

Desvendar o Potencial da Ciência Aberta: Perspectivas da Engenharia Biomédica e Ciência de Dados

Alexandre Sayal^{1,2}, Bruno Direito^{2,1}, Carolina Travassos^{1,2}, João Pereira^{1,2}

1 - Centro de Imagem Biomédica e Investigação Translacional, UC

2 - Centro de Informática e Sistemas, UC

*todos os autores contribuem de igual forma para o trabalho



This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Comunidades de Ciência Aberta @ Universidade de Coimbra
06 de Maio de 2024

Dados FAIR

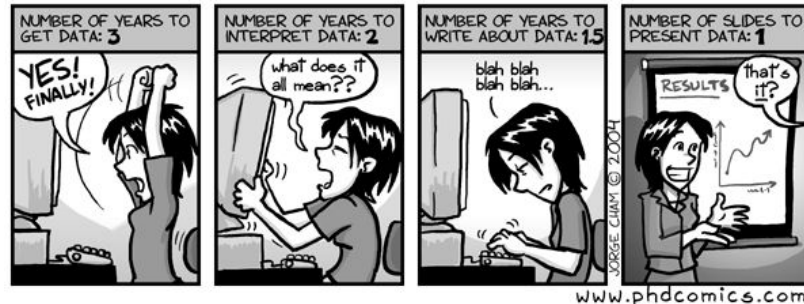
(...)Mais do que acesso aberto de dados e publicações, **Ciência Aberta é a construção de pontes no processo científico,** promovendo **transparência, acessibilidade e colaboração** na investigação científica.

A **interoperabilidade** (um dos princípios FAIR) é um aspecto chave na gestão dos dados de investigação.

IA e dados de investigação

“If **80 percent of our work is data preparation**, then ensuring **data quality is the important work** of a machine learning team.”

Andrew Ng, founder Coursera, Stanford University



Gestão de dados de investigação

A utilização de standards é fundamental para todo o ciclo de vida dos dados de investigação

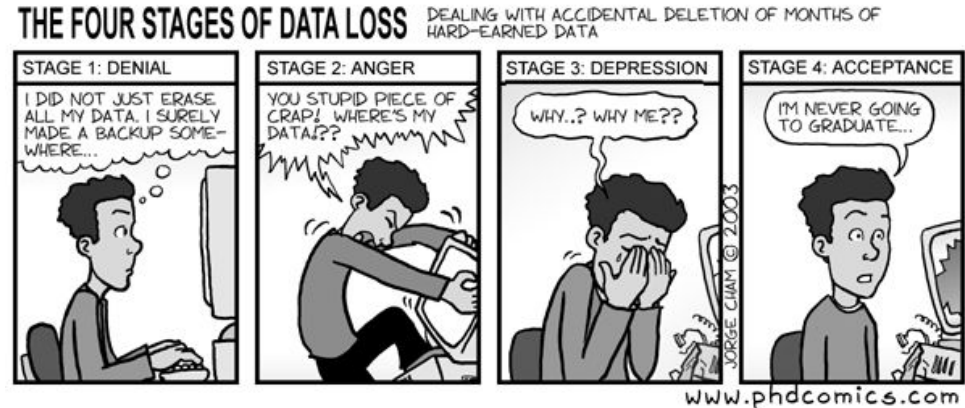
- Facilita a Gestão de Dados de Investigação
 - FAIR data
 - Findable, Accessible, Interoperable, Reusable
 - Potencia a reutilização
 - Explorar novas técnica/teorias



<https://rdmkit.elixir-europe.org/>

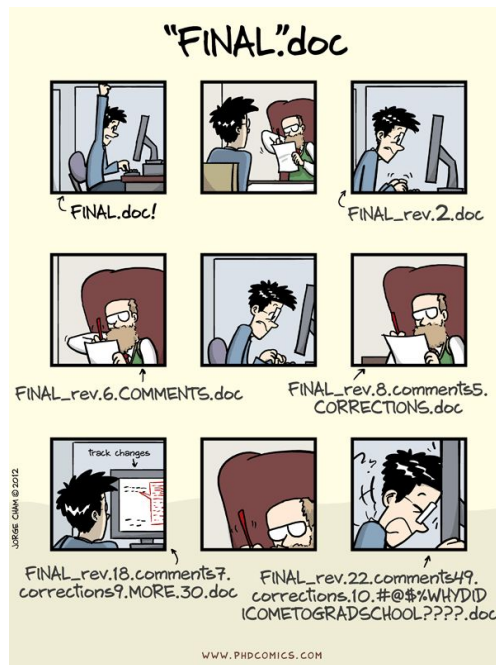
A imagem biomédica e as ciências biomédicas

