UNIVERSITY OF SURREY

ORGANISATION AND SHARING OF NEUROIMAGING DATA

BIDS and pieces – the theory

Tibor Auer

INTRODUCTION

- **Who is Tibor?**
  - Lecturer in Biological Psychology and Research Methods
- **My journey**
  - MD
    - Undergraduate research in neuropathology
  - Clinical and cognitive neuroscientist
    - PhD in clinical neuroscience: neuroimaging techniques in various clinical conditions
    - Post-doctoral experience in neuromodulation intervention (NFT, tACS)
  - Neuroinformatician (research computing, scalable pipelines, open-source development)
    - Developing, training, and supporting robust and reliable methods

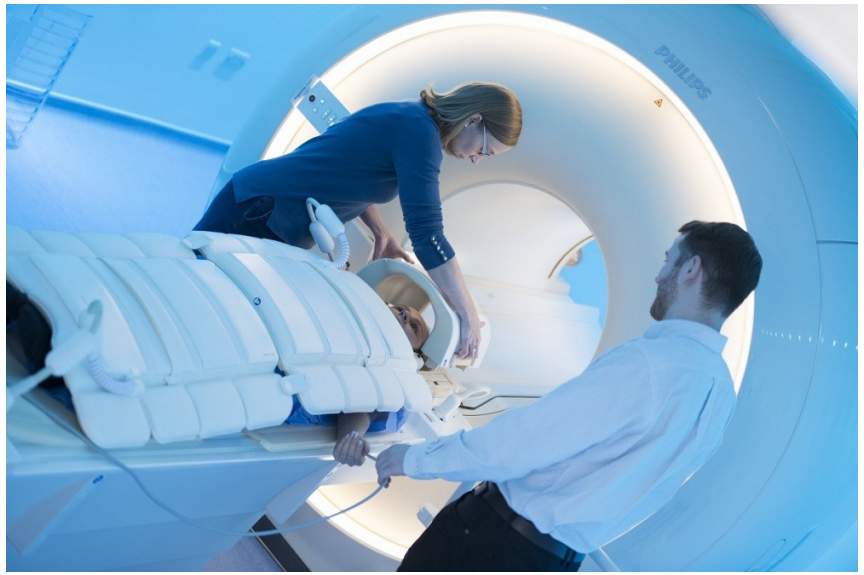
LEARNING OUTCOMES

- **Understand BIDS**
  - What it is
  - What the key components are
  - Where you can get help

- **Want to use BIDS**
  - Why you should use it
  - How it can help your research

LEARNING OUTCOMES

- **BIDS raw data for MRI**
  - fMRI (task and rest)
  - sMRI (qMRI)
  - DW-MRI
  - ASL

