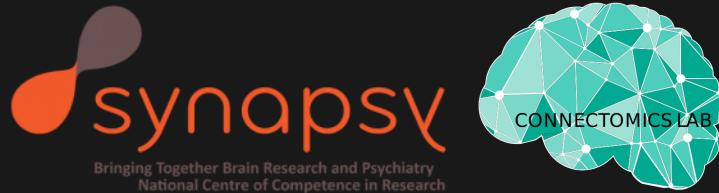


HARMONIZED NEUROIMAGING DATA MANAGEMENT IN SYNAPSY

Synapsy DMP Training @ Campus Biotech, Geneva

by [Sebastien Tourbier](#)

March 08, 2021



BIDS

BRAIN IMAGING DATA STRUCTURE

WHAT IS BIDS?

SPECIFICATIONS TO ORGANIZE AND DESCRIBE NEUROIMAGING DATA

SCIENTIFIC DATA

A graphic consisting of four rows of binary digits (0s and 1s) in blue, arranged vertically. The rows are: 110110, 0111101, 11011110, and 011101101.

OPEN

SUBJECT CATEGORIES

- » Data publication and archiving
- » Research data

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

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

Krzysztof J. Gorgolewski¹, Tibor Auer², Vince D. Calhoun^{3,4}, R. Cameron Craddock^{5,6}, Samir Das⁷, Eugene P. Duff⁸, Guillaume Flandin⁹, Satrajit S. Ghosh^{10,11}, Tristan Glatard^{7,12}, Yaroslav O. Halchenko¹³, Daniel A. Handwerker¹⁴, Michael Hanke^{15,16}, David Keator¹⁷, Xiangrui Li¹⁸, Zachary Michael¹⁹, Camille Maumet²⁰, B. Nolan Nichols^{21,22}, Thomas E. Nichols^{20,23}, John Pellman⁶, Jean-Baptiste Poline²⁴, Ariel Rokem²⁵, Gunnar Schaefer^{1,26}, Vanessa Sochat²⁷, William Triplett¹, Jessica A. Turner^{3,28}, Gaël Varoquaux²⁹ & Russell A. Poldrack¹

<https://doi.org/10.1038/sdata.2016.44>

COMPREHENSIBLE ORGANIZATION AND NAMING WITH WELL-ACCEPTED FORMATS

dicomdir/

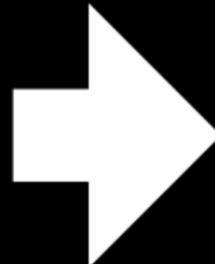
1208200617178_22/

- 1208200617178_22_8973.dcm
- 1208200617178_22_8943.dcm
- 1208200617178_22_2973.dcm
- 1208200617178_22_8923.dcm
- 1208200617178_22_4473.dcm
- 1208200617178_22_8783.dcm
- 1208200617178_22_7328.dcm
- 1208200617178_22_9264.dcm
- 1208200617178_22_9967.dcm
- 1208200617178_22_3894.dcm
- 1208200617178_22_3899.dcm