- Os dados são o elemento fundamental do processo científico
 - A sua partilha é chave
- Dados heterogêneos, não estruturados
 - Desafios: Partilha, automatização de processos, **probabilidade de erro na sua curadoria**



<https://phdcomics.com/comics/archive.php?comid=382>

Dados e datasets

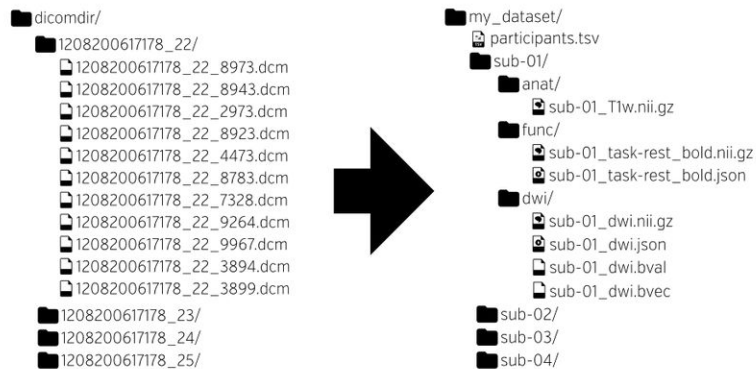


Utilização de standards na formatação e na estrutura de conjuntos de dados de investigação

- Facilita a partilha
 - as mesmas regras entre emissor e receptor - interoperabilidade
- Criação de pipelines de análise,
 - Reutilização
 - Multicêntricos / colaborativos
 - Utilização de novos métodos estado de arte desenvolvidos
- Facilita a curadoria
 - Desenvolvimento de regras, métodos
 - multicêntricos/colaborativos

BIDS e a Gestão de dados de investigação

- Brain Imaging Data Structure
 - Estrutura modular dos conjuntos de dados (source, raw, derivatives)
 - Especificação de uma estrutura de pastas
 - Nomenclatura de ficheiros
 - human-readable a machine-actionable
 - Formatos abertos para os vários ficheiros
 - Documentos e formato de metadados
 - Minimiza duplicação



<https://bids.neuroimaging.io/>

BIDS e a Gestão de dados de investigação

- Como?
 - Formato do standard

The screenshot shows the website for Brain Imaging Data Structure v1.6.0. The page is titled 'Common principles' and 'Definitions'. The left sidebar contains a navigation menu with items like 'The BIDS Specification', 'Introduction', 'Common principles', 'Modality agnostic files', 'Modality specific files', 'Derivatives', 'Longitudinal and multi-site studies', 'BIDS Extension Proposals', 'Appendix', 'Changelog', and 'The BIDS Starter Kit'. The main content area is divided into two sections: 'Common principles' and 'Definitions'. The 'Definitions' section contains three numbered items: 1. Dataset, 2. Subject, and 3. Session. A 'Table of contents' sidebar is visible on the right, listing various topics like 'Definitions', 'File name structure', 'Entity-linked file collections', etc.

Brain Imaging Data Structure v1.6.0

Search

Brain Imaging Data Structure v1.6.0

The BIDS Specification

Introduction

Common principles

Modality agnostic files

Modality specific files

Derivatives

Longitudinal and multi-site studies

BIDS Extension Proposals

Appendix

Changelog

The BIDS Starter Kit

Common principles

Definitions

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

Throughout this specification we use a list of terms and abbreviations. To avoid misunderstanding we clarify them here.