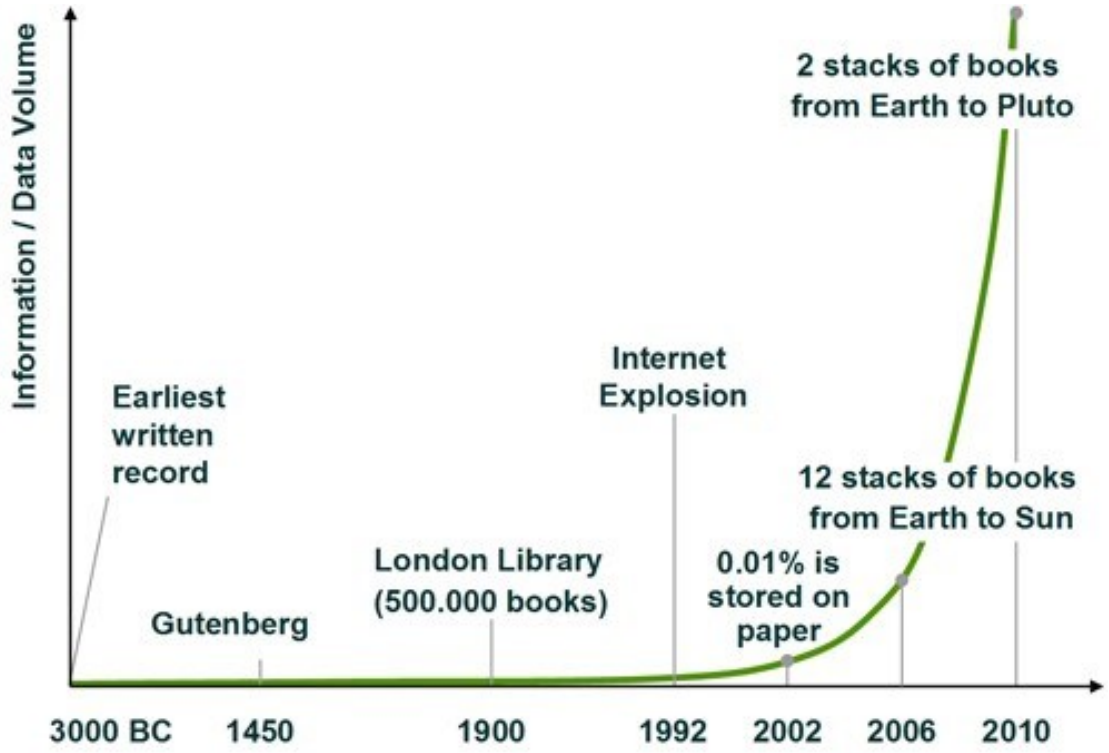


- **BIDS for …**
  - MEG, EEG, iEEG
  - PET
  - NIRS
  - Microscopy
  - Physiology
  - Psychophysics
  - Genetics

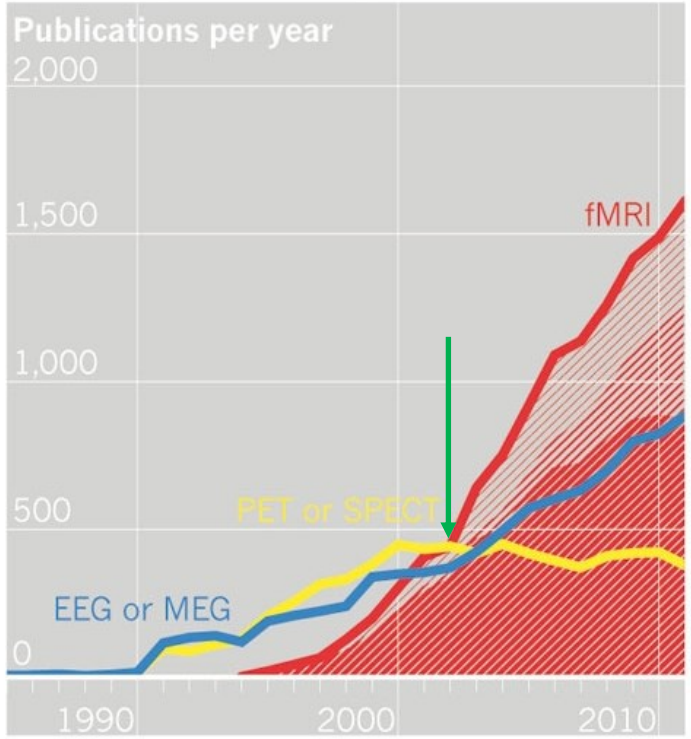
- *Derived data*

CHALLENGE

- **Information explosion**
  - Neuroimaging field generates an **increasing amount of data**



information explosion | Ethos3



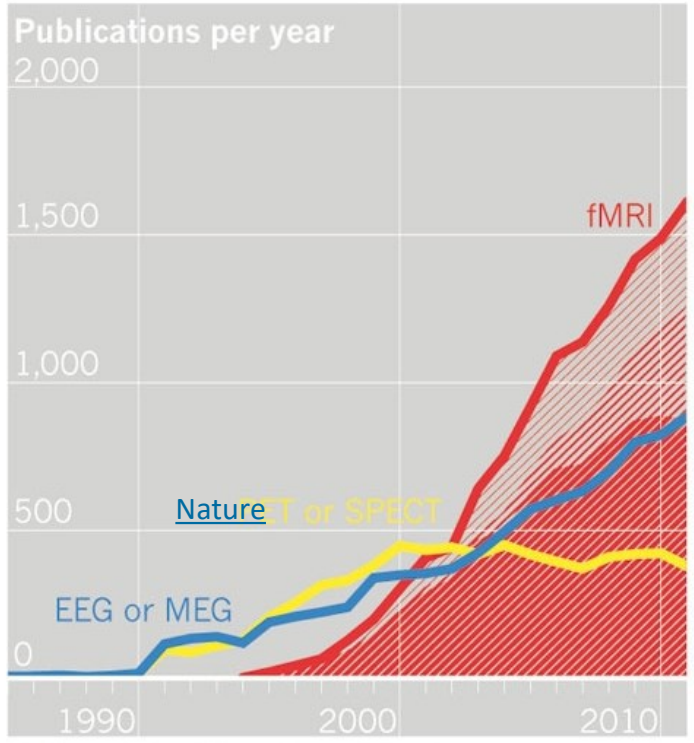
Nature

CHALLENGE

- **Information explosion**
  - Neuroimaging field generates an **increasing amount of data**
  - Most neuroimaging data are **complex** (i.e., multidimensional, hierarchical, structured)
  - No consensus on how to represent **common concepts and similarities**, e.g., in experimental designs and data types