1208200617178_23/

1208200617178_24/

1208200617178_25/



my_dataset/

participants.tsv

sub-01/

anat/

sub-01_T1w.nii.gz

func/

sub-01_task-rest_bold.nii.gz

sub-01_task-rest_bold.json

dwi/

sub-01_dwi.nii.gz

sub-01_dwi.json

sub-01_dwi.bval

sub-01_dwi.bvec

sub-02/

sub-03/

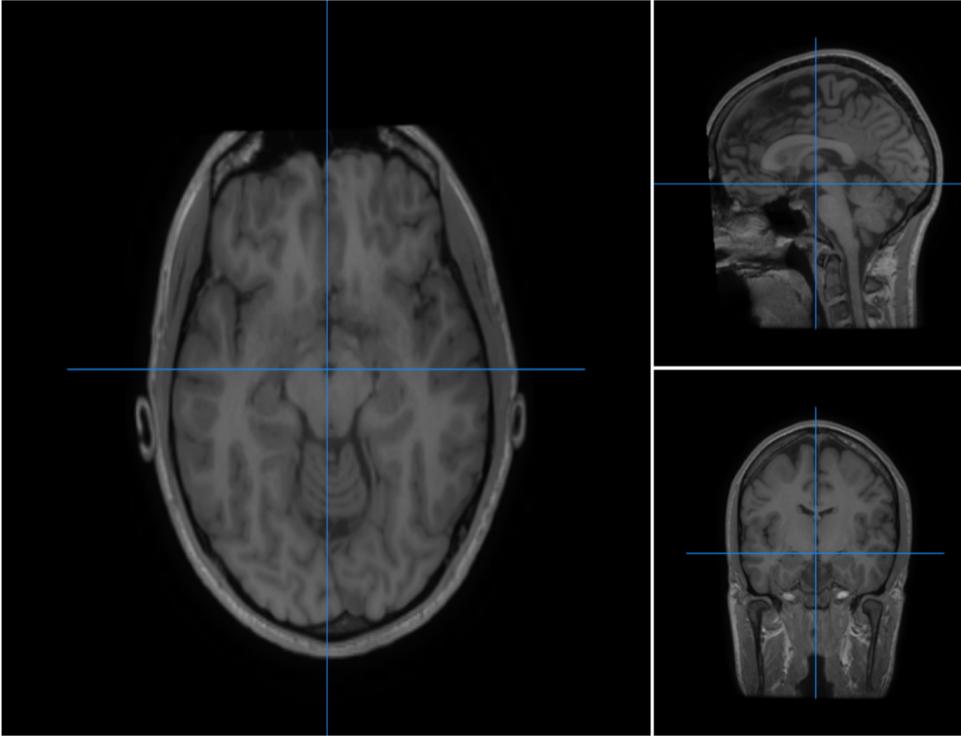
sub-04/

**DESIGNED WITH THE FINDABLE,
ACCESSIBLE, INTER-OPERABLE AND
RESEARCHABLE (FAIR) PRINCIPLES IN
MIND.**

<https://www.go-fair.org/fair-principles/>

IMAGE WITH ACQUISITION METADATA

DS003505 - SUB-01 - ANAT - SUB-01_T1W.NII.GZ [DOWNLOAD](#)



DS003505 - SUB-01 - ANAT - SUB-01_T1W.JSON [DOWNLOAD](#)

Tree

```
▼ root: {} 20 keys
  Manufacturer: "GE"
  ManufacturersModelName: "DISCOVERY_MR750"
  AcquisitionNumber: 1
  InstitutionName: "HFR_Hopital_Cantonal_Fribourg"
  DeviceSerialNumber: "000000475290MR01"
  SoftwareVersions: "25_LX_MR_Software_release:DV25.0_R01_1451.a"
  ScanningSequence: "GR"
  SequenceVariant: "SS_SK"
  SeriesDescription: "COR_FSPGR_BRAVO"
  ProtocolName: "COR_FSPGR_BRAVO"
▶ ImageType: [] 3 items
  AcquisitionTime: "19:04:55.000000"
  MagneticFieldStrength: 3
  FlipAngle: 9
  EchoTime: 0.002812
  RepetitionTime: 0.007276
  InversionTime: 0.9
  PhaseEncodingLines: 256
  ConversionSoftware: "dcm2niix"
  ConversionSoftwareVersion: "v1.0.20170624 (OpenJPEG build) MSC1900"
```

DATASET DOCUMENTATION AND METADATA

VEPCON: Source imaging of high-density visual evoked potentials with multi-scale brain parcellations and connectome:

[+ ADD FILES](#) [+ ADD DIRECTORY](#) [DELETE](#)

CHANGES

[DOWNLOAD](#) [VIEW](#) [UPDATE](#) [DELETE](#)

dataset_description.json

[DOWNLOAD](#) [VIEW](#) [UPDATE](#)

participants.json

[DOWNLOAD](#) [VIEW](#) [UPDATE](#) [DELETE](#)

participants.tsv

[DOWNLOAD](#) [VIEW](#) [UPDATE](#) [DELETE](#)

README

[DOWNLOAD](#) [VIEW](#) [UPDATE](#) [DELETE](#)

DS003505 - DATASET_DESCRIPTION.JSON [DOWNLOAD](#)

Tree

```

▼ root: {} 10 keys
  Name: "VEPCON: Source imaging of high-density vis
  with multi-scale brain parcellations and connec
  BIDSVersion: "1.4.2"
  License: "CC0"
  ▶ Authors: [] 6 items
  Acknowledgements: "say here what are your acknowl
  HowToAcknowledge: "Please cite the associated dat
  ▶ Funding: [] 1 item
  ▶ ReferencesAndLinks: [] 3 items
  DatasetDOI: "10.18112/openneuro.ds003505.v1.0.0"
  ▶ EthicsApprovals: [] 1 item

```

README

VEPCON: Source imaging of high-density visual evoked potentials with multi-scale brain parcellations and connectomes

OVERVIEW

The multimodal dataset VEP CON follows the BIDS standard. It contains high-density EEG, structural MRI and diffusion weighted MRI data. The dataset includes participants and derivatives.

Visual evoked potentials were recorded while participants viewed faces from scrambled faces, or coherently moving faces. Diffusion weighted MRI (DWI) were recorded in a separate session from the visual evoked potential recording.

The dataset contains pre-processed EEG of single subjects, structural MRIs, individual brain parcellations (regions), and corresponding structural connectomes. It includes measures of average fractional anisotropy and mean diffusion tensor eigenvalues, inverse solutions for source imaging based on independent component analysis, and code for deriving time-series of activity in each region.

This dataset can contribute to multimodal methods development, studying structure-function relations, as well as unimodal optimization of source imaging and graph analysis among many other possibilities.