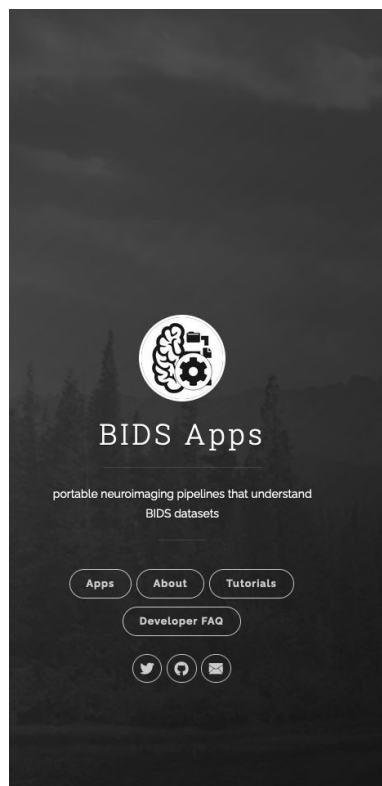
- Dataset** - a set of neuroimaging and behavioral data acquired for a purpose of a particular study. A dataset consists of data acquired from one or more subjects, possibly from multiple sessions.
- Subject** - a person or animal participating in the study. Used interchangeably with term **Participant**.
- Session** - a logical grouping of neuroimaging and behavioral data consistent across

Table of contents

- Definitions
- Compulsory, optional, and additional data and metadata
- File name structure
 - Entity-linked file collections
- Source vs. raw vs. derived data
- Storage of derived datasets
- Non-compliant derivatives
- The Inheritance Principle
- Good practice recommendations
- File Formation specification
- Imaging files
 - Tabular files
 - Key/value files (dictionaries)
- Participant names and other labels

BIDS e a Gestão de dados de investigação

- Como?
 - Identificação de ferramentas



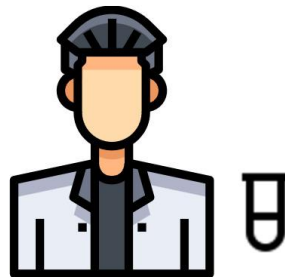
Available BIDS Apps

All apps

afni_proc	circleci failing	last commit last tuesday	open issues 4
antsCorticalThickness	circleci passing	last commit last tuesday	open issues 0
baracus	circleci failing	last commit last tuesday	open issues 4
brainiak-srm	circleci passing	last commit last tuesday	open issues 0
BrainSuite	circleci passing	last commit last tuesday	open issues 0
BROCCOLI	circleci failing	last commit last tuesday	open issues 5
CPAC	circleci failing	last commit last tuesday	open issues 10
DPARSF	circleci failing	last commit last tuesday	open issues 2
example	circleci passing	last commit last tuesday	open issues 6
freesurfer	circleci failing	last commit last tuesday	open issues 12
giga_connectome	test.yml passing	last commit today	open issues 13
HCPIPipelines	circleci failing	last commit last monday	open issues 16
hyperalignment	circleci failing	last commit last tuesday	open issues 1
MAGEtBrain	circleci failing	last commit last tuesday	open issues 3
mindboggle	circleci failing	last commit last tuesday	open issues 9
MRtrix3_connectome	circleci failing	last commit august 2023	open issues 32
ndmg	circleci failing	last commit last tuesday	open issues 2
nipypelines	circleci failing	last commit january 2017	open issues 1
opnri	circleci failing	last commit last tuesday	open issues 7

BIDS e a Gestão de dados de investigação

- Como?
 - Definir pipeline
 - Estruturar os dados seguindo as regras do standard
 - validar estrutura (ferramentas de validação disponível)
 - Identificar as ferramentas para o processamento / análise
 - Automatizar processos
 - Validar resultados