↓

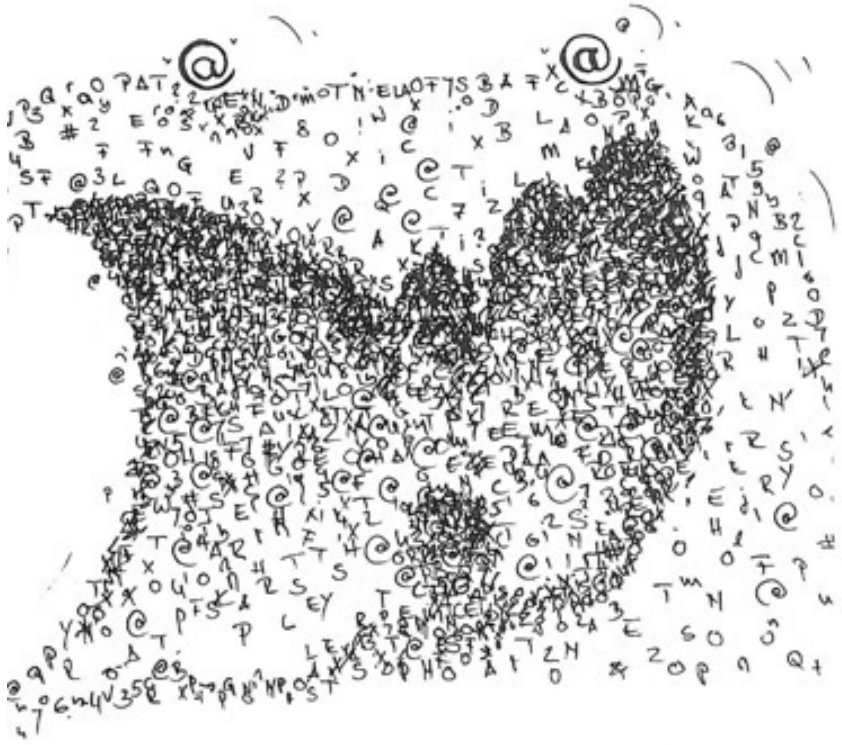
  - Software tools and researcher organize and describe their data **in their own way**



CHALLENGE

- **Fragmented data management**
  - Sharing requires rearranging/reformatting the data
  - Validating (accuracy, completeness) against what?

- **Data-unaware workflows**
  - Requires manual input
  - Analysis scripts must be adapted to the data
  - + Complex workflows (integrating different tools)



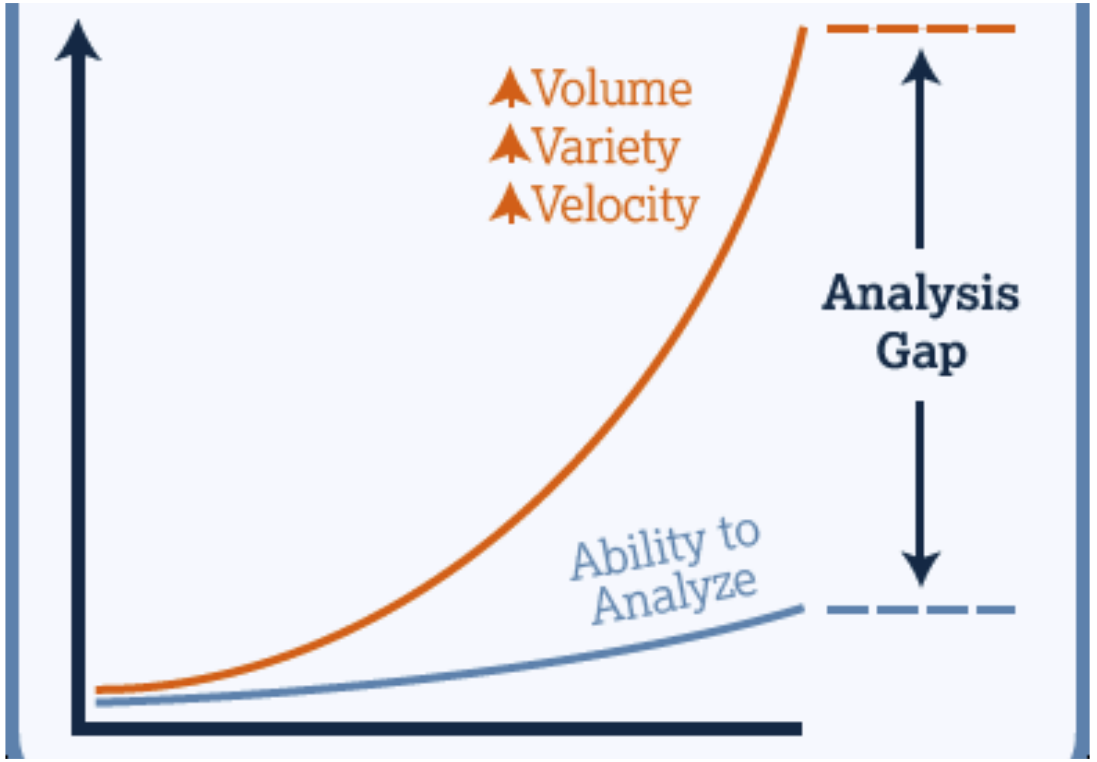
sonian.com

CHALLENGE

- **Reduced efficiency**
- **Lack of transparency (data and workflow)**
  - Difficult to document
  - Harder to detect errors

