended due to an implementation error (or bug) or 2/ Validation failure: when the software includes assumptions that were not explicit and is therefore behaving differently to what experts would expect. Providing solutions to more systematically test neuroimaging software is another open research direction that I would like to pursue in the future.

# My publications

A complete list of my publications in journals and conferences (full-texts only) since my PhD defense.

The names of people I supervised are underlined. In the field of neuroimaging the first author is typically the person who has undertaken the work (primary author) and the last author the supervisor (senior author). \* denotes a joint authorship with equal contributions (either as primary authors or senior authors).

## Preprints

Germani, Elodie, Elisa Fromont, Pierre Maurel, and **Camille Maumet** (2023b). “The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis”. Preprint in preparation.

**Maumet, Camille** and Thomas E. Nichols (2023). “Validity and Accuracy of Neuroimaging Image-Based Meta-Analysis with Z-statistics”. Preprint in preparation.

## Journal Articles

Acar, Freya, **Camille Maumet**, Talia Heuten, Maya Vervoort, Han Bossier, Ruth Seurinck, and Beatrijs Moerkerke (2023). “Improving the Eligibility of Task-Based fMRI Studies for Meta-Analysis: A Review and Reporting Recommendations”. In: *Neuroinformatics*. DOI: 10.1007/s12021-023-09643-5.

Germani, Elodie, Elisa Fromont\*, and **Camille Maumet**\* (2023c). “On the benefits of self-taught learning for brain decoding”. In: *GigaScience* 12, giad029. ISSN: 2047-217X. DOI: 10.1093/gigascience/giad029.

Kabbara, Aya, Nina Forde, **Camille Maumet**\*, and Mahmoud Hassan\* (2023). “Successful reproduction of a large EEG study across software packages”. In: *Neuroimage: Reports* 3.2, p. 100169. DOI: 10.1016/j.ynirp.2023.100169.

Acar, Freya, **Camille Maumet**, Talia Heuten, Maya Vervoort, Han Bossier, Ruth Seurinck, and Beatrijs Moerkerke (2022). “Review Paper: Reporting Practices for Task fMRI Studies”. In: *Neuroinformatics*. DOI: 10.1007/s12021-022-09606-2. URL: <https://www.hal.inserm.fr/inserm-03800029/document>.

- Niso, Guiomar, Rotem Botvinik-Nezer, Stefan Appelhoff, Alejandro de La Vega, Oscar Esteban, Joset Etzel, Karolina Finc, Melanie Ganz, Remi Gau, Yaroslav Halchenko, Peer Herholz, Agah Karakuzu, David Keator, Christopher Johnson Markiewicz, **Camille Maumet**, Cyril Pernet, Franco Pestilli, Nazek Queder, Tina Schmitt, Weronika Sójka, Adina Svenja Wagner, Kirstie Jane Whitaker, and Jochem Rieger (2022). “Open and reproducible neuroimaging: from study inception to publication”. In: *NeuroImage*. DOI: 10.1016/j.neuroimage.2022.119623.
- Bowring, Alexander, Thomas Nichols\*, and **Camille Maumet\*** (2021b). “Isolating the Sources of Pipeline-Variability in Group-Level Task-fMRI results”. In: *Human Brain Mapping*. DOI: 10.1002/hbm.25713.
- Clement, Patricia, Thomas Booth, Fran Borovečki, Kyrre E. Emblem, Patrícia Figueiredo, Lydiane Hirschler, Radim Jančálek, Vera C. Keil, **Camille Maumet**, Yelda Özsunar, Cyril Pernet, Jan Petr, Joana Pinto, Marion Smits, and Esther A. H. Warnert (2021). “GliMR: Cross-Border Collaborations to Promote Advanced MRI Biomarkers for Glioma”. In: *Journal of Medical and Biological Engineering* 41.2, pp. 115–125. DOI: 10.1007/s40846-020-00582-z.
- Gau, Rémi, Stephanie Noble, Katja Heuer, Katherine Bottenhorn, Isil Bilgin, Yu-Fang Yang, Julia Huntenburg, Johanna M.M. Bayer, Richard A.I. Bethlehem, Shawn Rhoads, Christoph Vogelbacher, Valentina Borghesani, Elizabeth Levitis, Hao-Ting Wang, Sofie van den Bossche, Xenia Kobeleva, Jon Hartz Legarreta, Samuel Guay, Selim Melvin Atay, Gael Varoquaux, Dorien Huijser, Malin Sandström, Peer Herholz, Samuel Nastase, Amanpreet Badhwar, Guillaume Dumas, Simon Schwab, Stefano Moia, Michael Dayan, Yasmine Bassil, Paula Brooks, Matteo Mancini, James Shine, David O’connor, Xihe Xie, Davide Poggiali, Patrick Friedrich, Anibal Heinsfeld, Lydia Riedl, Roberto Toro, César Caballero-Gaudes, Anders Eklund, Kelly Garner, Christopher Nolan, Damion Demeter, Fernando Barrios, Junaid Merchant, Elizabeth Mcdevitt, Robert Oostenfeld, R. Cameron Craddock, Ariel Rokem, Andrew Doyle, Satrajit Ghosh, Aki Nikolaidis, Olivia Stanley, Eneko Uruñuela, Nasim Anousheh, Aurina Arnatkeviciute, Guillaume Auzias, Dipankar Bachar, Elise Bannier, Ruggero Basanisi, Arshitha Basavaraj, Marco Bedini, Pierre Bellec, R. Austin Benn, Kathryn Berluti, Steffen Bollmann, Saskia Bollmann, Claire Bradley, Jesse Brown, Augusto Buchweitz, Patrick Callahan, Micaela Chan, Bramsh Chandio, Theresa Cheng, Sidhant Chopra, Ai Wern Chung, Thomas Close, Etienne Combrisson, Giorgia Cona, R. Todd Constable, Claire Cury, Kamalaker Dadi, Pablo Damasceno, Samir Das, Fabrizio de Vico Fallani, Krista Destasio, Erin Dickie, Lena Dorfschmidt, Eugene Duff, Elizabeth Dupre, Sarah Dziura, Nathalia Esper, Oscar Esteban, Shreyas Fadnavis, Guillaume Flandin, Jessica Flannery, John Flournoy, Stephanie Forkel, Alexandre Franco, Saampras Ganesan, Siyuan Gao, José García Alanis, Eleftherios Garyfallidis, Tristan Glatard, Enrico Glerean, Javier Gonzalez-Castillo, Cassandra Gould van Praag, Abigail Greene, Geetika Gupta, Catherine Alice Hahn, Yaroslav Halchenko, Daniel Handwerker, Thomas Hartmann, Valérie Hayot-Sasson, Stephan Heunis, Felix Hoffstaedter, Daniela Hohmann, Corey Horien, Horea-Ioan Ioanas, Alexandru Iordan, Chao Jiang, Michael Joseph, Jason Kai, Agah Karakuzu, David Kennedy, Anisha Keshavan, Ali Khan, Gregory Kiar, P. Christiaan Klink, Vincent Koppelmans, Serge Koudoro, Angela Laird, Georg Langs, Marissa Laws, Roxane Licandro, Sook-Lei Liew, Tomislav Lipic, Krisanne Litinas, Daniel Lurie, Désirée Lussier, Christopher Madan, Lea-Theresa Mais, Sina Mansour L, Jose Pedro Manzano-Patron, Dimitra Maoutsa, Matheus Marcon, Daniel Margulies, Giorgio Marinato, Daniele Marinazzo, Christopher Markiewicz, **Camille Maumet**, Felipe Meneguzzi,



- David Meunier, Michael Milham, Kathryn Mills, Davide Momi, Clara Moreau, Aysha Motala, Iska Moxon-Emre, Thomas E. Nichols, Dylan Nielson, Gustav Nilsson, Lisa Novello, Caroline O'Brien, Emily Olafson, Lindsay Oliver, John Onofrey, Edwina Orchard, Kendra Oudyk, Patrick Park, Mahboobeh Parsapoor, Lorenzo Pasquini, Scott Peltier, Cyril Pernet, Rudolph Pienaar, Pedro Pinheiro-Chagas, Jean-Baptiste Poline, Anqi Qiu, Tiago Quendera, Laura Rice, Joscelin Rocha-Hidalgo, Saige Rutherford, Mathias Scharinger, Dustin Scheinost, Deena Shariq, Thomas Shaw, Viviana Silless, Molly Simmonite, Nikoloz Sirmopilatze, Hayli Spence, Julia Sprenger, Anđrija Stajduhar, Martin Szinte, Sylvain Takerkart, Angela Tam, Link Tejavibulya, Michel Thiebaut de Schotten, Ina Thome, Laura Tomaz da Silva, Nicolas Traut, Lucina Uddin, Antonino Vallesi, John Vanmeter, Nandita Vijayakumar, Matteo Visconti Di Oleggio Castello, Jakub Vohryzek, Jakša Vukojević, Kirstie Jane Whitaker, Lucy Whitmore, Steve Wideman, Suzanne Witt, Hua Xie, Ting Xu, Chao-Gan Yan, Fang-Cheng Yeh, B.T. Thomas Yeo, and Xi-Nian Zuo (2021). "Brainhack: Developing a culture of open, inclusive, community-driven neuroscience". In: *Neuron* 109.11, pp. 1769–1775. DOI: 10.1016/j.neuron.2021.04.001.
- Levitis, Elizabeth, Cassandra Gould van Praag, Remi Gau, Stephan Heunis, Elizabeth Dupre, Greg Kiar, Katherine Bottenhorn, Tristan Glatard, Aki Nikolaidis, Kirstie Jane Whitaker, Matteo Mancini, Guiomar Niso, Soroosh Afyouni, Eva Alonso Ortiz, Stefan Appelhoff, Aurina Arnatkeviciute, Melvin Selim Atay, Tibor Auer, Giulia Baracchini, Johanna Margarete Marianne Bayer, Michael J. S. Beauvais, Janine Diane Bijsterbosch, Isil Poyraz Bilgin, Saskia Bollmann, Steffen Bollmann, Rotem Botvinik-Nezer, Molly Bright, Vince Calhoun, Xiao Chen, Sidhant Chopra, Hu Chuan-Peng, Thomas Close, Savannah Cookson, Cameron Craddock, Alejandro de La Vega, Benjamin de Leener, Damion Demeter, Paola Di Maio, Erin Dickie, Simon Eickhoff, Oscar Esteban, Karolina Finc, Matteo Frigo, Saampras Ganesan, Melanie Ganz, Kelly Garner, Eduardo Garza-Villarreal, Gabriel Gonzalez-Escamilla, Rohit Goswami, John David Griffiths, Tijn Grootswagers, Samuel Guay, Olivia Guest, Daniel Handwerker, Peer Herholz, Katja Heuer, Dorien Huijser, Vittorio Iacovella, Michael Joseph, Agah Karakuzu, David Keator, Xenia Kobleleva, Manoj Kumar, Angie Laird, Linda Larson-Prior, Alexandra Lautarescu, Alberto Lazari, Jon Hartz Legarreta Gorroño, Xue-Ying Li, Jinglei Lv, Sina Mansour, David Meunier, Dustin Moraczewski, Tulika Nandi, Samuel Nastase, Matthias Nau, Stephanie Noble, Martin Norgaard, Johnes Obungoloch, Robert Oostenveld, Edwina Orchard, Ana Luísa Pinho, Russell Poldrack, Anqi Qiu, Pradeep Reddy Raamana, Ariel Rokem, Saige Rutherford, Malvika Sharan, Thomas Shaw, Warda Syeda, Meghan Testerman, Roberto Toro, Sofie Valk, Sofie van den Bossche, Gael Varoquaux, Frantisek Vasa, Michele Veldsman, Jakub Vohryzek, Adina Svenja Wagner, Reubs Walsh, Tonya White, Fu-Te Wong, Xihe Xie, Chao-Gan Yan, Yu-Fang Yang, Yohan Yee, Gaston Zanitti, Ana van Gulick, Eugene Duff, and **Camille Maumet** (2021). "Centering inclusivity in the design of online conferences - An OHBM - Open Science perspective". In: *GigaScience* 10.8. DOI: 10.1093/gigascience/giab051.
- Open Brain Consent working group (2021). "The Open Brain Consent : Informing research participants and obtaining consent to share brain imaging data". In: *Human Brain Mapping* 42.7, pp. 1945–1951. DOI: 10.1002/hbm.25351.
- Botvinik-Nezer, Rotem, Felix Holzmeister, Colin F Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A Mumford, R. Alison Adcock, Paolo Avesani, Blazej M Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari,