2020-10-11

Initial upload of raw and cmp-derived MRI data

2020-14-12

Update with defaced raw and cmp-derived MRI data

2020-14-12

Upload of EEG data

2021-03-02

Upload of MRIQC derivatives for quality assessment of T1w MRI data

2021-04-02

Update main README, CHANGES and dataset_description.json

DS003505 - PARTICIPANTS.JSON

[DOWNLOAD](#)

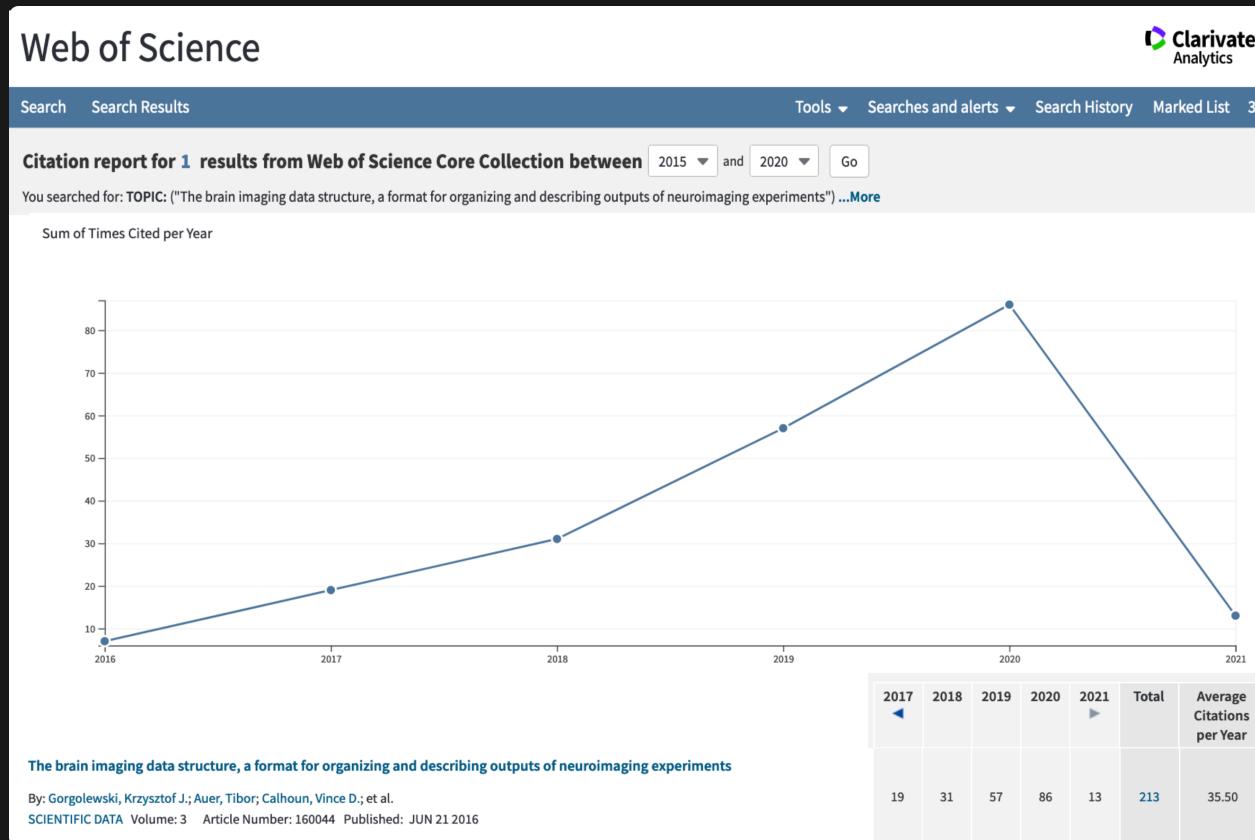
Tree

DS003505 - PARTICIPANTS.TSV

PARTICIPANT... SEX AGE

sub-01	M	26
sub-02	M	29
sub-03	F	23
sub-04	M	29
sub-05	F	20
sub-06	F	23
sub-07	F	32
sub-08	F	20
sub-09	F	20
sub-10	F	22
sub-11	F	24
sub-12	F	26
sub-13	F	24
sub-14	F	23
sub-15	F	22

STANDARD ADOPTED BY A GROWING NUMBER OF RESEARCHERS



Results from www.webofknowledge.com (Date: March 07, 2021)

**BIDS OFFICIAL
WEBSITE**

HOW IS BIDS USEFUL?

**MAKE DATA FULLY UNDERSTANDABLE BY
ITSELF THANKS TO ITS METADATA AND
DOCUMENTATION FILES**

**FACILITATE DATA SHARING BETWEEN LAB
MEMBERS AND COLLABORATORS PART
OF SYNAPSY**

**MAKE CODE INTEROPERABLE BETWEEN
PROJECTS, LAB MEMBERS, AND
COLLABORATORS PART OF SYNAPSY**

VERY LITTLE EFFORT TO PUBLISH DATASET TO DATABASES

The screenshot shows the OpenNeuro website. The header includes the OpenNeuro logo, navigation links for 'MY DASHBOARD', 'PUBLIC DASHBOARD', 'SUPPORT', 'FAQ', and 'UPLOAD DATASET', and a sub-header for 'Stanford Center for Reproducible Neuroscience'. The main content area features a section about 'Brain Imaging Data Structure (BIDS)' with a link to 'BIDS specifications'.