↓

Lack of reproducibility

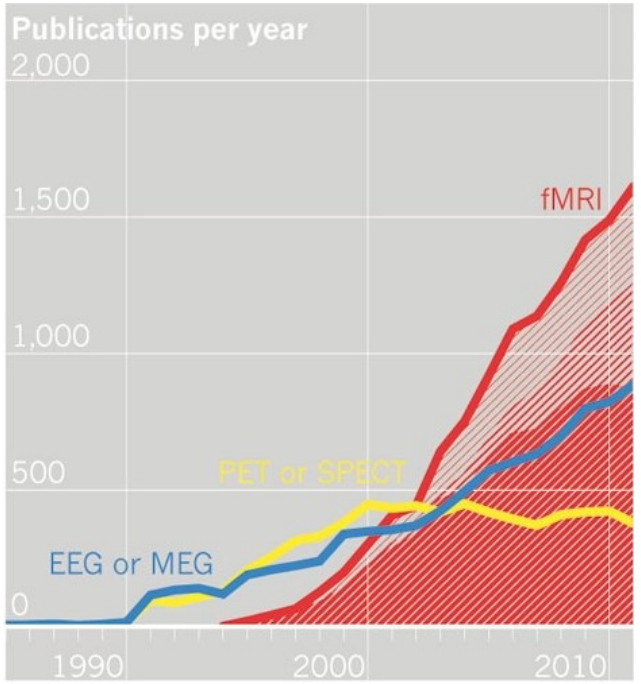


Big Data Driving Business | Segittur

CHALLENGE

- **Lack of reproducibility → Translation gap**
  - Increasing volume of neuroimaging findings in mental health

  - Limited success in, e.g., identifying reliable biomarkers
    (Nikolaidis et al., 2022, Botvinik-Nezer and Wager, 2022)
    - Limited guidance in methodology
      ↓
    - Unstable and **irreproducible** workflows
      ↓
    - Confounded variability
    - Limited confidence in the accuracy



[Smith, 2012]

Neuroimaging/psychology (Aarts et al., 2015)

Cancer research (Begley and Ellis, 2012)

Social sciences (Breznau et al., 2022)

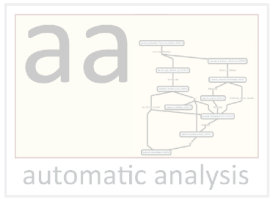
...

SOLUTIONS

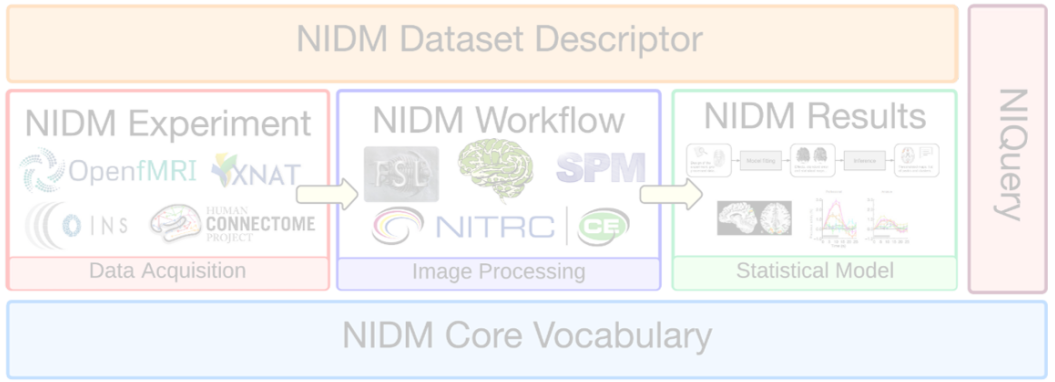
Field-specific solutions

- **Harmonization of neuroimaging**
  - Data management (e.g., BIDS, DataLad)

  - Data processing and analysis (e.g., aa, reproa, nipype)