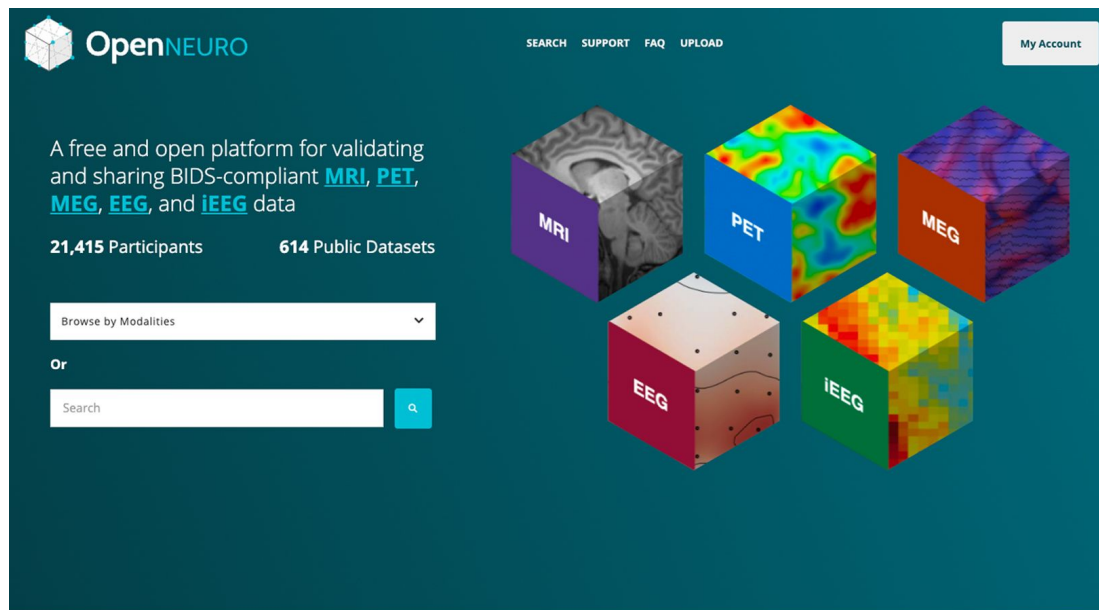


BIDS e a Investigação responsável e transparente

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

From [The Turing Way, Ch. 2](#); doi:10.5281/zenodo.3233853

BIDS e Repositórios de dados



OpenNEURO

SEARCH SUPPORT FAQ UPLOAD My Account

A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data

21,415 Participants 614 Public Datasets

Browse by Modalities

Or

Search

MRI PET MEG EEG iEEG



<https://openneuro.org/>

Ciência aberta e formatos standards como um desafio da comunidade

incf | NeuroStars

50+ results for BIDS

BIDS Topics/posts Q Search

Advanced filters

Sort by Relevance

- BIDS.BIDSLayout.build_path() removes dir- label from dwi images**
pybids
Sep '22 - I am using `BIDSLayout.build_path()` to build valid **BIDS**-paths for images that I am saving. When I try to build a path using modified entities for a DWI image, bl...
- BIDS-Apps Google Season of Docs information**
Announcements bids, bids-app
May '20 - Summary INCF is accepted as a <https://developers.google.com/season-of-docs> Google Seasons of Docs (GSoD) organization & **BIDS**-INCF is listed as one of the projects (<https://www.incf.org/activities/google-season-of-docs> see the proposal - project #10) GSoD is a program spon...
- BidsValidator Error when running fmriprep although browser bids-validator was succesful**
fmriprep, bids
Sep '21 - I am running fmriprep using singularity on a directory that passed `bids-validator` successfully. However, I get the following error: Making sure the input data is **BIDS** compliant (warnings can be ignored in most cases), 1...
- Bids-validator vs BIDS Specification for T2starw**
Software Support bids, bids-validator
Dec '23 - I'm trying to validate a **BIDS** dataset and I'm getting an error that the T2starw files are NOT_INCLUDED. The reason seems to be that the **BIDS** Specification allows for echo with T2S...
- EEGLAB: BIDS import of condition per participant**
bids, matlab, eeglab
24d - Dear NeuroStars, I'm new to the brain imaging data structure (**BIDS**: <https://bids.neuroimaging.io/>) practice and am attempting to organise an EEG dataset with four separate files per subject which reflect different e...
- BIDS-validator Error 1: [Code 0] INTERNAL ERROR**
bids-validator
Aug '20 - Hello everyone, I tried to validate OASIS3 dataset and faced with the problem. Below the error output of the online `bids-validator`: Error 1: [Code 0] INTERNAL ERROR Internal error. SOME VALIDATION STEPS MAY NOT HAVE OCCURRED! file NaN KB | Location:Reason: Maximum call...
- Bids-validator errors**
bids
Sep '19 - Dear all, I really appreciate your help. I used https://1526-37161308-gh.circle-artifacts.com/0/root/web_version/index.html (<https://bids-standard.github.io/bids-validator/> gave additional errors...) `Nifti_errors (5).txt` (32.3 KB) 57 File Path: Not all subjects contain the same files. Eac...
- Bids-database-dir ValueError**
pybids
Jan '22 - ...multi-echo resting-state data. The data lives on a network server but the processing is done by the workstation not a HPC. I created a separate test `bids` directory with one subject and successfully processed their data through fmriprep. Using the same command on the path to the actual directory but sti...

<https://neurostars.org/search?q=bids>

obrigado pela vossa atenção

Q&A