



**HAL**  
open science

# The Brain Imaging Data Structure: Overview of the project and what BIDS does and allows to do

Camille Maumet

## ► To cite this version:

Camille Maumet. The Brain Imaging Data Structure: Overview of the project and what BIDS does and allows to do. GliMR Hybrid Networking Event, Sep 2021, Brno (hybrid), Czech Republic. , 2021. hal-04255147

**HAL Id: hal-04255147**

**<https://inria.hal.science/hal-04255147v1>**

Submitted on 23 Oct 2023

**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



# The Brain Imaging Data Structure

Overview of the project and what BIDS does and allows to do



**Camille Maumet**

Univ Rennes, Inria, CNRS, Inserm



# Glioma MR Imaging 2.0

- Build a pan-European and multidisciplinary network of **international experts in glioma research, patient organisations, data scientists, and MR imaging scientists**
- Use **advanced MR imaging for improved decision making** in diagnosis, patient monitoring, and assessment of treatment response in clinical trials and clinical practice.



Esther Warnert



Radim Jancalek



Jan Petr



Vera Keil



Lydiane Hirschler



## WG1 - Advanced MRI biomarkers for glioma characterisation

Working Group 1 focusses on coordinating the identification and quantification of advanced MRI biomarkers for the application in the field of glioma.



Kyrre Emblem



Marion Smits



Fran Borovecki



## WG4 – Stakeholder relations

Working Group 4 ensures representation of all relevant stakeholders within GliMR and acts as liaison from GliMR to all relevant stakeholders.



## WG2 - Multi-site data integration

Working Group 2 coordinates multi-site data integration and enables the creation of large datasets in glioma diagnostics via the creation of common GDPR- and BIDS-compliant forms and data structures.



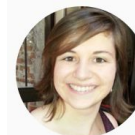
## WG5 – Dissemination

Working Group 5 disseminates the Action's goals and results in tailored manners to all relevant stakeholders.



## WG3 - Clinical translation

Working Group 3 focusses on fostering cross-border information exchange about past and current clinical trials and studies of glioma.