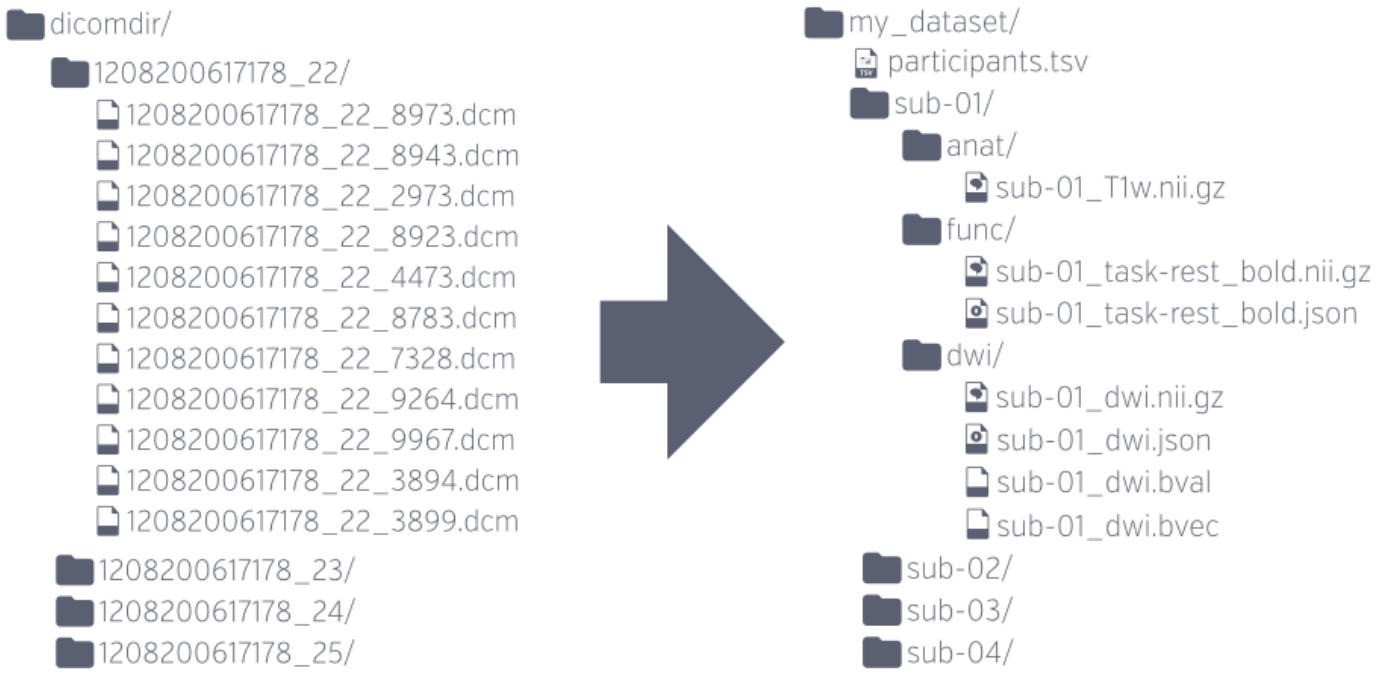
  - Reporting methodology (e.g., COBIDAS, CRED-nf, ContES)
  - Recording provenance (e.g., NIDM, BIDS-Prov)



BIDS

Introduction

- **Brain Imaging Data Structure (BIDS)**
  - Standard for organizing results of a human neuroimaging experiment.
  - http://bids.neuroimaging.io, **https://bids-specification.readthedocs.io**