- Nadège Bault, Derek Beaton, Julia Beitner, Roland G Benoit, Ruud M.W.J. Berkers, Jamil P Bhanji, Bharat B Biswal, Sebastian Bobadilla-Suarez, Tiago Bortolini, Katherine L Bottenhorn, Alexander Bowring, Senne Braem, Hayley R Brooks, Emily G Brudner, Cristian B Calderon, Julia A Camilleri, Jaime J Castellon, Luca Cecchetti, Edna C Cieslik, Zachary J Cole, Olivier Collignon, Robert W Cox, William A Cunningham, Stefan Czoschke, Kamalaker Dadi, Charles P Davis, Alberto De Luca, Mauricio R Delgado, Lysia Demetriou, Jeffrey B Dennison, Xin Di, Erin W Dickie, Ekaterina Dobryakova, Claire L Donnat, Juergen Dukart, Niall W Duncan, Joke Durnez, Amr Eed, Simon B. Eickhoff, Andrew Erhart, Laura Fontanesi, G. Matthew Fricke, Shiguang Fu, Adriana Galván, Remi Gau, Sarah Genon, Tristan Glatard, Enrico Glerean, Jelle J Goeman, Sergej A.E. Golowin, Carlos González-García, Krzysztof J Gorgolewski, Cheryl L Grady, Mikella A Green, João F Guassi Moreira, Olivia Guest, Shabnam Hakimi, J. Paul Hamilton, Roeland Hancock, Giacomo Handjaras, Bronson B Harry, Colin Hawco, Peer Herholz, Gabrielle Herman, Stephan Heunis, Felix Hoffstaedter, Jeremy Hogeveen, Susan Holmes, Chuan-Peng Hu, Scott A Huettel, Matthew E Hughes, Vittorio Iacovella, Alexandru D Iordan, Peder M Isager, Ayse I Isik, Andrew Jahn, Matthew R Johnson, Tom Johnstone, Michael J.E. Joseph, Anthony C Juliano, Joseph W Kable, Michalis Kassinopoulos, Cemal Koba, Xiang-Zhen Kong, Timothy R Kosciak, Nuri Erkut Kucukboyaci, Brice A Kuhl, Sebastian Kupek, Angela R Laird, Claus Lamm, Robert Langner, Nina Lauharatanahirun, Hongmi Lee, Sangil Lee, Alexander Leemans, Andrea Leo, Elise Lesage, Flora Li, Monica Y.C. Li, Phui Cheng Lim, Evan N Lintz, Schuyler W Liphardt, Annabel B Losecaat Vermeer, Bradley C Love, Michael L Mack, Norberto Malpica, Theo Marins, **Camille Maumet**, Kelsey Mcdonald, Joseph T Mcguire, Helena Melero, Adriana S Méndez Leal, Benjamin Meyer, Kristin N Meyer, Glad Mihai, Georgios D. Mitsis, Jorge Moll, Dylan M Nielson, Gustav Nilsson, Michael P Notter, Emanuele Olivetti, Adrian I Onicas, Paolo Papale, Kaustubh R Patil, Jonathan E Peelle, Alexandre Pérez, Doris Pischcedda, Jean-Baptiste Poline, Yanina Prystauka, Shruti Ray, Patricia A Reuter-Lorenz, Richard C Reynolds, Emiliano Ricciardi, Jenny R Rieck, Anais M Rodriguez-Thompson, Anthony Romyn, Taylor Salo, Gregory R Samanez-Larkin, Emilio Sanz-Morales, Margaret L Schlichting, Douglas H Schultz, Qiang Shen, Margaret A Sheridan, Jennifer A Silvers, Kenny Skagerlund, Alec Smith, David V Smith, Peter Sokol-Hessner, Simon R Steinkamp, Sarah M Tashjian, Bertrand Thirion, John N Thorp, Gustav Tinghög, Loreen Tisdall, Steven H Tompson, Claudio Toro-Serey, Juan Jesus Torre Tresols, Leonardo Tozzi, Vuong Truong, Luca Turella, Anna E van 't Veer, Tom Verguts, Jean M Vettel, Sagana Vijayarajah, Khoi Vo, Matthew B Wall, Wouter D Weeda, Susanne Weis, David J White, David Wisniewski, Alba Xifra-Porxas, Emily A Yearling, Sangsuk Yoon, Rui Yuan, Kenneth S.L. Yuen, Lei Zhang, Xu Zhang, Joshua E Zosky, Thomas E. Nichols, Russell A Poldrack, and Tom Schonberg (2020). “Variability in the analysis of a single neuroimaging dataset by many teams”. In: *Nature* 582.7810, pp. 84–88. DOI: 10.1038/s41586-020-2314-9. URL: <https://www.biorxiv.org/content/10.1101/843193>.
- Bowring, Alexander, **Camille Maumet**\*, and Thomas E. Nichols\* (2019). “Exploring the Impact of Analysis Software on Task fMRI Results”. In: *Human Brain Mapping* 40.11, pp. 3362–3384. DOI: 10.1002/hbm.24603.
- Gutierrez-Barragan, Felipe, Vamsi K. Ithapu, Chris Hinrichs, **Camille Maumet**, Sterling C. Johnson, Thomas E. Nichols, and Vikas Singh (2017). “Accelerating permutation testing in

- voxel-wise analysis through subspace tracking: A new plugin for SnPM”. In: *NeuroImage*. DOI: 10.1016/j.neuroimage.2017.07.025.
- Gorgolewski, Krzysztof J., Tibor Auer, Vince D. Calhoun, Cameron R. Craddock, Samir Das, Eugene P. Duff, Guillaume Flandin, Satrajit S. Ghosh, Tristan Glatard, Yaroslav O. Halchenko, Daniel A. Handwerker, Michael Hanke, David Keator, Xiangrui Li, Zachary Michael, **Camille Maumet**, Nolan B. Nichols, Thomas E. Nichols, John Pellman, Jean-Baptiste Poline, Ariel Rokem, Gunnar Schaefer, Vanessa Sochat, William Triplett, Jessica A. Turner, Gaël Varoquaux, and Russell A. Poldrack (2016b). “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments”. In: *Scientific Data* 3, p. 160044. DOI: 10.1038/sdata.2016.44.
- Maumet, Camille**, Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, Tristan Glatard, Krzysztof J Gorgolewski, Karl G Helmer, Mark Jenkinson, David B Keator, B. Nolan Nichols, Jean-Baptiste Poline, Richard Reynolds, Vanessa Sochat, Jessica Turner, and Thomas E. Nichols (2016a). “Sharing brain mapping statistical results with the neuroimaging data model”. In: *Scientific Data* 3. DOI: 10.1038/sdata.2016.102.
- Maumet, Camille**, Pierre Maurel, Jean-Christophe Ferré, and Christian Barillot (2016b). “An a contrario approach for the detection of patient-specific brain perfusion abnormalities with arterial spin labelling”. In: *NeuroImage* 134. DOI: 10.1016/j.neuroimage.2016.03.054. URL: <https://www.hal.inserm.fr/inserm-01291748/document>.
- Pauli, Ruth**, Alexander Bowring, Richard Reynolds, Gang Chen, Thomas E. Nichols, and **Camille Maumet** (2016). “Exploring fMRI Results Space: 31 Variants of an fMRI Analysis in AFNI, FSL, and SPM”. In: *Frontiers in Neuroinformatics* 10. DOI: 10.3389/fninf.2016.00024.
- Scarpazza, Cristina, Thomas E. Nichols, Donato Seramondi, **Camille Maumet**, Giuseppe Sartori, and Andrea Mechelli (2016b). “When the Single Matters more than the Group (II): Addressing the Problem of High False Positive Rates in Single Case Voxel Based Morphometry Using Non-parametric Statistics”. In: *Frontiers in Neuroscience* 10.6. DOI: 10.3389/fnins.2016.00006.
- Gorgolewski, Krzysztof J., Gaël Varoquaux, Gabriel Rivera, Yannick Schwartz, Vanessa V. Sochat, Satrajit S. Ghosh, **Camille Maumet**, Thomas E. Nichols, Jean-Baptiste Poline, Tal Yarkoni, Daniel S. Margulies, and Russell A. Poldrack (2015c). “NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain”. In: *NeuroImage* 124.Pt B, pp. 1242–1244. DOI: 10.1016/j.neuroimage.2015.04.016. URL: <https://www.hal.inserm.fr/inserm-01148969/document>.
- Gorgolewski, Krzysztof J., Gaël Varoquaux, Gabriel Rivera, Yannick Schwarz, Satrajit S. Ghosh, **Camille Maumet**, Vanessa V. Sochat, Thomas E. Nichols, Russell A. Poldrack, Jean-Baptiste Poline, Tal Yarkoni, and Daniel S. Margulies (2015d). “NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain”. In: *Frontiers in Neuroinformatics* 9.8. DOI: 10.3389/fninf.2015.00008.
- Maumet, Camille**, Pierre Maurel, Jean-Christophe Ferré, and Christian Barillot (2014b). “Robust estimation of the cerebral blood flow in arterial spin labelling”. In: *Magnetic Resonance Imaging* 32.5, pp. 497–504. DOI: 10.1016/j.mri.2014.01.016. URL: <https://www.hal.inserm.fr/inserm-00942814/document>.

