THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Open Data with Novel Techniques

Leveraging the NeuroDesk platform to Enhance Reproducible Workflows with Open Data

Steffen Bollmann
Senior Research Fellow
School of Information Technology and Electrical Engineering, The University of Queensland, Australia

Acknowledgement of **Country**

The University of Queensland (UQ) acknowledges the Traditional Owners and their custodianship of the lands on which we meet.

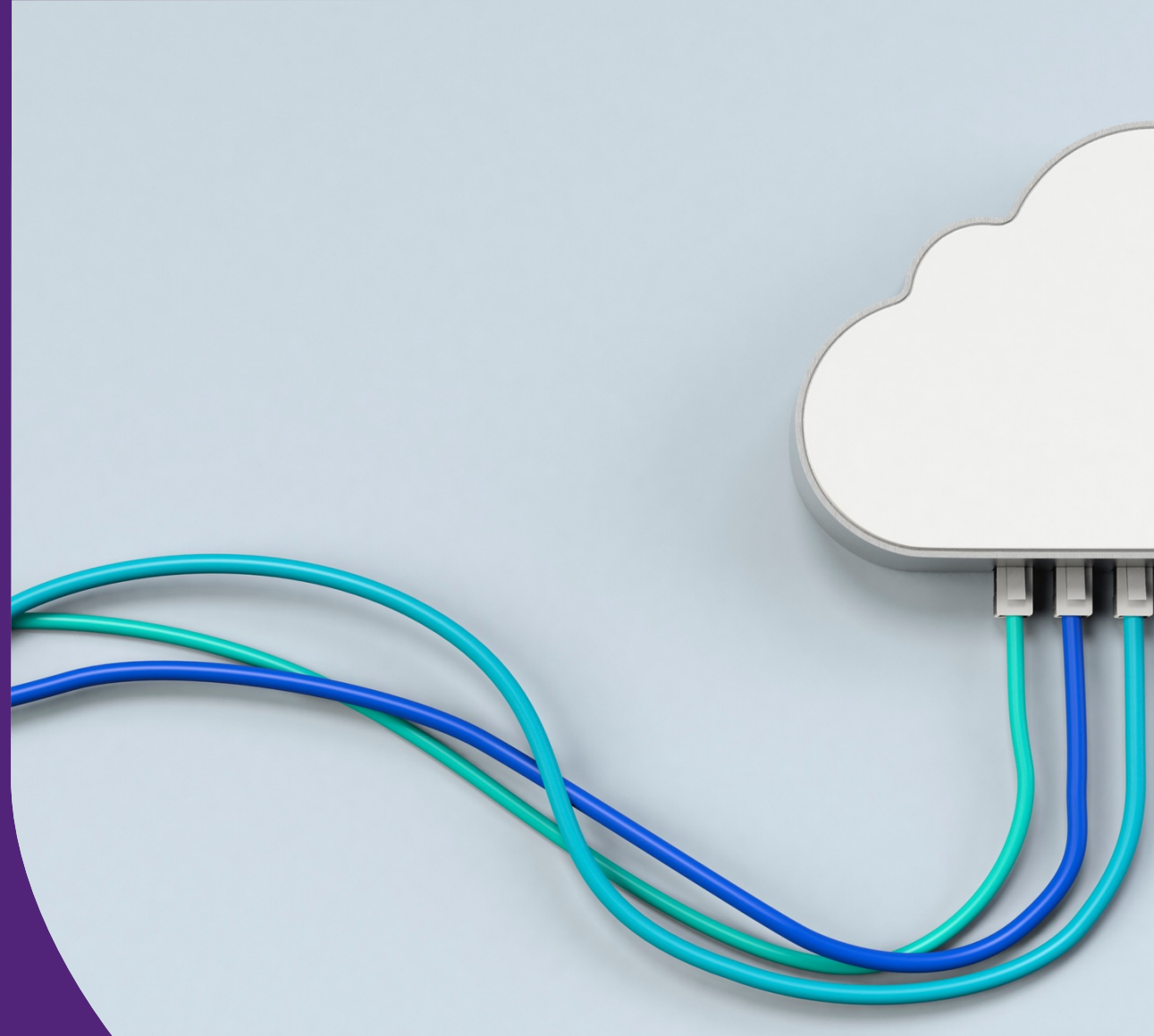
We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.