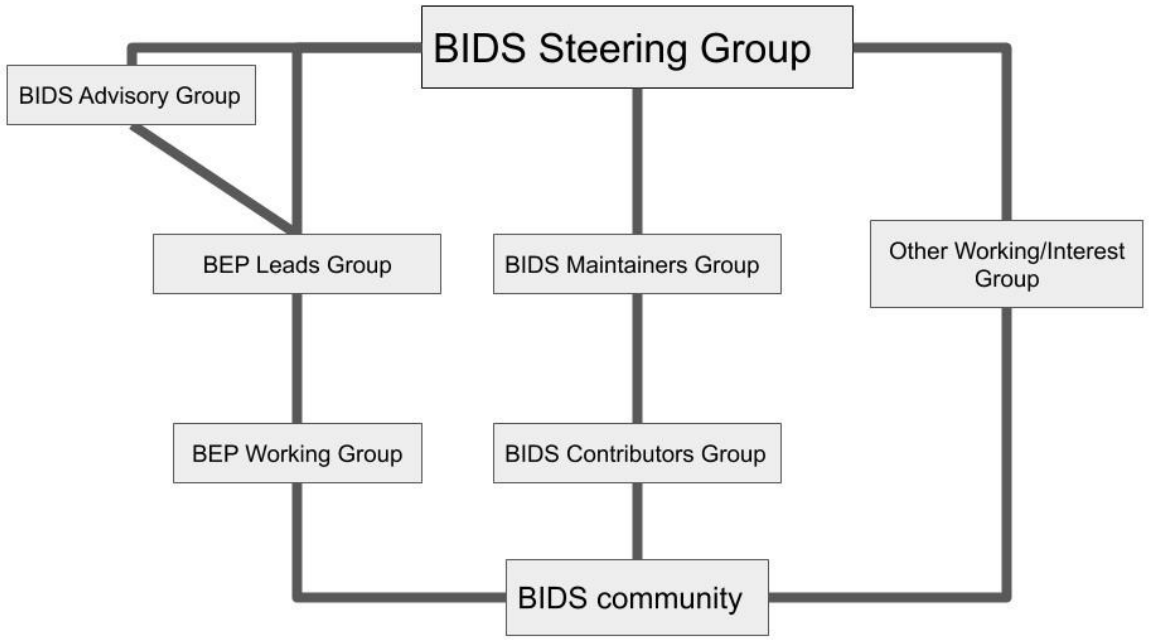


BIDS

Introduction

- **Brain Imaging Data Structure (BIDS)**
  - Standard for organizing results of a human neuroimaging experiment.
  - http://bids.neuroimaging.io, **https://bids-specification.readthedocs.io**

- **Governance**
  - Developed by the **community**
  - Managed by the Maintainers Group
  - Approved by the Steering group
  - Endorsed by the INCF



BIDS

Introduction

- **Advantages for**
  - User: Software aware of the data structure → less manual entry
  - Group leader: More than one person working on the same data over time
  - Developer: Data structure can be expected
  - Database: Easier to manage and share data
    - Most databases already accept BIDS: OpenNEURO
    - Some journals require data sharing
    - Validator tool

- ✓ Efficiency
- ✓ Transparency
- ✓ Reproducibility

BIDS

Principles

- **Definitions**
  - Organisation levels (folder/file structure)
    - Subject: “A person or animal participating in the study (participant).”
    - Session: “A logical grouping of neuroimaging and behavioural data consistent across subjects.”
    - Run: “An uninterrupted repetition of data acquisition with the same acquisition parameters.”
    - Data type: “A functional group of different types of data.” **≠ modality**
      - *anat, beh, dwi, eeg, fmap, func, meg, perf, …*

BIDS

Principles

- **Definitions**
  - Identifiers (folder/file name elements)
    - Suffix: Can correspond to modality, acquisition type, group, …
      - E.g.: T1w, bold, dwi, eeg, meg, events, physio
    - Entities: key-value pairs
      - Name: Context
      - Key: Compression of the entity name
      - Value: Specifier
        - Index: “A nonnegative integer with consistent indentation.” **ordinal rather than scalar**
        - Labels: An alphanumeric value
      - E.g.: Subject (Name): “sub-control01”

BIDS

Principles

- **Motivations**
  - Some metadata is better than no metadata
  - Don’t rely on external software or complicated file formats
    - Clause for acquisitions: keep the original or use legacy/common format
  - Aim to capture most of experiments but give space to **extend** the standard format

BIDS

Implementation

- **Some metadata is better than no metadata (+ redundancy)**
  - Folder structure
  - Filename

```
sub-control01/
  anat/
    sub-control01_T1w.nii.gz
    sub-control01_T1w.json
    sub-control01_T2w.nii.gz
    sub-control01_T2w.json
  func/
    sub-control01_task-nback_bold.nii.gz
    sub-control01_task-nback_bold.json
    sub-control01_task-nback_events.tsv
    sub-control01_task-nback_cont-physio.tsv
    sub-control01_task-nback_cont-physio.json
    sub-control01_task-nback_sbref.nii.gz
  dwi/
    sub-control01_dwi.nii.gz
    sub-control01_dwi.bval
    sub-control01_dwi.bvec
  fmap
    sub-control01_phasediff.nii.gz
    sub-control01_phasediff.json
    sub-control01_magnitude1.nii.gz
  sub-control01_scans.tsv
README
CHANGES
dataset_description.json
participants.tsv
```

BIDS

Implementation

- **Don’t rely on external software or complicated file formats**
  - JSON-files for key-value pairs (MUST have for some acquisitions, e.g., fMRI)

```
{
    "RepetitionTime": 3.0,
    "EchoTime": 0.03,
    "FlipAngle": 78,
    "SliceTiming": [0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4,
                    1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8],
    "InPlanePhaseEncodingDirection": "AP“
    "TaskName": “nback"
}
```