## Conference Articles

Rolland, Xavier, Pierre Maurel, and **Camille Maumet** (2022). “Towards efficient fmri data re-use: can we run between-group analyses with datasets processed differently with spm ?” In: *IEEE International Symposium on Biomedical Imaging (ISBI)*. Calcutta, India: IEEE, pp. 1–4. URL: <https://hal.science/hal-03607384>.

Cao, Fang, Olivier Commowick, **Camille Maumet**, and Christian Barillot (2014). “Quantitative Relaxation Templates for the Human Brain at 3T”. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*. China, pp. 57–60. URL: <https://www.hal.inserm.fr/inserm-00950843>.

# Bibliography

- Acar, Freya, Camille Maumet, Talia Heuten, Maya Vervoort, Han Bossier, Ruth Seurinck, and Beatrijs Moerkerke (2022). “Review Paper: Reporting Practices for Task fMRI Studies”. In: *Neuroinformatics*. DOI: 10.1007/s12021-022-09606-2. URL: <https://www.hal.inserm.fr/inserm-03800029/document>.
- Acher, Mathieu (2022). “Reproducible Science and Deep Software Variability”. In: *VaMoS 2022 - 16th International Working Conference on Variability Modelling of Software-Intensive Systems*. Florence, Italy, pp. 1–2. URL: <https://hal.inria.fr/hal-03528889>.
- Acel, Balazs, Barnabas Szaszi, Gustav Nilsson, Olmo van den Akker, Casper Albers, Marcel A. L. M. van Assen, Jojanneke A. Bastiaansen, Daniel Jacob Benjamin, Udo Boehm, Rotem Botvinik-Nezer, Laura Bringmann, Niko Busch, Emmanuel Caruyer, Andrea M. Cataldo, Nelson Cowan, Andrew Delios, Noah van Dongen, Chris Donkin, Johnny van Doorn, Anna Dreber Almenberg, Gilles Dutilh, Gary F. Egan, Morton Ann Gernsbacher, Rink Hoekstra, Sabine Hoffmann, Felix Holzmeister, Magnus Johannesson, Kai Jonas, Alexander Kindel, Michael Kirchler, Yoram Kevin Kunkels, D. Stephen Lindsay, Jan-Francois Mangin, Dora Matzke, Marcus Munafo, Ben Newell, Brian A. Nosek, Russell Poldrack, Don van Ravenzwaaij, Jörg Rieskamp, Matthew Salganik, Alexandra Sarafoglou, Tom Schonberg, Martin Schweinsberg, David Shanks, Raphael Silberzahn, Daniel J. Simons, Bobbie Spellman, Jeffrey Starns, Samuel St-Jean, Eric Luis Uhlmann, Jelte Wicherts, and Eric-Jan Wagenmakers (2021). *Guidance for conducting and reporting multi-analyst studies*. Tech. rep. type: article. MetaArXiv. DOI: 10.31222/osf.io/5ecnh. URL: <https://osf.io/preprints/metaarxiv/5ecnh/> (visited on 07/09/2021).
- Allen, Emily J, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, et al. (2022). “A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence”. In: *Nature neuroscience* 25.1, pp. 116–126.
- Apel, Sven, Don Batory, Christian Kästner, and Gunter Saake (2016). *Feature-oriented software product lines*. Springer.
- Barba, Lorena A. (2018). “Terminologies for Reproducible Research”. en. In: *arXiv:1802.03311 [cs]*. arXiv: 1802.03311. URL: <http://arxiv.org/abs/1802.03311> (visited on 03/28/2022).
- Barr, Earl T., Mark Harman, Phil McMinn, Muzammil Shahbaz, and Shin Yoo (2015). “The Oracle Problem in Software Testing: A Survey”. In: *IEEE Transactions on Software Engineering* 41.5. Conference Name: IEEE Transactions on Software Engineering, pp. 507–525. ISSN: 1939-3520. DOI: 10.1109/TSE.2014.2372785.

- Begley, C. Glenn and Lee M. Ellis (2012). “Raise standards for preclinical cancer research”. en. In: *Nature* 483.7391, pp. 531–533. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/483531a. URL: <http://www.nature.com/articles/483531a> (visited on 10/14/2022).
- Biswal, Bharat B., Maarten Mennes, Xi-Nian Zuo, Suril Gohel, Clare Kelly, Steve M. Smith, Christian F. Beckmann, Jonathan S. Adelstein, Randy L. Buckner, Stan Colcombe, Anne-Marie Dagonowski, Monique Ernst, Damien Fair, Michelle Hampson, Matthew J. Hoptman, James S. Hyde, Vesa J. Kiviniemi, Rolf Kötter, Shi-Jiang Li, Ching-Po Lin, Mark J. Lowe, Clare Mackay, David J. Madden, Kristoffer H. Madsen, Daniel S. Margulies, Helen S. Mayberg, Katie McMahon, Christopher S. Monk, Stewart H. Mostofsky, Bonnie J. Nagel, James J. Pekar, Scott J. Peltier, Steven E. Petersen, Valentin Riedl, Serge A. R. B. Rombouts, Bart Rypma, Bradley L. Schlaggar, Sein Schmidt, Rachael D. Seidler, Greg J. Siegle, Christian Sorg, Gao-Jun Teng, Juha Veijola, Arno Villringer, Martin Walter, Lihong Wang, Xu-Chu Weng, Susan Whitfield-Gabrieli, Peter Williamson, Christian Windischberger, Yu-Feng Zang, Hong-Ying Zhang, F. Xavier Castellanos, and Michael P. Milham (2010). “Toward discovery science of human brain function”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.10, pp. 4734–4739. ISSN: 1091-6490. DOI: 10.1073/pnas.0911855107.
- Bloom, Paul Alexander, Michelle VanTieghem, Laurel Gabard-Durnam, Dylan G. Gee, Jessica Flannery, Christina Caldera, Bonnie Goff, Eva H. Telzer, Kathryn L. Humphreys, Dominic S. Fareri, Mor Shapiro, Sameah Algharazi, Niall Bolger, Mariam Aly, and Nim Tottenham (2022). “Age-related change in task-evoked amygdala—prefrontal circuitry: A multiverse approach with an accelerated longitudinal cohort aged 4–22 years”. en. In: *Human Brain Mapping* 43.10. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25847>, pp. 3221–3244. ISSN: 1097-0193. DOI: 10.1002/hbm.25847. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25847> (visited on 10/14/2022).
- Botvinik-Nezer, Rotem, Felix Holzmeister, Colin F. Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A. Mumford, R. Alison Adcock, Paolo Avesani, Blazej M. Baczowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari, Nadège Bault, Derek Beaton, Julia Beitner, Roland G. Benoit, Ruud M. W. J. Berkers, Jamil P. Bhanji, Bharat B. Biswal, Sebastian Bobadilla-Suarez, Tiago Bortolini, Katherine L. Bottenhorn, Alexander Bowring, Senne Braem, Hayley R. Brooks, Emily G. Brudner, Cristian B. Calderon, Julia A. Camilleri, Jaime J. Castellon, Luca Cecchetti, Edna C. Cieslik, Zachary J. Cole, Olivier Collignon, Robert W. Cox, William A. Cunningham, Stefan Czoschke, Kamalaker Dadi, Charles P. Davis, Alberto De Luca, Mauricio R. Delgado, Lysia Demetriou, Jeffrey B. Dennison, Xin Di, Erin W. Dickie, Ekaterina Dobryakova, Claire L. Donnat, Juergen Dukart, Niall W. Duncan, Joke Durnez, Amr Eed, Simon B. Eickhoff, Andrew Erhart, Laura Fontanesi, G. Matthew Fricke, Shiguang Fu, Adriana Galván, Remi Gau, Sarah Genon, Tristan Glatard, Enrico Glerean, Jelle J. Goeman, Sergej A. E. Golowin, Carlos González-García, Krzysztof J. Gorgolewski, Cheryl L. Grady, Mikella A. Green, João F. Guassi Moreira, Olivia Guest, Shabnam Hakimi, J. Paul Hamilton, Roeland Hancock, Giacomo Handjaras, Bronson B. Harry, Colin Hawco, Peer Herholz, Gabrielle Herman, Stephan Heunis, Felix Hoffstaedter, Jeremy Hogeveen, Susan Holmes, Chuan-Peng Hu, Scott A. Huettel, Matthew E. Hughes, Vittorio Iacovella, Alexandru D. Iordan, Peder M. Isager, Ayse I. Isik, Andrew Jahn, Matthew R. Johnson, Tom Johnstone, Michael J. E. Joseph, Anthony C. Juliano, Joseph W. Kable, Michalis Kassinopoulos, Cemal Koba, Xiang-Zhen Kong, Timothy R.