A free and open platform for sharing MRI, MEG, EEG, iEEG, and ECoG data

View more information about
[Stanford Center for Reproducible Neuroscience](#)

Brain Imaging Data Structure (BIDS)

A Validator for the Brain Imaging Data Structure
Read more about the [BIDS specifications](#)

Want to contribute to BIDS?
Visit the [Google discussion group](#) to contribute.

Databases such as [OpenNeuro](#), [LORIS](#), [COINs](#), [XNAT](#), [SciTran](#) and others **accept and export datasets organized following BIDS**

BENEFIT OF DEDICATED AND WELL DOCUMENTED TOOLS

FOR BIDS DATASET CREATION

HeuDiConv

a heuristic-centric DICOM converter



About

`heudiconv` is a flexible DICOM converter for organizing brain imaging data into structured directory layouts.

- it allows flexible directory layouts and naming schemes through customizable heuristics implementations
- it only converts the necessary DICOMs, not everything in a directory
- you can keep links to DICOM files in the participant layout
- using dcm2niix under the hood, it's fast
- it can track the provenance of the conversion from DICOM to NIfTI in W3C PROV format
- it provides assistance in converting to [BIDS](#).
- it integrates with [DataLad](#) to place converted and original data under git/git-annex version control, while automatically annotating files with sensitive information (e.g., non-defaced anatomicals, etc)

<https://github.com/nipy/heudiconv>

**FOR VALIDATION AND
DATA CURATION
SUPPORT**

FOR DATASET QUERY

- In Python: <https://github.com/bids-standard/pybids>

```
>>> from bids import BIDSLayout  
>>> layout = BIDSLayout('/home/data/ds000117')  
>>> layout.get_subjects()
```

- In Matlab: <https://github.com/bids-standard/bids-matlab>

```
>>> BIDS = bids.layout('/home/data/ds000117');  
>>> bids.query(BIDS, 'subjects')
```

FOR ANALYSIS

A number of processing pipelines handling BIDS datasets (BIDS Apps) are available, ranging from quality control to preprocessing, connectome mapping, and statistical analysis - and maybe one of yours in the future!



BIDS Apps

portable neuroimaging pipelines that
understand BIDS datasets

[About](#)

[Tutorials](#)

[Apps](#)

poldracklab/mriqc

version v0.16.1

open issues 131

BIDS-Apps/QAP

version v0.0.1

open issues 5

BIDS-Apps/CPAC

version v1.0.1a_22

open issues 10

BIDS-Apps/hyperalignment

version v0.0.5

open issues 0

BIDS-Apps/mindboggle

version v0.0.4-1

open issues 8

BIDS-Apps/MRtrix3_connectome

version v0.5.0

open issues 18

BIDS-Apps/rs_signal_extract

version v0.1

open issues 0

BIDS-Apps/aa

version v0.2.0

open issues 2

BIDS-Apps/niak

version v1.0

open issues 3

BIDS-Apps/oppni

version v0.7.0-1

open issues 5

poldracklab/fmriprep

version v20.2.1

open issues 226

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

HOW TO GET STARTED WITH BIDS?

ONLINE BIDS SPECIFICATIONS

OFFICIAL BIDS TUTORIALS

Heudiconv Example



Search or jump to... | Pull requests Issues Marketplace Explore

bids-standard / bids-starter-kit

Code Issues 27 Pull requests 2 Actions Projects 1 Wiki

Tutorials

James Kent edited this page on 26 Feb 2020 · 7 revisions

BIDS Tutorial Series:

- Part 1a: Introductory Walkthrough
- Part 1b: Automate the Introductory Walkthrough
- Part 2a: HeuDiConv Walkthrough