BIDS

Implementation

- **Don’t rely on external software or complicated file formats**
  - Use of tab separated files for tabular data (participant, sessions, scans and event descriptor).

```
participant_id age sex
sub-001        34  M
Sub-002        12  F
Sub-003        33  F
```

```
filename    acq_time
func/sub-control01_task-nback_bold.nii.gz   1877-06-15T13:45:30
func/sub-control01_task-motor_bold.nii.gz   1877-06-15T13:55:33
meg/sub-control01_task-rest_split-01_meg.nii.gz 1877-06-15T12:15:27
meg/sub-control01_task-rest_split-02_meg.nii.gz 1877-06-15T12:15:27
```

```
onset  duration trial_type ResponseTime
1.2    0.6      go         1.435
5.6    0.6      stop       1.739
…
```

BIDS

Implementation

- **Don’t rely on external software or complicated file formats**
  - Use of legacy text file formats for b vectors/values

```
-0.552828 0.0480948 0.980937 -0.24275 -0.292642 0.085518 0.470646
-0.865701 0.226775 0.334443 0.727534 -0.552828 -0.625823 0.353667 -
0.853084 0.516586 -0.766369 -0.125754 -0.149251 0.937341 -0.124163
-0.36458 -0.552828 0.579767 0.445586 -0.905294 -0.282606 -0.731609
0.676757 0.171374 -0.59121 -0.46017 0.845645 -0.552828 -0.130717
0.975624 -0.288147 0.655193 -0.442479 -0.471845 0.638596 -0.642432
0.850936 -0.240808 -0.552828 -0.578162 0.100487 -0.924592 -
0.0210952 -0.764669 0.197294 0.405727 0.859032 0.380277 -0.6891 -
0.552828 0.430722 0.0366712 -0.540564 0.775224 0.0646129 -0.978577
0.199971 0.678143 -0.448761 0.849148 -0.552828
```

BIDS

Implementation

- **Don’t rely on external software or complicated file formats**
  - Keep the original or use legacy/common format for acquisitions
    - (MRI) imaging data: NIfTI (can be gzipped)
    - MEG data: native format
      - .ds folder for CTF
      - (split) .fif for Elekta
    - EEG data: one of the common formats
      - **EDF, BrainVision**, EEGLAB, Biosemi
    - Behavioural: TSV

BIDS

Special cases

- **ASL**
  - Data:
    - *_asl.nii[.gz]: **4D data**
    - *_m0scan.nii[.gz]: calibration scan (can be part of the *_asl)
    - *_aslcontext.tsv: annotation of volumes

```
volume_type
control
label
…
```

  - Metadata (MUST have): JSON

```
{
    "ArterialSpinLabelingType": "PCASL",
    "PostLabelingDelay": 0,
    "BackgroundSuppression": "true",
    "M0Type": "Separate",
    "LabelingDuration": 1,
    "SliceTiming": [
            0.1,
            0.2,
            0.3,
            …
        ],
    "RepetitionTimePreparation": 4.55
}
```

BIDS

Special cases

- **Physiology**
  - Data: **headerless gzipped** TSV

```
34      110      0
44      112      0
23      100      1
…
```

  - **Required** metadata fields
    - SamplingFrequency: in Hz
    - StartTime: in seconds, in relation to the start of neural data acquisition (can be negative)
  - Metadata (MUST have): JSON

```
{
    "SamplingFrequency": 100.0,
    "StartTime": -22.345,
    "Columns": ["cardiac", "respiratory“, “trigger”],
    "cardiac": {
        "Description": "continuous pulse measurement",
        "Units": "mV"
        },
    "respiratory": {
        "Description": "continuous breathing measurement",
        "Units": "mV"
        },
    “trigger": {
        "Description": "continuous measurement of the scanner pulse",
        "Units": “arbitrary"
        }
}
```

BIDS

Implementation

- **Aim to capture most of experiments but give space to extend the standard format**
  - Make certain folder hierarchy levels optional for simplicity
    - e.g.: single session, inherited parameters and events
  - Supports multiple types of anatomical scans
  - Supports fMRI: both task based and resting state
  - Supports sparse fMRI(via slice timing)
  - Supports multiple fieldmap formats for dual-echo and dual phase-encoding direction (top-up)
  - Allows for arbitrary files not covered by the spec to be included
  - **BIDS Extension Proposal – Check first before going creative!**
    https://bids.neuroimaging.io/get_involved.html#extending-the-bids-specification
    https://bids-extensions.readthedocs.io

BIDS

Support

- **Official BIDS Starter Kit: https://bids-standard.github.io/bids-starter-kit**
  - Videos, tutorials
  - Pointers to resources

- **Look for examples in databases**
  - COINS, LORIS , **OpenfMRI**, SciTran , XNAT, …

- **Community**
  - Neurostars: https://neurostars.org/tag/bids
  - Google Mailing List: https://groups.google.com/g/bids-discussion