- Koscik, Nuri Erkut Kucukboyaci, Brice A. Kuhl, Sebastian Kupek, Angela R. Laird, Claus Lamm, Robert Langner, Nina Lauharatanahirun, Hongmi Lee, Sangil Lee, Alexander Leemans, Andrea Leo, Elise Lesage, Flora Li, Monica Y. C. Li, Phui Cheng Lim, Evan N. Lintz, Schuyler W. Liphardt, Annabel B. Losecaat Vermeer, Bradley C. Love, Michael L. Mack, Norberto Malpica, Theo Marins, Camille Maumet, Kelsey McDonald, Joseph T. McGuire, Helena Melero, Adriana S. Méndez Leal, Benjamin Meyer, Kristin N. Meyer, Glad Mihai, Georgios D. Mitsis, Jorge Moll, Dylan M. Nielson, Gustav Nilsson, Michael P. Notter, Emanuele Olivetti, Adrian I. Onicas, Paolo Papale, Kaustubh R. Patil, Jonathan E. Peelle, Alexandre Pérez, Doris Pischcedda, Jean-Baptiste Poline, Yanina Prystauka, Shruti Ray, Patricia A. Reuter-Lorenz, Richard C. Reynolds, Emiliano Ricciardi, Jenny R. Rieck, Anais M. Rodriguez-Thompson, Anthony Romyn, Taylor Salo, Gregory R. Samanez-Larkin, Emilio Sanz-Morales, Margaret L. Schlichting, Douglas H. Schultz, Qiang Shen, Margaret A. Sheridan, Jennifer A. Silvers, Kenny Skagerlund, Alec Smith, David V. Smith, Peter Sokol-Hessner, Simon R. Steinkamp, Sarah M. Tashjian, Bertrand Thirion, John N. Thorp, Gustav Tinghög, Loreen Tisdall, Steven H. Tompson, Claudio Toro-Serey, Juan Jesus Torre Tresols, Leonardo Tozzi, Vuong Truong, Luca Turella, Anna E. van 't Veer, Tom Verguts, Jean M. Vettel, Sagana Vijayarajah, Khoi Vo, Matthew B. Wall, Wouter D. Weeda, Susanne Weis, David J. White, David Wisniewski, Alba Xifra-Porxas, Emily A. Yearling, Sangsuk Yoon, Rui Yuan, Kenneth S. L. Yuen, Lei Zhang, Xu Zhang, Joshua E. Zosky, Thomas E. Nichols, Russell A. Poldrack, and Tom Schonberg (2020). “Variability in the analysis of a single neuroimaging dataset by many teams”. en. In: *Nature* 582.7810. Number: 7810 Publisher: Nature Publishing Group, pp. 84–88. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2314-9. URL: <https://www.nature.com/articles/s41586-020-2314-9> (visited on 09/15/2020).
- Botvinik-Nezer, Rotem, Roni Iwanir, Felix Holzmeister, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Anna Dreber, Colin F Camerer, Russell A Poldrack, and Tom Schonberg (2019). “fMRI data of mixed gambles from the Neuroimaging Analysis Replication and Prediction Study”. In: *Scientific data* 6.1, p. 106.
- Bowring, Alexander, Camille Maumet, and Thomas E. Nichols (2019). “Exploring the impact of analysis software on task fMRI results”. en. In: *Human Brain Mapping* 40.11. Number: 11, pp. 3362–3384. ISSN: 1065-9471, 1097-0193. DOI: 10.1002/hbm.24603. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24603> (visited on 01/19/2021).
- Bowring, Alexander, Thomas E. Nichols, and Camille Maumet (2021). “Isolating the sources of pipeline-variability in group-level task-fMRI results”. en. In: *Human Brain Mapping* 43.3. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25713>, pp. 1112–1128. ISSN: 1097-0193. DOI: 10.1002/hbm.25713. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25713> (visited on 06/07/2022).
- Brakewood, Beth and Russell A. Poldrack (2013). “The ethics of secondary data analysis: considering the application of Belmont principles to the sharing of neuroimaging data”. eng. In: *NeuroImage* 82, pp. 671–676. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2013.02.040.
- Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, Muna Adem, Jule Adriaans, Amalia Alvarez-Benjumea, Henrik Kenneth Andersen, Daniel Auer, Flavio Azevedo, Oke Bahnsen, Dave Balzer, Gerrit Bauer, Paul C. Bauer, Markus Baumann, Sharon Baute, Verena Benoit, Julian Bernauer, Carl Berning, Anna Berthold, Felix Bethke, Thomas Biegert, Katharina Blinzler, Johannes Blumenberg, Licia Bobzien, Andrea Bohman, Thijs Bol, Amie Bostic, Zuzanna Brzozowska,

- Katharina Burgdorf, Kaspar Burger, Kathrin Busch, Juan Carlos Castillo, Pablo Christmann, Roxanne Connelly, Christian S. Czymara, Elena Damian, Alejandro Ecker, Achim Edelmann, Maureen A. Eger, Simon Ellerbrock, Anna Forke, Andrea Forster, Chris Gaasendam, Konstantin Gavras, Vernon Gayle, Theresa Gessler, Timo Gnambs, Amélie Godefroidt, Max Grömping, Martin Groß, Stefan Gruber, Tobias Gummer, Andreas Hadjar, Jan Paul Heisig, Sebastian Hellmeier, Stefanie Heyne, Magdalena Hirsch, Mikael Hjerm, Oshrat Hochman, Andreas Hövermann, Sophia Hunger, Christian Hunkler, Nora Huth, Zsofia Ignacz, Laura Jacobs, Jannes Jacobsen, Bastian Jaeger, Sebastian Jungkunz, Nils Jungmann, Mathias Kauff, Manuel Kleinert, Julia Klinger, Jan-Philipp Kolb, Marta Kołczyńska, John Seungmin Kuk, Katharina Kunißen, Dafina Kurti Sinatra, Alexander Greinert, Philipp M. Lersch, Lea-Maria Löbel, Philipp Lutscher, Matthias Mader, Joan Madia, Natalia Malancu, Luis Maldonado, Helge Marahrens, Nicole Martin, Paul Martinez, Jochen Mayerl, Oscar Jose Mayorga, Patricia A. McManus, Kyle Wagner, Cecil Meeusen, Daniel Meierrieks, Jonathan Mellon, Friedolin Merhout, Samuel Merk, Daniel Meyer, Leticia Micheli, Jonathan J. B. Mijs, Cristóbal Moya, Marcel Neunhoffer, Daniel Nüst, Olav Nygård, Fabian Ochsenfeld, Gunnar Otte, Anna Pechenkina, Christopher Prosser, Louis Raes, Kevin Ralston, Miguel Ramos, Arne Roets, Jonathan Rogers, Guido Ropers, Robin Samuel, Gregor Sand, Ariela Schachter, Merlin Schaeffer, David Schieferdecker, Elmar Schlueter, Katja M. Schmidt, Regine Schmidt, Alexander Schmidt-Catran, Claudia Schmiedeberg, Jürgen Schneider, Martijn Schoonvelde, Julia Schulte-Cloos, Sandy Schumann, Reinhard Schunck, Jürgen Schupp, Julian Seuring, Henning Silber, Willem Slegers, Nico Sonntag, Alexander Staudt, Nadia Steiber, Nils Steiner, Sebastian Sternberg, Dieter Stiers, Dragana Stojmenovska, Nora Storz, Erich Striessnig, Anne-Kathrin Stroppe, Janna Teltemann, Andrey Tibajev, Brian B. Tung, Giacomo Vagni, Jasper Van Assche, Meta van der Linden, Jolanda van der Noll, Arno Van Hooategem, Stefan Vogtenhuber, Bogdan Voicu, Fieke Wagemans, Nadja Wehl, Hannah Werner, Brenton M. Wiernik, Fabian Winter, Christof Wolf, Yuki Yamada, Nan Zhang, Conrad Ziller, Stefan Zins, Tomasz Żółtak, and Hung H. V. Nguyen (2021). *Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty*. en-us. DOI: 10.31222/osf.io/cd5j9. URL: <https://osf.io/preprints/metaarxiv/cd5j9/> (visited on 10/20/2022).
- Burns, Edwin J., Taylor Arnold, and Cindy M. Bukach (2019). “P-curving the fusiform face area: Meta-analyses support the expertise hypothesis”. en. In: *Neuroscience & Biobehavioral Reviews* 104, pp. 209–221. ISSN: 0149-7634. DOI: 10.1016/j.neubiorev.2019.07.003. URL: <https://www.sciencedirect.com/science/article/pii/S014976341830798X> (visited on 10/14/2022).
- Button, Katherine S, John P a Ioannidis, Claire Mokrysz, Brian a Nosek, Jonathan Flint, Emma S J Robinson, and Marcus R Munafò (2013a). “Power failure: why small sample size undermines the reliability of neuroscience.” In: *Nature reviews. Neuroscience* 14.5, pp. 365–76. ISSN: 1471-0048. DOI: 10.1038/nrn3475.
- Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò (2013b). “Power failure: why small sample size undermines the reliability of neuroscience”. en. In: *Nature Reviews Neuroscience* 14.5, pp. 365–376. ISSN: 1471-0048. DOI: 10.1038/nrn3475. URL: <https://www.nature.com/articles/nrn3475> (visited on 11/07/2018).
- Carp, Joshua (2012). “On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments”. en. In: *Frontiers in Neuroscience*, p. 13.