Video tutorials

- "DICOM to BIDS conversion" by "math et al"
- "Heudiconv Example" by James Kent

<https://github.com/bids-standard/bids-starter-kit/wiki/Tutorials>

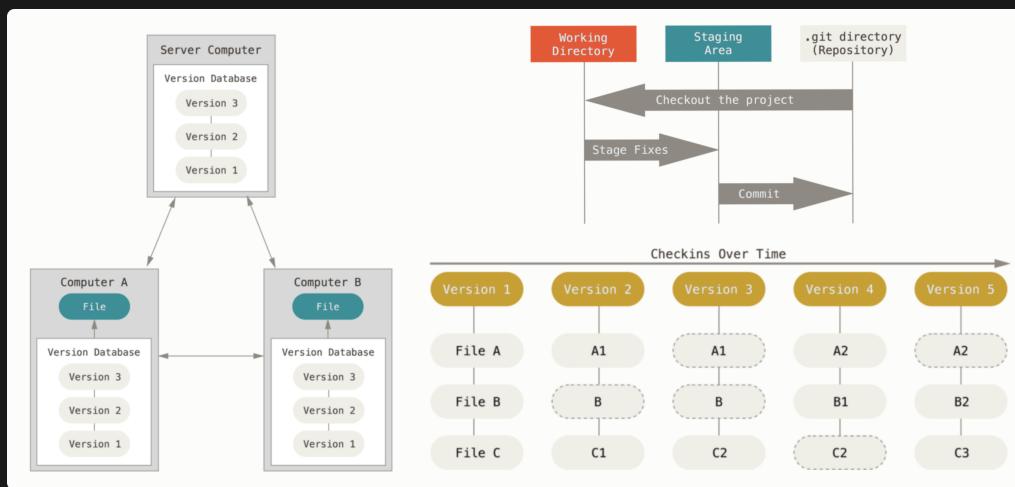
DATALAD

WHAT IS DATALAD?

DISTRIBUTED DATA MANAGEMENT SYSTEM

- Built on top of Git and git-annex

<https://git-scm.com/book/en/v2>



It allows you to keep track of dataset with large file content as you do with text file in Git.

- But it much more than that!

WHAT IS DATA LAD CAPABLE OF?

KEEP TRACK OF YOUR DATASET HISTORY

KEEP TRACK OF YOUR DATASET HISTORY

- Create an empty DataLad dataset:

```
datalad create (-c yoda, -c text2git)
```

A dataset has a history to track files and their modifications that is explored with Git:

```
git log
```

- Record with a descriptive message the dataset or file state to the history

```
datalad save -m "messsage"
```

Concise commit messages should summarize the change for future you and others.

- Report the current dataset state:

```
datalad status
```

A clean status is good practice.

DATASET CONSUMPTION AND SHARING

DATASET CONSUMPTION AND SHARING

- Install a DataLad dataset:

```
datalad clone
```

A dataset (sub-dataset) can be installed inside a super-dataset (dataset nesting):



Only small files and metadata about file availability are present locally after its installation.

- To download the actual content of large files:

```
datalad get
```

It allows us to efficiently handle large datasets and download file content on demand.

DATASET CONSUMPTION AND SHARING

- Publish your dataset to a remote dataset repository:
 - Create a dataset sibling on a UNIX-like Shell (local or SSH)-accessible machine:

```
datalad create-sibling
```

It creates a remote dataset repository and configures it as a dataset sibling to be used as a publication target.

- Publish the dataset:

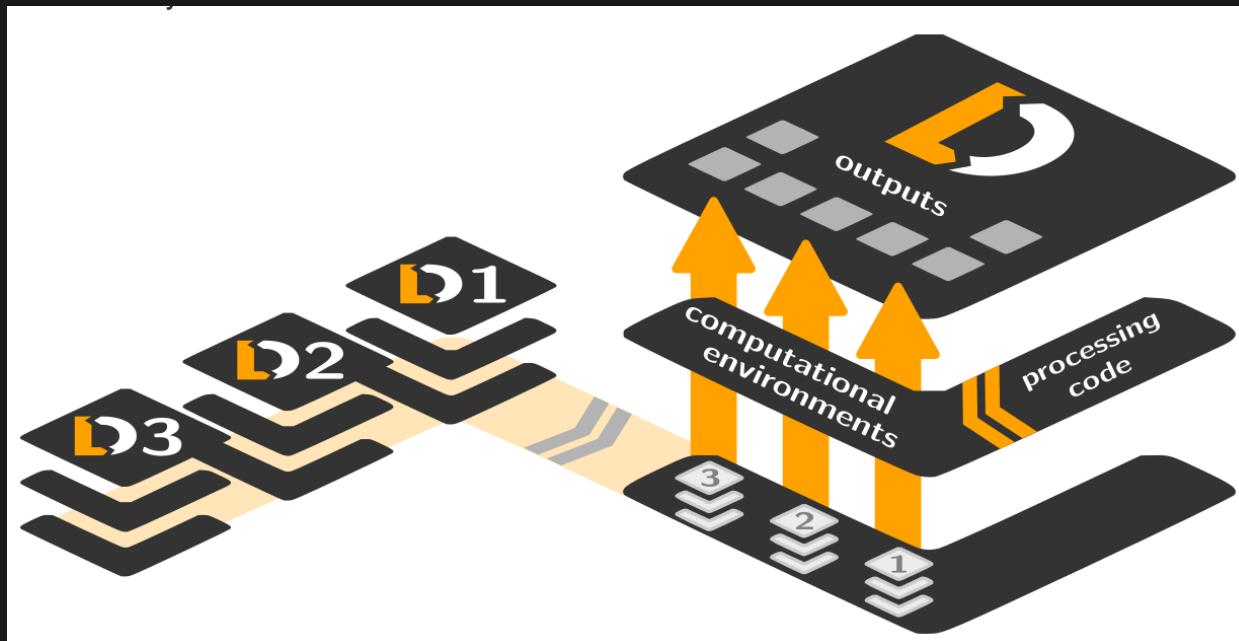
```
datalad push
```

It updates all your local changes saved and annexed data to the remote dataset repository.