BIDS

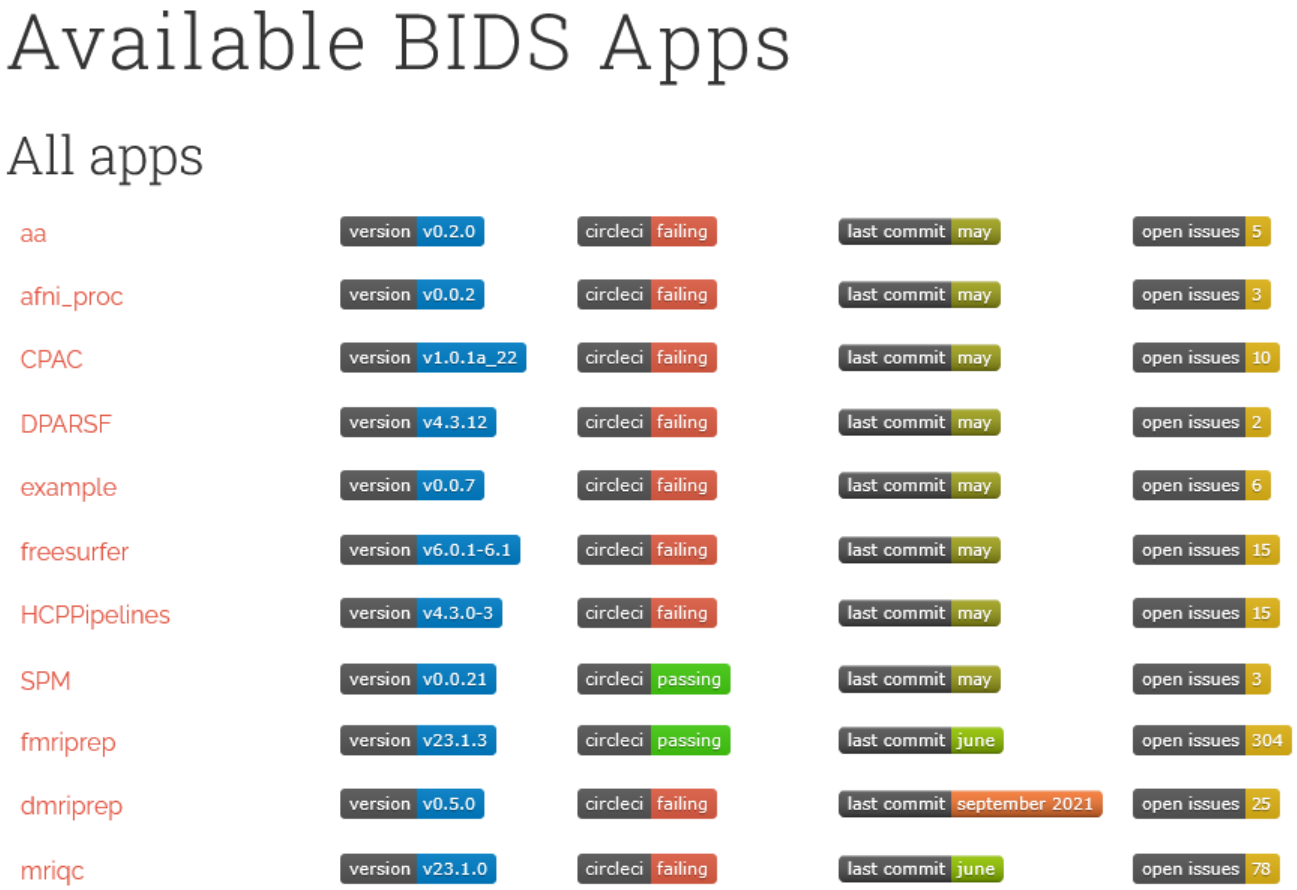
Support

- **Tools and pipelines**
  - Conversion
    - Comprehensive solution: HeuDiConv – works best for site with naming conventions
    - Converters producing BIDS-compatible metadata: dcmi2nix – allows more flexibility
  - Validation: https://bids-standard.github.io/bids-starter-kit/validator.html
    - Browser-based: https://bids-standard.github.io/bids-validator
    - Command line tool: https://github.com/bids-standard/bids-validator
  - Processing
    - SPM, EEGLAB, FieldTrip, …
    - Automatic Analysis, C-PAC, Nipype, Reproducibility Analysis, …
    - BIDS Apps: http://bids-apps.neuroimaging.io

BIDS

Support

- **BIDS Apps**
  - A container image
    - Contains the whole pipeline
    - Contains all dependencies
    - Preconfigured and ready-to-run
      - Docker – PC
      - Singularity – Cluster
  - Standardised
    - BIDS dataset as input
    - Same core set of arguments
  - Open with version control
      - https://github.com/BIDS-Apps
      - https://hub.docker.com/r/bids



QUESTIONS?