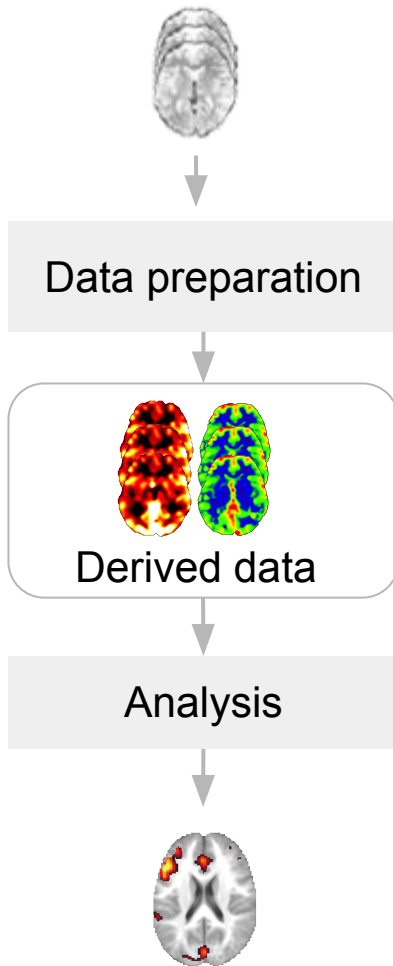
Patricia Clement



Joana Pinto



# Gliomas research in Europe

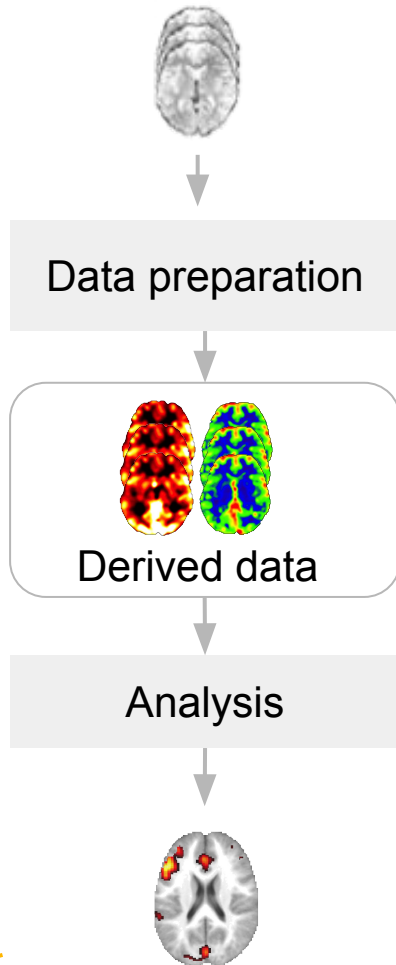


- **50,000 new cases** of gliomas per year in Europe

- **Low incidence** rate

>> Limited samples sizes

# Why do we need large samples?



## Reproducibility crisis

SCIENCE

### A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



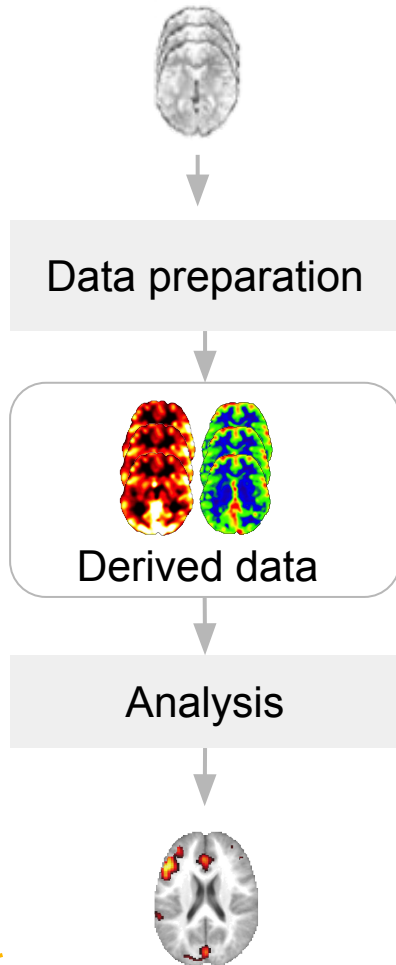
SEAN NEL / SHUTTERSTOCK

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

It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

By Ed Yong, The Atlantic Science, May 2019, [link](#).

# Why do we need large samples?



## Reproducibility crisis

SCIENCE

### A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



SEAN NEL / SHUTTERSTOCK

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

It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

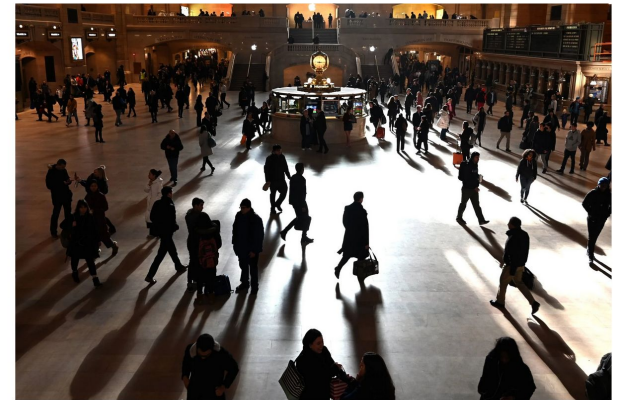
By Ed Yong, The Atlantic Science, May 2019, [link](#).

## Algorithmic injustice

The New York Times

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

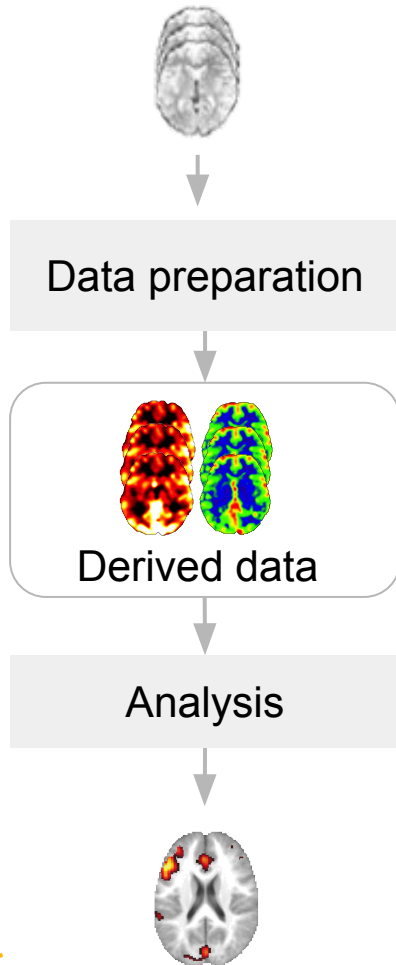
Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



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

By Natasha Singer and Cade Metz, The New York Times, Dec. 2019, [link](#).