COMPUTATIONALLY REPRODUCIBLE DATASET ANALYSIS

COMPUTATIONALLY REPRODUCIBLE DATASET ANALYSIS

- DataLad can track source code and computing environment that led to the generation of study results as well as all the generated outputs



<http://www.repronim.org/ohbm2018-training/03-01-reproin/>

(COMPUTATIONALLY) REPRODUCIBLE DATASET ANALYSIS

- Execute and track input, output and source code:

```
datalad run
```

It links datasets (as subdatasets) and source code, records data origin and command execution, and collect and store provenance of all contents of a dataset created.

- The analysis step can be re-executed with:

```
datalad rerun
```

COMPUTATIONALLY REPRODUCIBLE DATASET ANALYSIS

- Execute and track input, output, source code, and computing environment (in the form of software containers) with the datalad-containers extension:

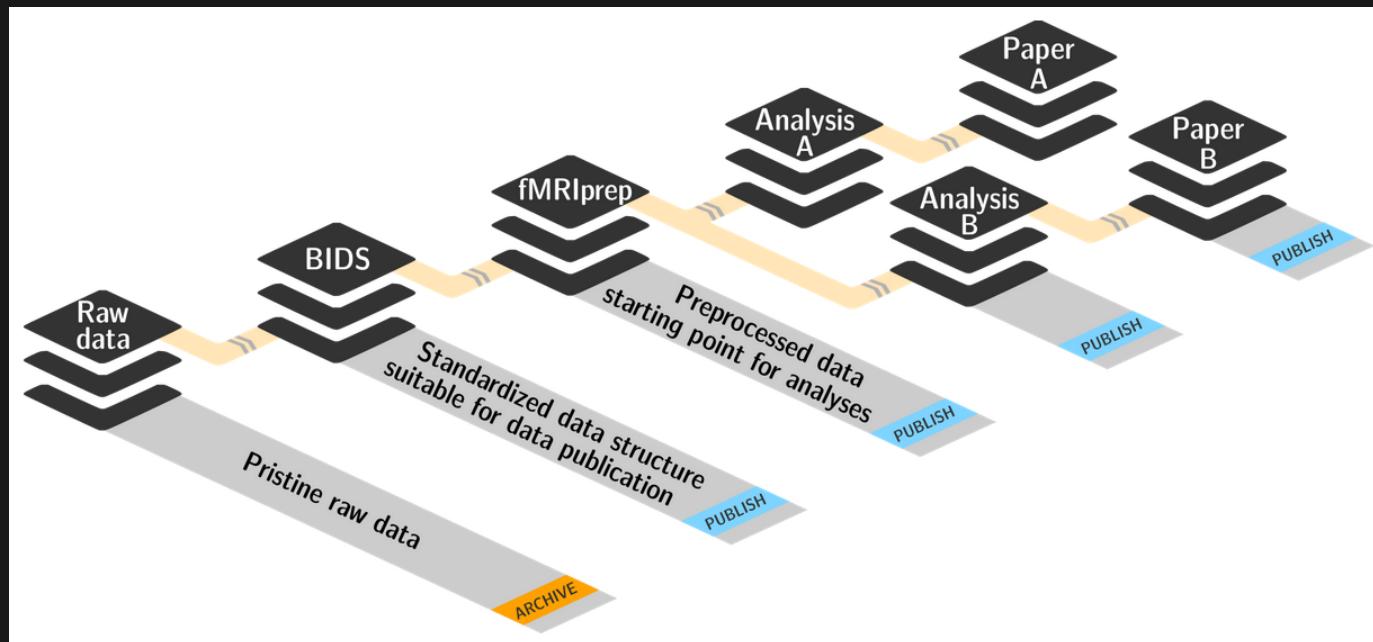
```
datalad run-containers
```

It stores the software container in the dataset, links datasets (as subdatasets) and the software container, records data origin and command execution, and collect and store provenance of all contents of a dataset created.

→ Fully computationally reproducible analysis

(COMPUTATIONALLY) REPRODUCIBLE DATASET ANALYSIS

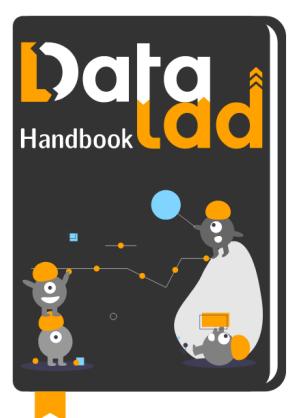
- DataLad can comprehensively represent a typical neuroimaging study workflow from raw data to paper results



<http://www.repronim.org/ohbm2018-training/03-01-reproin/>

HOW TO GET STARTED WITH DATALAD?