- Casey, B. J., Tariq Cannonier, May I. Conley, Alexandra O. Cohen, Deanna M. Barch, Mary M. Heitzeg, Mary E. Soules, Theresa Teslovich, Danielle V. Dellarco, Hugh Garavan, Catherine A. Orr, Tor D. Wager, Marie T. Banich, Nicole K. Speer, Matthew T. Sutherland, Michael C. Riedel, Anthony S. Dick, James M. Bjork, Kathleen M. Thomas, Bader Chaarani, Margie H. Mejia, Donald J. Hagler, M. Daniela Cornejo, Chelsea S. Sicut, Michael P. Harms, Nico U. F. Dosenbach, Monica Rosenberg, Eric Earl, Hauke Bartsch, Richard Watts, Jonathan R. Polimeni, Joshua M. Kuperman, Damien A. Fair, and Anders M. Dale (2018). “The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites”. en. In: *Developmental Cognitive Neuroscience*. The Adolescent Brain Cognitive Development (ABCD) Consortium: Rationale, Aims, and Assessment Strategy 32, pp. 43–54. ISSN: 1878-9293. DOI: 10.1016/j.dcn.2018.03.001. URL: <https://www.sciencedirect.com/science/article/pii/S1878929317301214> (visited on 06/06/2023).
- Cashman, Mikaela, Myra B. Cohen, Priya Ranjan, and Robert W. Cottingham (2018). “Navigating the maze: the impact of configurability in bioinformatics software”. en. In: *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*. Montpellier France: ACM, pp. 757–767. ISBN: 978-1-4503-5937-5. DOI: 10.1145/3238147.3240466. URL: <https://dl.acm.org/doi/10.1145/3238147.3240466> (visited on 10/20/2022).
- Chambers, Christopher D., Zoltan Dienes, Robert D. McIntosh, Pia Rotshtein, and Klaus Willmes (2015). “Registered Reports: Realigning incentives in scientific publishing”. en. In: *Cortex* 66, A1–A2. ISSN: 00109452. DOI: 10.1016/j.cortex.2015.03.022. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0010945215001136> (visited on 03/20/2021).
- Chen, Gang, Paul A. Taylor, and Robert W. Cox (2017). “Is the statistic value all we should care about in neuroimaging?” In: *NeuroImage* 147, pp. 952–959. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2016.09.066. URL: <https://www.sciencedirect.com/science/article/pii/S1053811916305432> (visited on 10/24/2023).
- Clement, Patricia, Marco Castellaro, Thomas W Okell, David L Thomas, Pieter Vandemaele, Sara Elgayar, Aaron Oliver-Taylor, Thomas Kirk, Joseph G Woods, Sjoerd B Vos, et al. (2022). “ASL-BIDS, the brain imaging data structure extension for arterial spin labeling”. In: *Scientific Data* 9.1, p. 543.
- Cox, Robert W (1996). “AFNI: software for analysis and visualization of functional magnetic resonance neuroimages”. In: *Computers and Biomedical research* 29.3, pp. 162–173.
- Cummings, P. (2004). “Meta-analysis based on standardized effects is unreliable.” In: *Archives of pediatrics & adolescent medicine* 158.6, pp. 595–7. ISSN: 1072-4710.
- Dafflon, Jessica, Pedro F. Da Costa, František Váša, Ricardo Pio Monti, Danilo Bzdok, Peter J. Hellyer, Federico Turkheimer, Jonathan Smallwood, Emily Jones, and Robert Leech (2022). “A guided multiverse study of neuroimaging analyses”. en. In: *Nature Communications* 13.1. Number: 1 Publisher: Nature Publishing Group, p. 3758. ISSN: 2041-1723. DOI: 10.1038/s41467-022-31347-8. URL: <https://www.nature.com/articles/s41467-022-31347-8> (visited on 10/14/2022).
- De Martino, Benedetto, Colin F Camerer, and Ralph Adolphs (2010). “Amygdala damage eliminates monetary loss aversion”. In: *Proceedings of the National Academy of Sciences* 107.8, pp. 3788–3792.

- Di Martino, Adriana, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. (2014). “The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism”. In: *Molecular psychiatry* 19.6, pp. 659–667.
- Eklund, Anders, Hans Knutsson, and Thomas E. Nichols (2018). “Cluster Failure Revisited: Impact of First Level Design and Data Quality on Cluster False Positive Rates”. en. In: *bioRxiv*, p. 296798. DOI: 10.1101/296798. URL: <https://www.biorxiv.org/content/10.1101/296798v2> (visited on 01/24/2019).
- Eklund, Anders, Thomas E. Nichols, and Hans Knutsson (2016a). “Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates”. en. In: *Proceedings of the National Academy of Sciences* 113.28, pp. 7900–7905. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1602413113. URL: <http://www.pnas.org/lookup/doi/10.1073/pnas.1602413113> (visited on 01/31/2019).
- (2016b). “Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates”. en. In: *Proceedings of the National Academy of Sciences* 113.28. Number: 28, pp. 7900–7905. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1602413113. URL: <http://www.pnas.org/lookup/doi/10.1073/pnas.1602413113> (visited on 02/02/2021).
- Esteban, Oscar, Christopher J. Markiewicz, Ross W. Blair, Craig A. Moodie, A. Ilkay Isik, Asier Erramuzpe, James D. Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit S. Ghosh, Jessey Wright, Joke Durnez, Russell A. Poldrack, and Krzysztof J. Gorgolewski (2019). “fMRIPrep: a robust preprocessing pipeline for functional MRI”. en. In: *Nature Methods* 16.1, pp. 111–116. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-018-0235-4. URL: <http://www.nature.com/articles/s41592-018-0235-4> (visited on 01/31/2019).
- Finn, Emily S., Xilin Shen, Dustin Scheinost, Monica D. Rosenberg, Jessica Huang, Marvin M. Chun, Xenophon Papademetris, and R. Todd Constable (2015). “Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity”. en. In: *Nature Neuroscience* 18.11. Number: 11 Publisher: Nature Publishing Group, pp. 1664–1671. ISSN: 1546-1726. DOI: 10.1038/nn.4135. URL: <https://www.nature.com/articles/nn.4135> (visited on 12/02/2023).
- Fisher, R.A. (1932). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.
- Flandin, Guillaume and Karl J Friston (2019). “Analysis of family-wise error rates in statistical parametric mapping using random field theory”. In: *Human brain mapping* 40.7, pp. 2052–2054.
- Ganz, Melanie, Sune H Holm, and Aasa Feragen (2021). “Assessing Bias in Medical AI”. en. In.
- Gau, Rémi, Stephanie Noble, Katja Heuer, Katherine L. Bottenhorn, Isil P. Bilgin, Yu-Fang Yang, Julia M. Huntenburg, Johanna M. M. Bayer, Richard A. I. Bethlehem, Shawn A. Rhoads, Christoph Vogelbacher, Valentina Borghesani, Elizabeth Levitis, Hao-Ting Wang, Sofie Van Den Bossche, Xenia Kobeleva, Jon Hartz Legarreta, Samuel Guay, Selim Melvin Atay, Gael P. Varoquaux, Dorien C. Huijser, Malin S. Sandström, Peer Herholz, Samuel A. Nastase, AmanPreet Badhwar, Guillaume Dumas, Simon Schwab, Stefano Moia, Michael Dayan, Yasmine Bassil, Paula P. Brooks, Matteo Mancini, James M. Shine, David O’Connor, Xihe Xie, Davide Poggiali, Patrick Friedrich, Anibal S. Heinsfeld, Lydia Riedl, Roberto Toro, César Caballero-Gaudes, Anders Eklund, Kelly G. Garner, Christopher R. Nolan, Damion V. Demeter, Fernando A. Barrios, Junaid S. Merchant, Elizabeth A. McDevitt, Robert Oostenveld, R. Cameron Craddock, Ariel Rokem, Andrew Doyle, Satrajit S. Ghosh, Aki Nikolaidis, Olivia W. Stanley, Eneko Uruñuela, and Brainhack Community

- (2021). “Brainhack: Developing a culture of open, inclusive, community-driven neuroscience”. eng. In: *Neuron* 109.11, pp. 1769–1775. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2021.04.001.
- Germani, Elodie, Elisa Fromont, and Camille Maumet (2022). *On the benefits of self-taught learning for brain decoding*. en. URL: <https://hal.inria.fr/hal-03769993> (visited on 11/11/2022).
- Germani, Elodie, Elisa Fromont, Pierre Maurel, and Camille Maumet (2023). “The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis”. working paper or preprint.
- Glatard, Tristan, Lindsay B. Lewis, Rafael Ferreira da Silva, Reza Adalat, Natacha Beck, Claude Lepage, Pierre Rioux, Marc-Etienne Rousseau, Tarek Sherif, Ewa Deelman, Najmeh Khalili-Mahani, and Alan C. Evans (2015). “Reproducibility of neuroimaging analyses across operating systems”. en. In: *Frontiers in Neuroinformatics* 9. ISSN: 1662-5196. DOI: 10.3389/fninf.2015.00012. URL: <http://journal.frontiersin.org/article/10.3389/fninf.2015.00012/abstract> (visited on 02/06/2021).
- Goble, Carole (2016). “Keynote: What is Reproducibility? The R\* Brouhaha”. en-us. In: DOI: [10/g75x4](https://doi.org/10.1007/978-1-4939-9754-4_10).
- Gorgolewski, Krzysztof, Daniel Margulies, and Michael Milham (2013). “Making Data Sharing Count: A Publication-Based Solution”. In: *Frontiers in Neuroscience* 7. ISSN: 1662-453X. URL: <https://www.frontiersin.org/articles/10.3389/fnins.2013.00009> (visited on 11/27/2023).
- Gorgolewski, Krzysztof J, Fidel Alfaro-Almagro, Tibor Auer, Pierre Bellec, Mihai Capotă, M Mallar Chakravarty, Nathan W Churchill, Alexander Li Cohen, R Cameron Craddock, Gabriel A Devenyi, et al. (2017). “BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods”. In: *PLoS computational biology* 13.3, e1005209.
- Gorgolewski, Krzysztof J., Tibor Auer, Vince D. Calhoun, R. Cameron Craddock, Samir Das, Eugene P. Duff, Guillaume Flandin, Satrajit S. Ghosh, Tristan Glatard, Yaroslav O. Halchenko, Daniel A. Handwerker, Michael Hanke, David Keator, Xiangrui Li, Zachary Michael, Camille Maumet, B. Nolan Nichols, Thomas E. Nichols, John Pellman, Jean-Baptiste Poline, Ariel Rokem, Gunnar Schaefer, Vanessa Sochat, William Triplett, Jessica A. Turner, Gaël Varoquaux, and Russell A. Poldrack (2016). “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments”. en. In: *Scientific Data* 3.1. Number: 1 Publisher: Nature Publishing Group, p. 160044. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.44. URL: <https://www.nature.com/articles/sdata201644> (visited on 08/11/2020).
- Gorgolewski, Krzysztof J., Gael Varoquaux, Gabriel Rivera, Yannick Schwartz, Vanessa V. Sochat, Satrajit S. Ghosh, Camille Maumet, Thomas E. Nichols, Jean-Baptiste Poline, Tal Yarkoni, Daniel S. Margulies, and Russell A. Poldrack (2015a). “NeuroVault.org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain”. In: *NeuroImage* 124.Pt B. Publisher: Elsevier, pp. 1242–1244. DOI: 10.1016/j.neuroimage.2015.04.016. URL: <https://www.hal.inserm.fr/inserm-01148969>.
- Gorgolewski, Krzysztof J., Gael Varoquaux, Gabriel Rivera, Yannick Schwarz, Satrajit S. Ghosh, Camille Maumet, Vanessa V. Sochat, Thomas E. Nichols, Russell A. Poldrack, Jean-Baptiste Poline, Tal Yarkoni, and Daniel S. Margulies (2015b). “NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain”. In: *Frontiers*