# Why do we need large samples?



## Reproducibility crisis

SCIENCE

### A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019



SEAN NEL / SHUTTERSTOCK

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

It was a blockbuster discovery at the time. The team found that a less active version of the gene was more common among 454 people who had mood disorders than in 570 who did not. In theory, anyone who had this particular gene variant could be at higher risk for depression, and that finding, they said, might help in diagnosing such disorders, assessing suicidal behavior, or even

By Ed Yong, The Atlantic Science, May 2019, [link](#).

## Algorithmic injustice

The New York Times

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

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



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

By Natasha Singer and Cade Metz, The New York Times, Dec. 2019, [link](#).

As a **community**, we need to create and work on **bigger & more representative and diverse datasets**

# GliMR WG2 **Multi-site data integration**



## WG2 - Multi-site data integration

Working Group 2 coordinates multi-site data integration and enables the creation of large datasets in glioma diagnostics via the creation of common GDPR- and BIDS-compliant forms and data structures.



Camille  
Maumet

Cyril  
Pernet



**Data Privacy**



**Data infrastructure**



**Data portability**

Image credits: "Privacy lock", "Transport map", "Mini bus"



## WG2 - Multi-site data integration

Working Group 2 coordinates multi-site data integration and enables the creation of large datasets in glioma diagnostics via the creation of common GDPR- and BIDS-compliant forms and data structures.



Camille  
Maumet

Cyril  
Pernet



**Data Privacy**



**Data infrastructure**



**Data portability**

Image credits: "Privacy lock", "Transport map", "Mini bus"



# Data portability

How differences in acquisition parameters can be conveyed with enough details to support future post-processing?

- **Extend the international BIDS standard** in setting up the structures for sharing advanced MR imaging data.



## Brain Imaging Data Structure (BIDS):

A global community of 200+ members!



Guiomar  
Niso



Melanie  
Ganz



Robert  
Oostenveld



Russ  
Poldrack



Kirstie  
Whitaker

# Why BIDS?

- Data sharing and exchange
  - With other labs
  - Within your lab
  - With yourself in 6 months :)
- Automatic pipeline processing

# What is BIDS?



- It's a data **structure** ; nothing to do with **format** per se
- It's about:
  - how you organize data in a folder
  - how you name files
  - how you document metadata
  - using community standards and dictionaries to do all the above
- It cares about imaging data but also behaviour/cognition

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

# What is BIDS?

## CREDITS

# SCIENTIFIC DATA

## OPEN

### SUBJECT CATEGORIES

- » Data publication and archiving
- » Research data

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski<sup>1</sup>, Tibor Auer<sup>2</sup>, Vince D. Calhoun<sup>3,4</sup>, R. Cameron Craddock<sup>5,6</sup>, Samir Das<sup>7</sup>, Eugene P. Duff<sup>8</sup>, Guillaume Flandin<sup>9</sup>, Satrajit S. Ghosh<sup>10,11</sup>, Tristan Glatard<sup>7,12</sup>, Yaroslav O. Halchenko<sup>13</sup>, Daniel A. Handwerker<sup>14</sup>, Michael Hanke<sup>15,16</sup>, David Keator<sup>17</sup>, Xiangrui Li<sup>18</sup>, Zachary Michael<sup>19</sup>, Camille Maumet<sup>20</sup>, B. Nolan Nichols<sup>21,22</sup>, Thomas E. Nichols<sup>20,23</sup>, John Pellman<sup>6</sup>, Jean-Baptiste Poline<sup>24</sup>, Ariel Rokem<sup>25</sup>, Gunnar Schaefer<sup>1,26</sup>, Vanessa Sochat<sup>27</sup>, William Triplett<sup>1</sup>, Jessica A. Turner<sup>3,28</sup>, Gaël Varoquaux<sup>29</sup> & Russell A. Poldrack<sup>1</sup>