FOLLOW THE GREAT DATALAD HANDBOOK



The screenshot shows the GitHub page for the DataLad Handbook. At the top left is the repository icon, which is a dark blue square with a white 'D' and 'L' logo. To its right is the repository name 'DataLad Handbook'. Below the icon is a small image of three cartoonish characters (two grey and one white) interacting with a network of nodes and lines. A yellow ribbon bookmark is positioned below the image. To the right of the bookmark is a 'Star' button with the number '59' next to it. Below the star button is a short description: 'The DataLad handbook will supply you with everything you need to get started and break new grounds with DataLad.' At the bottom left is a graphic featuring a yellow circle with a white 'D', two dark blue mountain peaks, and a yellow banner that says 'Get the gist!'.



The screenshot shows the main landing page of the DataLad Handbook. It features a large, stylized 'DataLad' logo where the 'D' and 'L' are in orange and the 'a' and 't' are in black. Below the logo is the text 'The Handbook'. Underneath that is a 'Welcome!' heading. The page then describes the handbook as a living resource for learning DataLad, mentioning 'Basics', 'use cases', and a 'companion repository'. It also notes that the handbook is based on at least version 0.12 of DataLad.

Welcome!

This handbook is a living resource about why and – more importantly – *how* to use DataLad. It aims to provide novices and advanced users of all backgrounds with both the basics of DataLad and start-to-end use cases of specific applications. If you want to get hands-on experience and learn DataLad, the *Basics* part of this book will teach you. If you want to know what is possible, the *use cases* will show you. And if you want to help others to get started with DataLad, the [companion repository](#) provides [free and open source teaching material](#) tailored to the handbook.

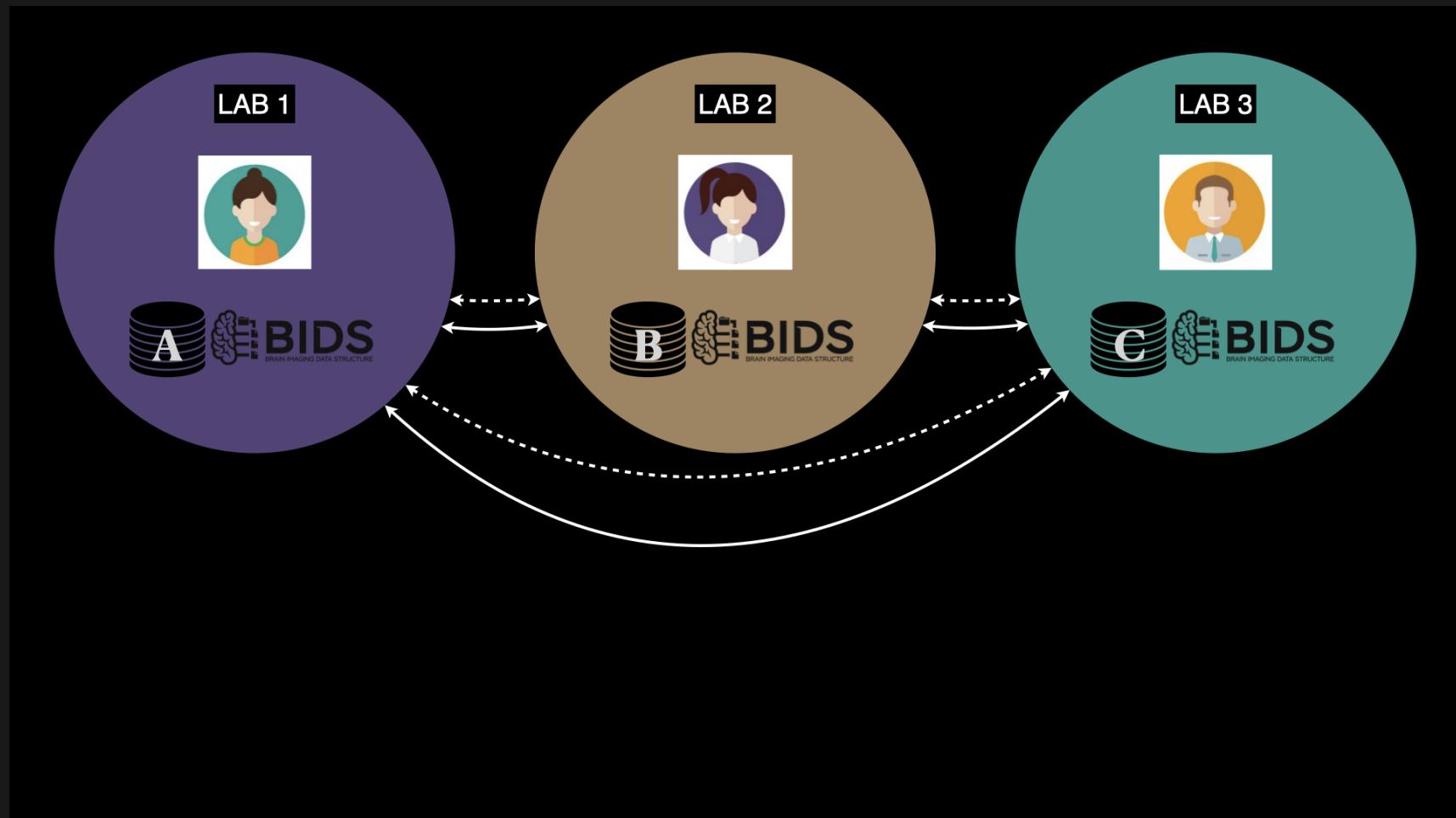
Before you read on, please note that the handbook is based on **at least** DataLad version **0.12**, and the higher your version of DataLad is, the better. The section [Installation and configuration](#) will set you up with what you need if you currently do not have DataLad 0.12 or higher installed.