We recognise their valuable contributions to Australian and global society.

Declaration of Potential Conflicts of Interest

I receive research funding from:

1. Oracle for Research
   a. partially fund NeuroDesk project via cloud credits
   b. I will talk about a project from Oracle for Research, called “Oracle Open Data”
2. Siemens Healthineers



Talk Outline

- Introducing an Open Data example
- NeuroDesk & The Open Data workflow
- Publishing Open Data
- Accessing Open Data
- Re-executable papers linking Code and Open Data

Open Data is easy, right?

What if this free service stops?

Download a multiple GB file just to see which functions were used?

DATA AVAILABILITY STATEMENT

All data and code used in this paper can be found here:
https://www.dropbox.com/s/ijkxmopv088e4iu/matlab_code_data_mri_paper2022.zip?dl=0

Will the code still work in the next version? Does the reader have a matlab license?

I found a bug and need to update this ….

Interactively running in browser – no setup needed

Platform can be changed later

DATA AVAILABILITY STATEMENT

We facilitate the reproducibility of our study by providing an interactive version of our implementation on a publicly accessible cloud-based platform. The readers can explore the implementation of the model (neural network), train the model with different hyper-parameters and architectures, investigate the stability of the training process, and reproduce our results with the identical model used in this manuscript (https://github.com/sbollmannMRI/scout2B1, 320a6ab). We anonymized and stored the input data (localizer, SA2RAGE B_1^+) of 28 participants in OSF (OSF, Center for Open Science, Inc., Virginia, USA) accessible via https://osf.io/y5cq9/.

Provide source code in an easy accessible way

This commit was used for the paper, but bug fixes possible

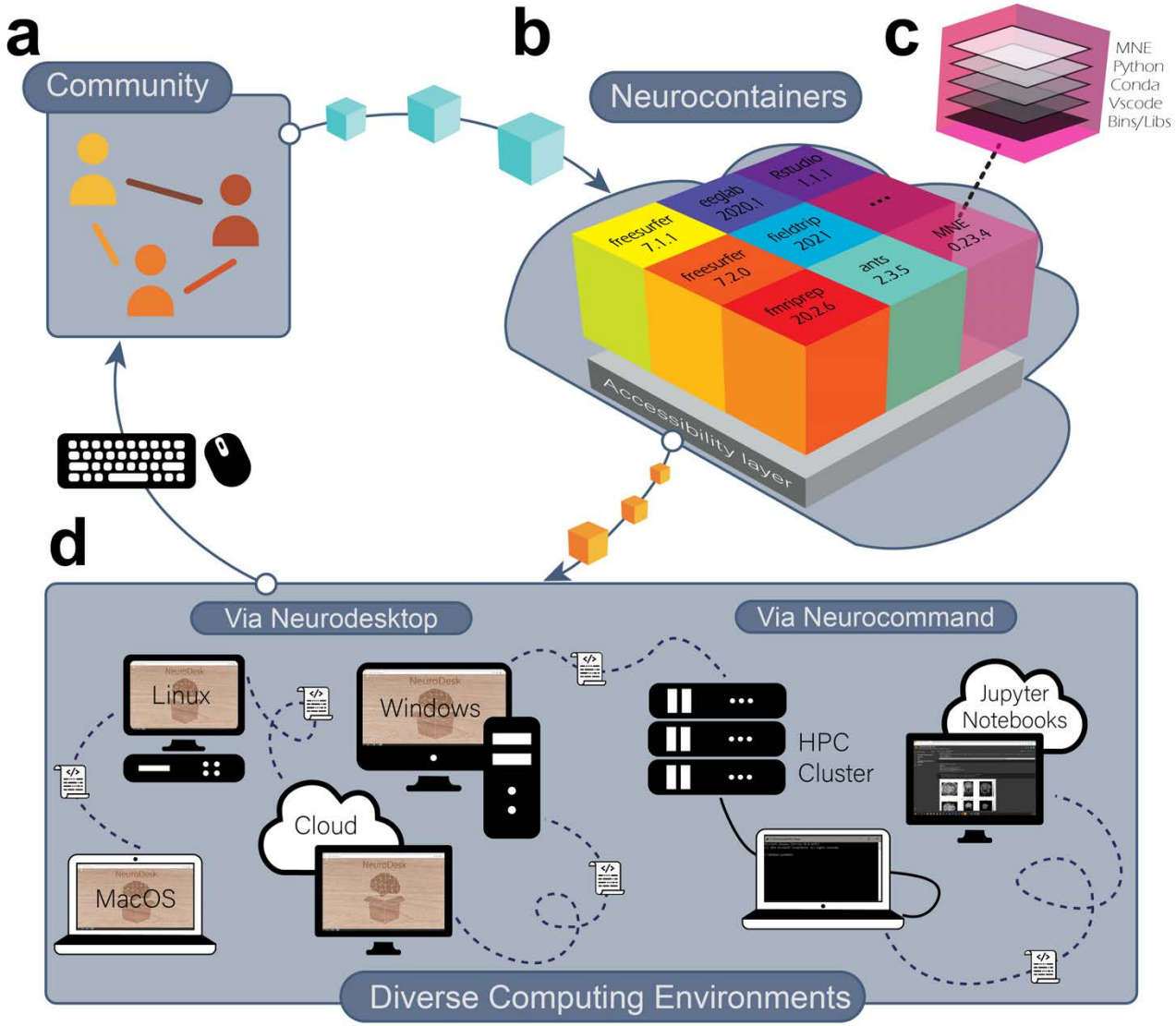
Data and links can be updated if bugs found or services move