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016



Data Sharing Task Force



The Poldrack Lab @ Stanford

# What is BIDS?

## Evolution of BIDS

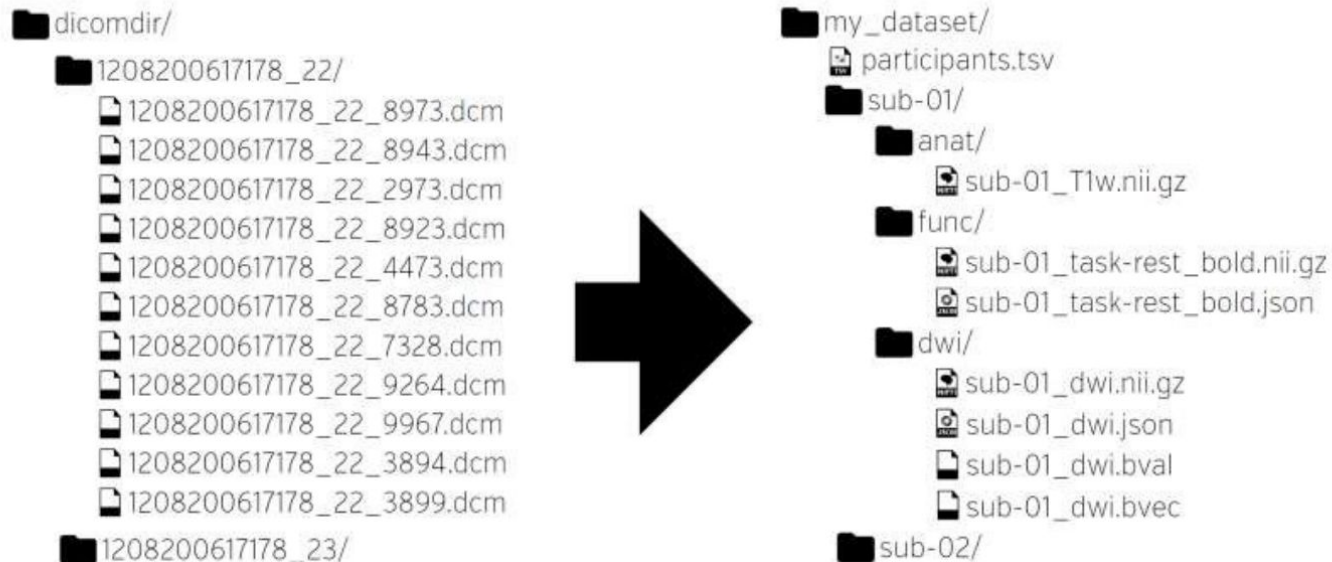
1. Kickoff meeting at Stanford in Spring 2015
2. Meeting at OHBM 2015 (June)
3. Introduced to neuroinformatics community at INCF Congress 2015 (August)
4. First release candidate and public call for comments (September)
5. Version 1.0.0 published along the introductory paper

→ Initially covered structural MRI and fMRI, now all sort of MRI, PET, EEG, MEG, iEEG, extensions for animal, connectivity, imaging genomics ..

# What is BIDS?

## Essence of BIDS

- Directory hierarchy and file naming system
  - Conveys essential meta-data in filenames alone
- Remaining meta-data in “sidecar” files





# What is BIDS?

## Metadata as text files (tsv, json) with standard dictionary

Name	Date modified	Type
stimuli	21/03/2018 21:58	File folder
sub-002	24/03/2018 08:56	File folder
sub-003	24/03/2018 08:56	File folder
sub-004	24/03/2018 08:56	File folder
sub-005	24/03/2018 08:56	File folder
sub-006	24/03/2018 08:56	File folder
sub-007	24/03/2018 08:56	File folder
sub-008	24/03/2018 08:56	File folder
sub-009	24/03/2018 08:56	File folder
sub-010	24/03/2018 08:56	File folder
sub-011	24/03/2018 08:56	File folder
sub-012	24/03/2018 08:56	File folder
sub-013	24/03/2018 08:56	File folder
sub-014	24/03/2018 08:56	File folder
sub-015	24/03/2018 08:56	File folder
sub-016	24/03/2018 08:56	File folder
sub-017	24/03/2018 08:56	File folder
sub-018	24/03/2018 08:56	File folder
sub-019	24/03/2018 08:56	File folder
dataset_description.json	15/03/2018 11:30	JSON File
participants.tsv	19/03/2018 20:21	TSV File
README.txt	15/03/2018 11:33	TXT File

```
participant_id age sex
sub-002        34  M
sub-003        12  F
sub-004        33  F
```