WHAT ARE THE STEPS?

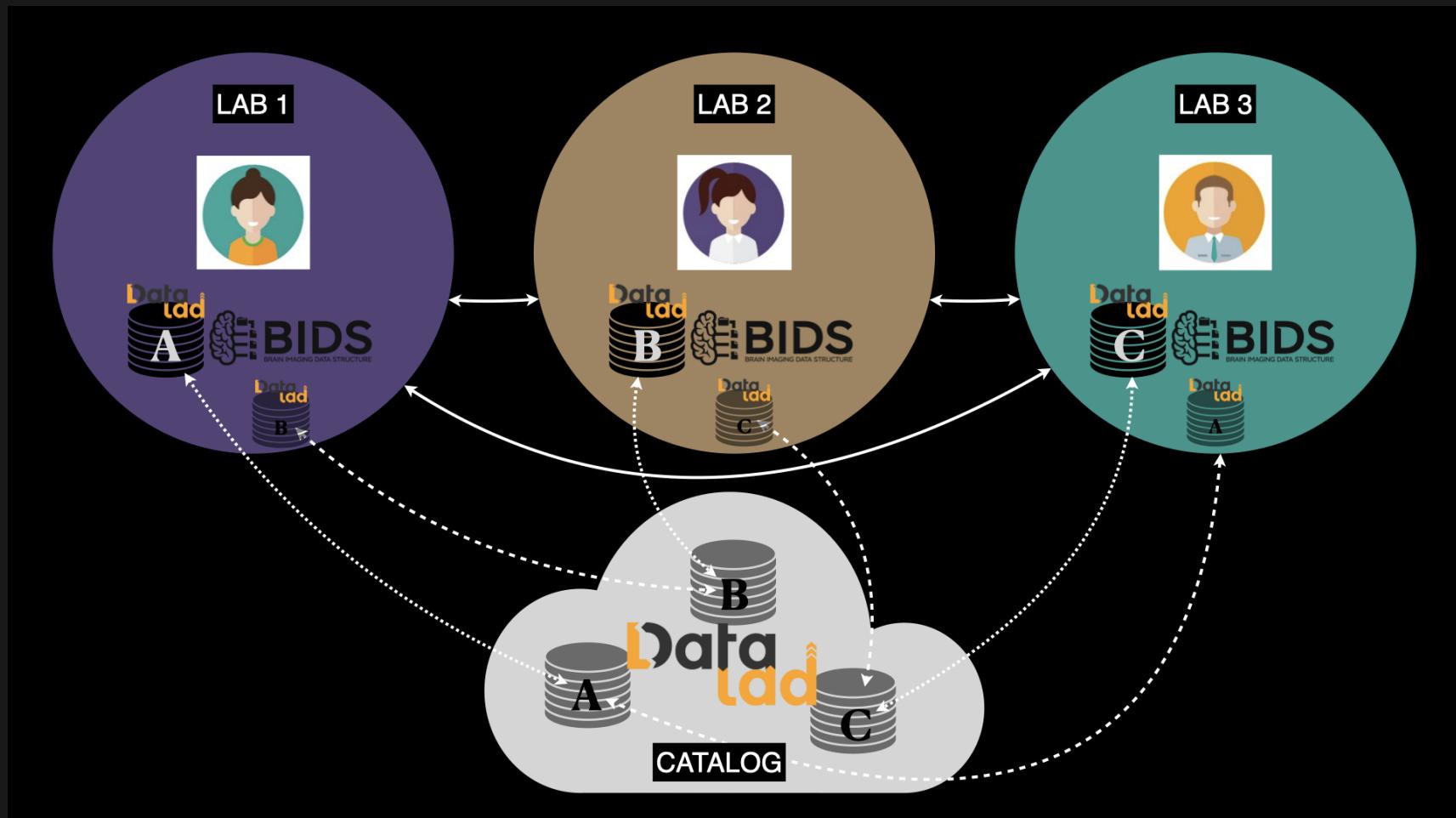
STEP 1: UNDERSTANDING HOW DATASETS ARE ORGANIZED IN EACH LAB



STEP 2: CREATION OF BIDS DATASETS FOR RAW DATA



STEP 3: ADOPTION OF DATALAD AND CREATION OF DATASET CATALOG



THANK YOU FOR YOUR ATTENTION!