- in Neuroinformatics* 9.8. Publisher: Frontiers. DOI: 10.3389/fninf.2015.00008. URL: <https://www.hal.inserm.fr/inserm-01134573>.
- Greene, WH William H. (2012). *Econometric analysis*. Vol. 97, p. 1189. ISBN: 9780131395381. DOI: 10.1198/jasa.2002.s458. arXiv: arXiv:1011.1669v3. URL: <http://www.mendeley.com/research/econometric-analysis-195/> [https://scholar.google.com/scholar?hl=en&7B%5C&7Dq=econometric+analysis%7B%5C&7DbtnG=%7B%5C&7Das%7B%5C\\_%7Dsdt=1,10%7B%5C&7Das%7B%5C\\_%7Dsdtp=%7B%5C%7D1](https://scholar.google.com/scholar?hl=en&7B%5C&7Dq=econometric+analysis%7B%5C&7DbtnG=%7B%5C&7Das%7B%5C_%7Dsdt=1,10%7B%5C&7Das%7B%5C_%7Dsdtp=%7B%5C%7D1).
- Hoffmann, Sabine, Felix Schönbrodt, Ralf Elsas, Rory Wilson, Ulrich Strasser, and Anne-Laure Boulesteix (2021). “The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines”. en. In: *Royal Society Open Science* 8.4, rsos.201925, 201925. ISSN: 2054-5703. DOI: 10.1098/rsos.201925. URL: <https://royalsocietypublishing.org/doi/10.1098/rsos.201925> (visited on 08/24/2021).
- Hofstra, Bas, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland (2020). “The Diversity–Innovation Paradox in Science”. en. In: *Proceedings of the National Academy of Sciences* 117.17. ISBN: 9781915378118 Publisher: National Academy of Sciences Section: Social Sciences, pp. 9284–9291. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1915378117. URL: <https://www.pnas.org/content/117/17/9284> (visited on 06/29/2020).
- Holdgraf, Christopher, Stefan Appelhoff, Stephan Bickel, Kristofer Bouchard, Sasha D’Ambrosio, Olivier David, Orrin Devinsky, Benjamin Dichter, Adeen Flinker, Brett L Foster, et al. (2019). “iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology”. In: *Scientific data* 6.1, p. 102.
- Holmes, A. P., R. C. Blair, G. J.D. Watson, and I. Ford (1996). “Nonparametric Analysis of Statistic Images from Functional Mapping Experiments”. In: *Journal of Cerebral Blood Flow and Metabolism* 16, pp. 7–22. DOI: 10.1097/00004647.
- Jenkinson, Mark, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith (2012). “Fsl”. In: *Neuroimage* 62.2, pp. 782–790.
- Kabbara, Aya, Nina Forde, Camille Maumet, and Mahmoud Hassan (2023). “Successful reproduction of a large EEG study across software packages”. In: *Neuroimage: Reports*. Publisher: Elsevier, pp. 1–37. DOI: 10.1101/2022.08.03.502683. URL: <https://www.hal.inserm.fr/inserm-03747289>.
- Keator, D.B., K. Helmer, J. Steffener, J.A. Turner, T.G.M. Van Erp, S. Gadde, N. Ashish, G.A. Burns, and B.N. Nichols (2013). “Towards structured sharing of raw and derived neuroimaging data across existing resources”. en. In: *NeuroImage* 82, pp. 647–661. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2013.05.094. URL: <https://linkinghub.elsevier.com/retrieve/pii/S105381191300596X> (visited on 06/11/2019).
- Kennedy, David N, Christian Haselgrove, Jon Riehl, Nina Preuss, and Robert Buccigrossi (2016). “The NITRC image repository”. In: *NeuroImage* 124, pp. 1069–1073.
- Kiar, Gregory, Yohan Chatelain, Pablo de Oliveira Castro, Eric Petit, Ariel Rokem, Gaël Varoquaux, Bratislav Misic, Alan C. Evans, and Tristan Glatard (2020). “Numerical Instabilities in Analytical Pipelines Lead to Large and Meaningful Variability in Brain Networks”. en. In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 2020.10.15.341495. DOI: 10.1101/2020.10.15.341495. URL: <https://www.biorxiv.org/content/10.1101/2020.10.15.341495v3> (visited on 11/13/2020).

- Knight, Christopher G., Sylvia H. E. Knight, Neil Massey, Tolu Aina, Carl Christensen, Dave J. Frame, Jamie A. Kettleborough, Andrew Martin, Stephen Pascoe, Ben Sanderson, David A. Stainforth, and Myles R. Allen (2007). “Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models”. In: *Proceedings of the National Academy of Sciences* 104.30. Publisher: Proceedings of the National Academy of Sciences, pp. 12259–12264. DOI: 10.1073/pnas.0608144104. URL: <https://www.pnas.org/doi/10.1073/pnas.0608144104> (visited on 03/28/2022).
- Larrazabal, Agostina J., Nicolás Nieto, Victoria Peterson, Diego H. Milone, and Enzo Ferrante (2020). “Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis”. In: *Proceedings of the National Academy of Sciences* 117.23. Publisher: Proceedings of the National Academy of Sciences, pp. 12592–12594. DOI: 10.1073/pnas.1919012117. URL: <https://www.pnas.org/doi/full/10.1073/pnas.1919012117> (visited on 11/10/2022).
- Longo, Dan L. and Jeffrey M. Drazen (2016). “Data Sharing”. In: *New England Journal of Medicine* 374.3. Publisher: Massachusetts Medical Society \_eprint: <https://doi.org/10.1056/NEJMe1516564>, pp. 276–277. ISSN: 0028-4793. DOI: 10.1056/NEJMe1516564. URL: <https://doi.org/10.1056/NEJMe1516564> (visited on 11/27/2023).
- Lundgren, Anders and Upulee Kanewala (2016). “Experiences of testing bioinformatics programs for detecting subtle faults”. en. In: *Proceedings of the International Workshop on Software Engineering for Science*. Austin Texas: ACM, pp. 16–22. ISBN: 978-1-4503-4167-7. DOI: 10.1145/2897676.2897678. URL: <https://dl.acm.org/doi/10.1145/2897676.2897678> (visited on 09/27/2022).
- Markiewicz, Christopher J, Krzysztof J Gorgolewski, Franklin Feingold, Ross Blair, Yaroslav O Halchenko, Eric Miller, Nell Hardcastle, Joe Wexler, Oscar Esteban, Mathias Goncavles, et al. (2021). “The OpenNeuro resource for sharing of neuroscience data”. In: *eLife* 10, e71774.
- Maumet, Camille, Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, Tristan Glatard, Krzysztof J Gorgolewski, Karl G Helmer, Mark Jenkinson, David B Keator, B. Nolan Nichols, Jean-Baptiste Poline, Richard Reynolds, Vanessa Sochat, Jessica Turner, and Thomas E Nichols (2016). “Sharing brain mapping statistical results with the neuroimaging data model”. In: *Scientific Data* 3. Publisher: Nature Publishing Group. DOI: 10.1038/sdata.2016.102. URL: <https://www.hal.inserm.fr/inserm-01411025>.
- Maumet, Camille, Satrajit Ghosh, Yaroslav O. Halchenko, Dorota Jarecka, Nolan B. Nichols, Jean-Baptiste Poline, and Michael Hanke (2019). “The best of both worlds: using semantic web with JSON-LD. An example with NIDM-Results & Datalad”. In: *OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping*. Rome, Italy. URL: <https://www.hal.inserm.fr/inserm-01972649>.
- McKiernan, Erin C, Philip E Bourne, C Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, Brian A Nosek, Karthik Ram, Courtney K Soderberg, Jeffrey R Spies, Kaitlin Thaney, Andrew Updegrove, Kara H Woo, and Tal Yarkoni (2016). “How open science helps researchers succeed”. In: *eLife* 5. Ed. by Peter Rodgers. Publisher: eLife Sciences Publications, Ltd, e16800. ISSN: 2050-084X. DOI: 10.7554/eLife.16800. URL: <https://doi.org/10.7554/eLife.16800> (visited on 11/27/2023).