**Subjects info are shared at the root - easy to figure populations, age, and other basic demographics**

# What is BIDS? A community!



**Technical challenges:** Standards, Databases, Good practices (version control, etc.)

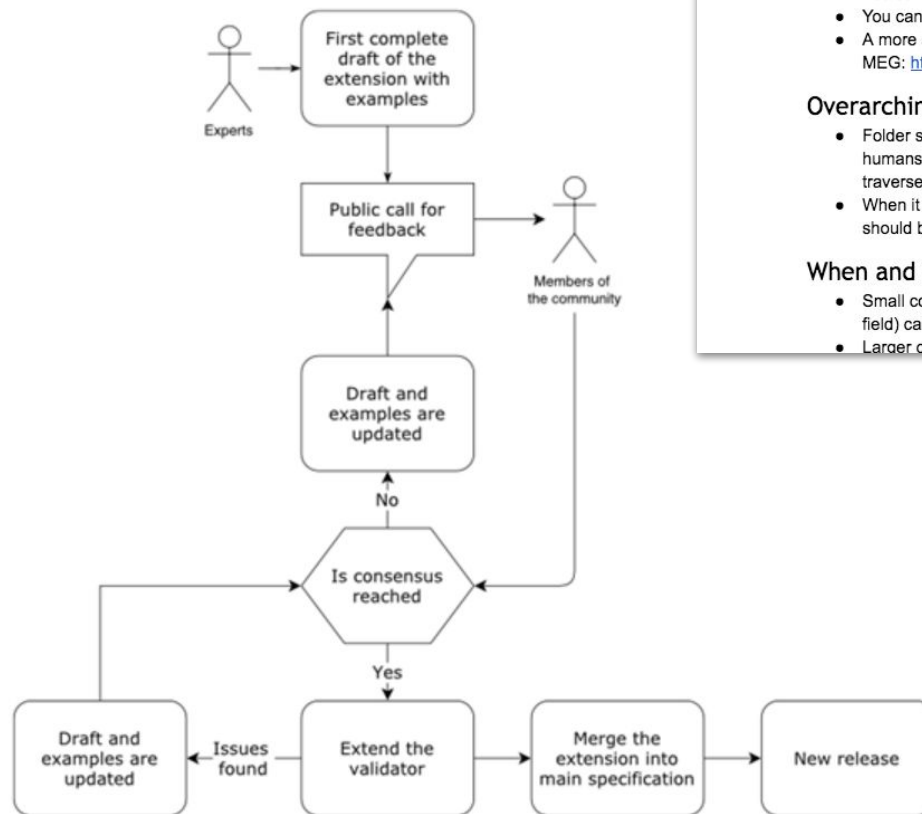
## A culture change: towards community-driven research

- Making resources **available to the scientific community**
- **Reusing** existing resources (datasets, tools) that were created by others
- **Developing tools collaboratively** with a community of contributors

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

# Extending BIDS: the BEPs

## BIDS Extension Proposals



### BIDS Extension Proposals: A Guide

#### Why contribute to BIDS?

- You can make something that is good even better without building it from scratch.
- You can interact with and get to know other experts in the field.
- A more substantial extension can lead to standalone publications (see for instance BIDS MEG: <https://www.nature.com/articles/sdata2018110>)

#### Overarching principles in the design of BIDS

- Folder structure must be not only machine-readable but also easy to interpret by humans. In other words, readability is as important as the ability to programmatically traverse the folder structure.
- When it comes to file names within folders, redundancy trumps concision – the files should be identifiable by their names even if removed from the folder structure.

#### When and how to start a BIDS Extension Proposal?

- Small contributions (typos, rephrasing of a description, adding a single new metadata field) can be just added as a [Pull Request on GitHub](#)
- Larger contributions that are expected to involve longer and more involved discussions



# The Brain Imaging Data Structure

Overview of the project and what BIDS does and allows to do

Camille Maumet



# Thank you!