- Menuet, Romuald, Raphael Meudec, Jérôme Dockès, Gael Varoquaux, and Bertrand Thirion (2022). “Comprehensive decoding mental processes from Web repositories of functional brain images”. In: *Scientific Reports* 12.1, p. 7050.
- Milham, Michael P., R. Cameron Craddock, Jake J. Son, Michael Fleischmann, Jon Clucas, Helen Xu, Bonhwang Koo, Anirudh Krishnakumar, Bharat B. Biswal, F. Xavier Castellanos, Stan Colcombe, Adriana Di Martino, Xi-Nian Zuo, and Arno Klein (2018). “Assessment of the impact of shared brain imaging data on the scientific literature”. en. In: *Nature Communications* 9.1, pp. 1–7. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04976-1. URL: <https://www.nature.com/articles/s41467-018-04976-1> (visited on 12/16/2019).
- Miller, Karla L., Fidel Alfaro-Almagro, Neal K. Bangerter, David L. Thomas, Essa Yacoub, Junqian Xu, Andreas J. Bartsch, Saad Jbabdi, Stamatiou N. Sotiropoulos, Jesper L. R. Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Thomas W. Okell, Peter Weale, Iulius Dragonu, Steve Garratt, Sarah Hudson, Rory Collins, Mark Jenkinson, Paul M. Matthews, and Stephen M. Smith (2016). “Multimodal population brain imaging in the UK Biobank prospective epidemiological study”. en. In: *Nature Neuroscience* 19.11, pp. 1523–1536. ISSN: 1546-1726. DOI: 10.1038/nn.4393. URL: <https://www.nature.com/articles/nn.4393> (visited on 11/08/2018).
- Moran, Joseph M, Eshin Jolly, and Jason P Mitchell (2012). “Social-cognitive deficits in normal aging”. In: *Journal of neuroscience* 32.16, pp. 5553–5561.
- Mumford, J. A. and T. E. Nichols (2009a). “Simple group fMRI modeling and inference.” In: *NeuroImage* 47.4, pp. 1469–75. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2009.05.034.
- Mumford, Jeanette A. and Thomas Nichols (2009b). “Simple group fMRI modeling and inference”. en. In: *NeuroImage* 47.4, pp. 1469–1475. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2009.05.034. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1053811909005473> (visited on 03/25/2020).
- Nichols, T. E. and A. P Holmes (2002). “Nonparametric permutation tests for functional neuroimaging: a primer with examples.” In: *Human brain mapping* 15.1, pp. 1–25. ISSN: 1065-9471.
- Nichols, Thomas (2012). *SPM plot units*. en. URL: <https://blog.nisox.org/2012/07/31/spm-plot-units/> (visited on 12/11/2023).
- Nichols, Thomas E, Samir Das, Simon B Eickhoff, Alan C Evans, Tristan Glatard, Michael Hanke, Nikolaus Kriegeskorte, Michael P Milham, Russell A Poldrack, Jean-baptiste Poline, Erika Proal, Bertrand Thirion, David C Van Essen, Tonya White, and B T Thomas Yeo (2017a). “Best practices in data analysis and sharing in neuroimaging using MRI”. In: *Nature Neuroscience* 20.3, pp. 299–303. ISSN: 1097-6256. DOI: 10.1038/nn.4500. arXiv: arXiv:1011.1669v3. URL: <http://www.nature.com/doi/10.1038/nn.4500>.
- Nichols, Thomas E., Samir Das, Simon B. Eickhoff, Alan C. Evans, Tristan Glatard, Michael Hanke, Nikolaus Kriegeskorte, Michael P. Milham, Russell A. Poldrack, Jean-Baptiste Poline, Erika Proal, Bertrand Thirion, David C. Van Essen, Tonya White, and B. T. Thomas Yeo (2017b). “Best practices in data analysis and sharing in neuroimaging using MRI”. en. In: *Nature Neuroscience* 20.3. Number: 3 Publisher: Nature Publishing Group, pp. 299–303. ISSN: 1546-1726. DOI: 10.1038/nn.4500. URL: <https://www.nature.com/articles/nn.4500> (visited on 10/11/2022).
- Niso, Guiomar, Krzysztof J Gorgolewski, Elizabeth Bock, Teon L Brooks, Guillaume Flandin, Alexandre Gramfort, Richard N Henson, Mainak Jas, Vladimir Litvak, Jeremy T Moreau, et al.

- (2018). “MEG-BIDS, the brain imaging data structure extended to magnetoencephalography”. In: *Scientific data* 5.1, pp. 1–5.
- Nooner, Kate Brody, Stanley J Colcombe, Russell H Tobe, Maarten Mennes, Melissa M Benedict, Alexis L Moreno, Laura J Panek, Shaquanna Brown, Stephen T Zavitz, Qingyang Li, et al. (2012). “The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry”. In: *Frontiers in neuroscience* 6, p. 152.
- Norgaard, Martin, Granville J Matheson, Hanne D Hansen, Adam Thomas, Graham Searle, Gaia Rizzo, Mattia Veronese, Alessio Giacomel, Maqsood Yaqub, Matteo Tonietto, et al. (2022). “PET-BIDS, an extension to the brain imaging data structure for positron emission tomography”. In: *Scientific data* 9.1, p. 65.
- Obungoloch, Johnes, Joshua R Harper, Steven Consevage, Igor M Savukov, Thomas Neuberger, Srinivas Tadigadapa, and Steven J Schiff (2018). “Design of a sustainable prepolarizing magnetic resonance imaging system for infant hydrocephalus”. In: *Magnetic Resonance Materials in Physics, Biology and Medicine* 31, pp. 665–676.
- Open Brain Consent working group (2021). “The Open Brain Consent : Informing research participants and obtaining consent to share brain imaging data”. In: *Human Brain Mapping* 42.7, pp. 1945–1951. DOI: 10.1002/hbm.25351.
- Open Science Collaboration (2015). “Estimating the reproducibility of psychological science”. In: *Science* 349.6251. Publisher: American Association for the Advancement of Science, aac4716. DOI: 10.1126/science.aac4716. URL: <https://www.science.org/doi/10.1126/science.aac4716> (visited on 04/05/2023).
- Padmanabhan, Aarthi, Charles F Geier, Sarah J Ordaz, Theresa Teslovich, and Beatriz Luna (2011). “Developmental changes in brain function underlying the influence of reward processing on inhibitory control”. In: *Developmental cognitive neuroscience* 1.4, pp. 517–529.
- Pauli, Ruth, Alexander Bowring, Richard Reynolds, Gang Chen, Thomas E Nichols, and Camille Maumet (2016). “Exploring fMRI results space: 31 variants of an fMRI analysis in AFNI, FSL, and SPM”. In: *Frontiers in neuroinformatics* 10, p. 24.
- Penny, William D, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.
- Pereira, Juliana Alves, Mathieu Acher, Hugo Martin, Jean-Marc Jézéquel, Goetz Botterweck, and Anthony Ventresque (2021). “Learning software configuration spaces: A systematic literature review”. In: *Journal of Systems and Software* 182, p. 111044.
- Perkel, Jeffrey M. (2020). “Challenge to scientists: does your ten-year-old code still run?” en. In: *Nature* 584.7822. Bandiera\_abtest: a Cg\_type: Technology Feature Number: 7822 Publisher: Nature Publishing Group Subject\_term: Computational biology and bioinformatics, Computer science, Research data, Software, pp. 656–658. DOI: 10.1038/d41586-020-02462-7. URL: <https://www.nature.com/articles/d41586-020-02462-7> (visited on 11/28/2023).
- Pernet, Cyril (2014). “Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers”. In: *Frontiers in Neuroscience* 8. ISSN: 1662-453X. URL: <https://www.frontiersin.org/articles/10.3389/fnins.2014.00001> (visited on 12/11/2023).
- Pernet, Cyril, Marta I Garrido, Alexandre Gramfort, Natasha Maurits, Christoph M Michel, Elizabeth Pang, Riitta Salmelin, Jan Mathijs Schoffelen, Pedro A Valdes-Sosa, and Aina Puce (2020).

- “Issues and recommendations from the OHBM COBIDAS MEEG committee for reproducible EEG and MEG research”. In: *Nature neuroscience* 23.12, pp. 1473–1483.
- Pernet, Cyril R, Stefan Appelhoff, Krzysztof J Gorgolewski, Guillaume Flandin, Christophe Phillips, Arnaud Delorme, and Robert Oostenveld (2019). “EEG-BIDS, an extension to the brain imaging data structure for electroencephalography”. In: *Scientific data* 6.1, p. 103.
- Pinho, Ana Luisa, Alexis Amadon, Torsten Ruest, Murielle Fabre, Elvis Dohmatob, Isabelle Denghien, Chantal Ginisty, Séverine Becuwe-Desmidt, Séverine Roger, Laurence Laurier, et al. (2018). “Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping”. In: *Scientific data* 5.1, pp. 1–15.
- Piwowar, Heather A and Todd J Vision (2013). “Data reuse and the open data citation advantage”. In: *PeerJ* 1, e175.
- Poldrack, Russell A, Deanna M Barch, Jason Mitchell, Tor Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael Milham (2013a). “Toward open sharing of task-based fMRI data: the OpenfMRI project”. In: *Frontiers in neuroinformatics* 7, p. 12.
- Poldrack, Russell A., Deanna M. Barch, Jason P. Mitchell, Tor D. Wager, Anthony D. Wagner, Joseph T. Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael P. Milham (2013b). “Toward open sharing of task-based fMRI data: the OpenfMRI project”. eng. In: *Frontiers in Neuroinformatics* 7, p. 12. ISSN: 1662-5196. DOI: 10.3389/fninf.2013.00012.
- Poldrack, Russell A., Timothy O. Laumann, Oluwasanmi Koyejo, Brenda Gregory, Ashleigh Hover, Mei-Yen Chen, Krzysztof J. Gorgolewski, Jeffrey Luci, Sung Jun Joo, Ryan L. Boyd, Scott Hunicke-Smith, Zack Booth Simpson, Thomas Caven, Vanessa Sochat, James M. Shine, Evan Gordon, Abraham Z. Snyder, Babatunde Adeyemo, Steven E. Petersen, David C. Glahn, D. Reese McKay, Joanne E. Curran, Harald H. H. Göring, Melanie A. Carless, John Blangero, Robert Dougherty, Alexander Leemans, Daniel A. Handwerker, Laurie Frick, Edward M. Marcotte, and Jeanette A. Mumford (2015). “Long-term neural and physiological phenotyping of a single human”. In: *Nature Communications* 6, p. 8885. URL: <https://doi.org/10.1038/ncomms9885>.
- Poline, Jean-Baptiste, Janis L. Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O. Halchenko, Michael Hanke, Christian Haselgrove, Karl G. Helmer, David B. Keator, Daniel S. Marcus, Russell A. Poldrack, Yannick Schwartz, John Ashburner, and David N. Kennedy (2012). “Data sharing in neuroimaging research”. In: *Frontiers in Neuroinformatics* 6. ISSN: 1662-5196. DOI: 10.3389/fninf.2012.00009. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3319918/> (visited on 11/08/2018).
- Queder, Nazek, Vivian B Tien, Sanu Ann Abraham, Sebastian Georg Wenzel Urchs, Karl G Helmer, Derek Chaplin, Theo GM van Erp, David N Kennedy, Jean-Baptiste Poline, Jeffrey S Grethe, et al. (2023). “NIDM-Terms: community-based terminology management for improved neuroimaging dataset descriptions and query”. In: *Frontiers in Neuroinformatics* 17.
- Rae, Charlotte L., Martin Farley, Kate J. Jeffery, and Anne E. Urai (2022). “Climate crisis and ecological emergency: Why they concern (neuro)scientists, and what we can do”. en. In: *Brain and Neuroscience Advances* 6. Publisher: SAGE Publications Ltd STM, p. 23982128221075430. ISSN: 2398-2128. DOI: 10.1177/23982128221075430. URL: <https://doi.org/10.1177/23982128221075430> (visited on 11/27/2023).



- Research Parasite Award* (2023). en. Page Version ID: 1184064738. URL: [https://en.wikipedia.org/w/index.php?title=Research\\_Parasite\\_Award&oldid=1184064738](https://en.wikipedia.org/w/index.php?title=Research_Parasite_Award&oldid=1184064738) (visited on 11/27/2023).
- Ricci Lara, María Agustina, Rodrigo Echeveste, and Enzo Ferrante (2022). “Addressing fairness in artificial intelligence for medical imaging”. en. In: *Nature Communications* 13.1, p. 4581. ISSN: 2041-1723. DOI: 10.1038/s41467-022-32186-3. URL: <https://www.nature.com/articles/s41467-022-32186-3> (visited on 08/22/2022).
- Rolland, Xavier, Pierre Maurel, and Camille Maumet (2022). “Towards efficient fmri data re-use: can we run between-group analyses with datasets processed differently with spm ?” In: *ISBI 2022 - IEEE International Symposium on Biomedical Imaging*. Calcutta, India: IEEE, pp. 1–4. URL: <https://hal.archives-ouvertes.fr/hal-03607384> (visited on 11/11/2022).
- Salimi-Khorshidi, Gholamreza, Stephen M Smith, John R Keltner, Tor D Wager, and Thomas E Nichols (2009). “Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies”. In: *Neuroimage* 45.3, pp. 810–823.
- Schonberg, Tom, Craig R Fox, Jeanette A Mumford, Eliza Congdon, Christopher Trepel, and Russell A Poldrack (2012). “Decreasing ventromedial prefrontal cortex activity during sequential risk-taking: an fMRI investigation of the balloon analog risk task”. In: *Frontiers in neuroscience* 6, p. 80.
- Silberzahn, R, E L Uhlmann, D P Martin, P Anselmi, F Aust, E Awtrey, Š Bahník, F Bai, C Bannard, E Bonnier, R Carlsson, F Cheung, G Christensen, R Clay, M A Craig, A Dalla Rosa, L Dam, M H Evans, I Flores Cervantes, N Fong, M Gamez-Djokic, A Glenz, S Gordon-McKeon, T J Heaton, K Hederos, M Heene, A J Hofelich Mohr, F Högden, K Hui, M Johannesson, J Kalodimos, E Kaszubowski, D M Kennedy, R Lei, T A Lindsay, S Liverani, C R Madan, D Molden, E Molleman, R D Morey, L B Mulder, B R Nijstad, N G Pope, B Pope, J M Prenoveau, F Rink, E Robusto, H Roderique, A Sandberg, E Schlüter, F D Schönbrodt, M F Sherman, S A Sommer, K Sotak, S Spain, C Spörlein, T Stafford, L Stefanutti, S Tauber, J Ullrich, and M Vianello (2018). “Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results”. en. In: *PSYCHOLOGICAL SCIENCE*, p. 21.
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn (2011). “False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant”. eng. In: *Psychological Science* 22.11, pp. 1359–1366. ISSN: 1467-9280. DOI: 10.1177/0956797611417632.
- Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson (2020). “Specification curve analysis”. en. In: *Nature Human Behaviour* 4.11, pp. 1208–1214. ISSN: 2397-3374. DOI: 10.1038/s41562-020-0912-z. URL: <http://www.nature.com/articles/s41562-020-0912-z> (visited on 01/04/2021).
- Steege, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel (2016). “Increasing Transparency Through a Multiverse Analysis”. en. In: *Perspectives on Psychological Science* 11.5, pp. 702–712. ISSN: 1745-6916, 1745-6924. DOI: 10.1177/1745691616658637. URL: <http://journals.sagepub.com/doi/10.1177/1745691616658637> (visited on 11/06/2020).
- Stouffer, S., L. DeVinney, and E. Suchmen (1949). *The American Soldier: Adjustment During Army Life*. Vol. 1. Princeton University Press, Princeton, NJ.
- Strother, Stephen, Stephen La Conte, Lars Kai Hansen, Jon Anderson, Jin Zhang, Sujit Pulapura, and David Rottenberg (2004). “Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis”. en. In: *NeuroImage*. Math-

- ematics in *Brain Imaging* 23, S196–S207. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2004.07.022. URL: <https://www.sciencedirect.com/science/article/pii/S1053811904003945> (visited on 10/14/2022).
- Tom, Sabrina M, Craig R Fox, Christopher Trepel, and Russell A Poldrack (2007). “The neural basis of loss aversion in decision-making under risk”. In: *Science* 315.5811, pp. 515–518.
- Turing Way, The, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, Anna Krystalli, Alexander Morley, Martin O’Reilly, and Kirstie Whitaker (2019). “The Turing Way: A Handbook for Reproducible Data Science”. In: *Zenodo*. ADS Bibcode: 2019zndo...3233986W. DOI: 10.5281/zenodo.3233986. URL: <https://ui.adsabs.harvard.edu/abs/2019zndo...3233986W> (visited on 11/28/2023).
- UNESCO (2021). *UNESCO Recommendation on Open Science*. Tech. rep. UNESCO. DOI: 10.54677/MNMH8546. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000379949> (visited on 12/02/2023).
- Van Essen, David C., Stephen M. Smith, Deanna M. Barch, Timothy E. J. Behrens, Essa Yacoub, Kamil Ugurbil, and WU-Minn HCP Consortium (2013). “The WU-Minn Human Connectome Project: an overview”. eng. In: *NeuroImage* 80, pp. 62–79. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2013.05.041.
- Van Horn, John Darrell and Michael S. Gazzaniga (2013). “Why share data? Lessons learned from the fMRIDC”. In: *NeuroImage* 82, pp. 677–682. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2012.11.010. URL: <https://www.sciencedirect.com/science/article/pii/S1053811912011068> (visited on 10/24/2023).
- Varoquaux, Gaël, Yannick Schwartz, Russell A. Poldrack, Baptiste Gauthier, Danilo Bzdok, Jean-Baptiste Poline, and Bertrand Thirion (2018). “Atlases of cognition with large-scale human brain mapping”. en. In: *PLOS Computational Biology* 14.11. Ed. by Jörn Diedrichsen, e1006565. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006565. URL: <https://dx.plos.org/10.1371/journal.pcbi.1006565> (visited on 10/25/2023).
- Whitaker, Kirstie and Olivia Guest (2020). “#bropenscience is broken science”. en. In: p. 4.
- White, Tonya, Elisabet Blok, and Vince D. Calhoun (2022). “Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed”. en. In: *Human Brain Mapping* 43.1. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25120>, pp. 278–291. ISSN: 1097-0193. DOI: 10.1002/hbm.25120. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25120> (visited on 11/27/2023).
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. ’t Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons (2016). “The FAIR Guiding Principles for scientific data management and stewardship”. en. In: *Scientific Data* 3.1. Number:

- 1 Publisher: Nature Publishing Group, p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18. URL: <https://www.nature.com/articles/sdata201618> (visited on 11/27/2023).
- Williams, Chad C., Thomas D. Ferguson, Cameron D. Hassall, Wande Abimbola, and Olave E. Krigolson (2021). "The ERP, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study". en. In: *Psychophysiology* 58.2. ISSN: 0048-5772, 1469-8986. DOI: 10.1111/psyp.13722. URL: <https://onlinelibrary.wiley.com/doi/10.1111/psyp.13722> (visited on 11/15/2021).
- Yarkoni, Tal (2019). *The Generalizability Crisis*. Tech. rep. type: article. PsyArXiv. DOI: 10.31234/osf.io/jqw35. URL: <https://psyarxiv.com/jqw35/> (visited on 03/11/2021).
- Yarkoni, Tal, Russell A. Poldrack, Thomas E. Nichols, David C. Van Essen, and Tor D. Wager (2011). "Large-scale automated synthesis of human functional neuroimaging data". en. In: *Nature Methods* 8.8, pp. 665–670. ISSN: 1548-7105. DOI: 10.1038/nmeth.1635. URL: <https://www.nature.com/articles/nmeth.1635> (visited on 11/08/2018).
- Zaykin, D V (2011). "Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis." In: *Journal of evolutionary biology* 24.8, pp. 1836–41. ISSN: 1420-9101. DOI: 10.1111/j.1420-9101.2011.02297.x.
- Zuo, Xi-Nian, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John Breitner, Randy L Buckner, Vince D Calhoun, F Xavier Castellanos, et al. (2014). "An open science resource for establishing reliability and reproducibility in functional connectomics". In: *Scientific data* 1.1, pp. 1–13.



# List of Acronyms

**BEP** BIDS extension proposal. 15, 38

**BIDS** Brain Imaging Data Structure. 14, 15, 16, 37, 38, 59

**FAIR** Findable, Accessible, Interoperable and Reusable. 12, 37, 39, 59

**FGLS** feasible generalized least squares. 43

**FWHM** full width at half maximum. 48

**GDPR** European General Data Protection Rule. 13, 22, 23

**GLM** General Linear Model. 25, 26, 42, 48

**HCP Young Adult** Human Connectome Project: Young Adult. 12, 47, 50

**HRF** Hemodynamic Response Function. 20, 31, 48

**IBMA** Image-Based Meta-Analyses. 42

**INCF** International Neuroinformatics Coordinating Facility. 10, 14

**INDI** International Neuroimaging Data-sharing Initiative. 11, 39

**MRI** Magnetic Resonance Imaging. 5, 6, 7, 8, 14, 15, 16, 54, 61

**NIDASH** NeuroImaging Data SHaring task force. 14, 18

**NIDM** Neuroimaging Data Model. 14, 15, 16, 37

**OHBM** Organization for Human Brain Mapping. 10, 54

**task-fMRI** task-based functional Magnetic Resonance Imaging. 5, 6, 11, 13, 15, 18, 19, 20, 21, 25, 26, 29, 30, 31, 32, 33, 34, 37, 40, 41, 42, 45, 47, 49, 50, 59



# Credits

The L<sup>A</sup>T<sub>E</sub>X template used in this manuscript was adapted from an original template by Pierre Maurel. The credits for the icons used for each section of the contributions are:

- Context: Creative Commons *brain side* by tezar tantular on the noun project.
- Data: Creative Commons MRI by LAFS on the noun project.
- Contributions: Creative Commons *Light* by Numero Uno on the noun project.
- Results: Creative Commons *Graph* by Nisa Varma on the noun project, the noun project.
- Publications: Creative Commons *Book* by Yayat Dayat on the noun project.
- Talks: Creative Commons *presentation* by Gregor Cresnar on the noun project.
- Open artefacts: Creative Commons *Omnichannel* by sevgenjory on the noun project.

Icons used in Fig. 4 are from the CC BY 4.0 Deed Olivia template from Slide Carnival. Other image credits are included within the caption of each figure.