NeuroDesk & The Open Data workflow

What is NeuroDesk?

Which problems does it solve?

The NeuroDesk.org Open-Source Platform



Challenges we tackle:

1. Research software is difficult to install (e.g. dependency conflicts, lack of packages/maintenance)
2. Variable compute environments and operating systems (e.g. HPCs, workstations, laptops, cloud …)
3. Large Datasets

Partners and Funders:

Australian Research Data Commons

National Imaging Facility

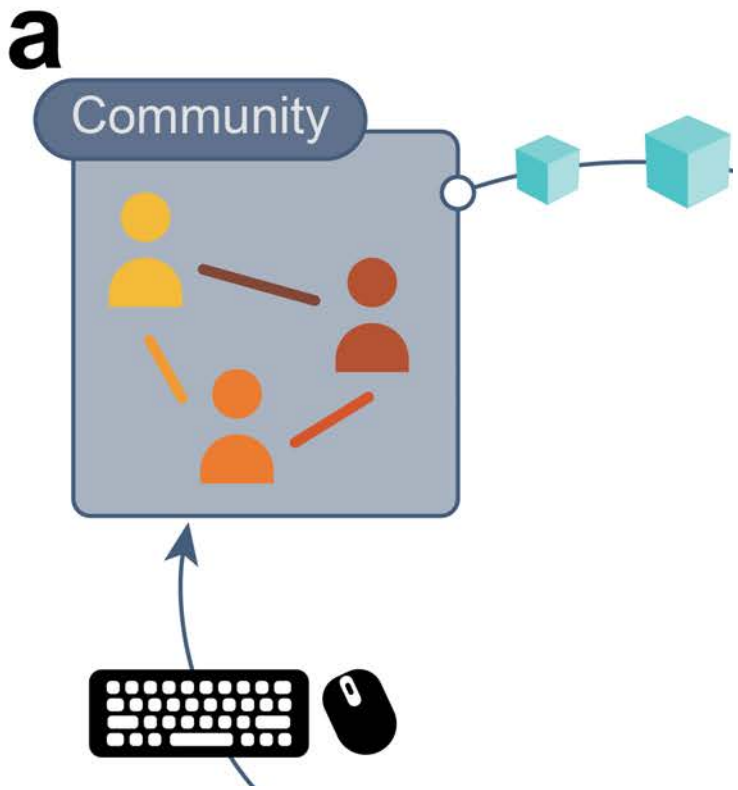
ORACLE for Research

Swinburne University of Technology

The University Of Queensland AUSTRALIA CREATE CHANGE

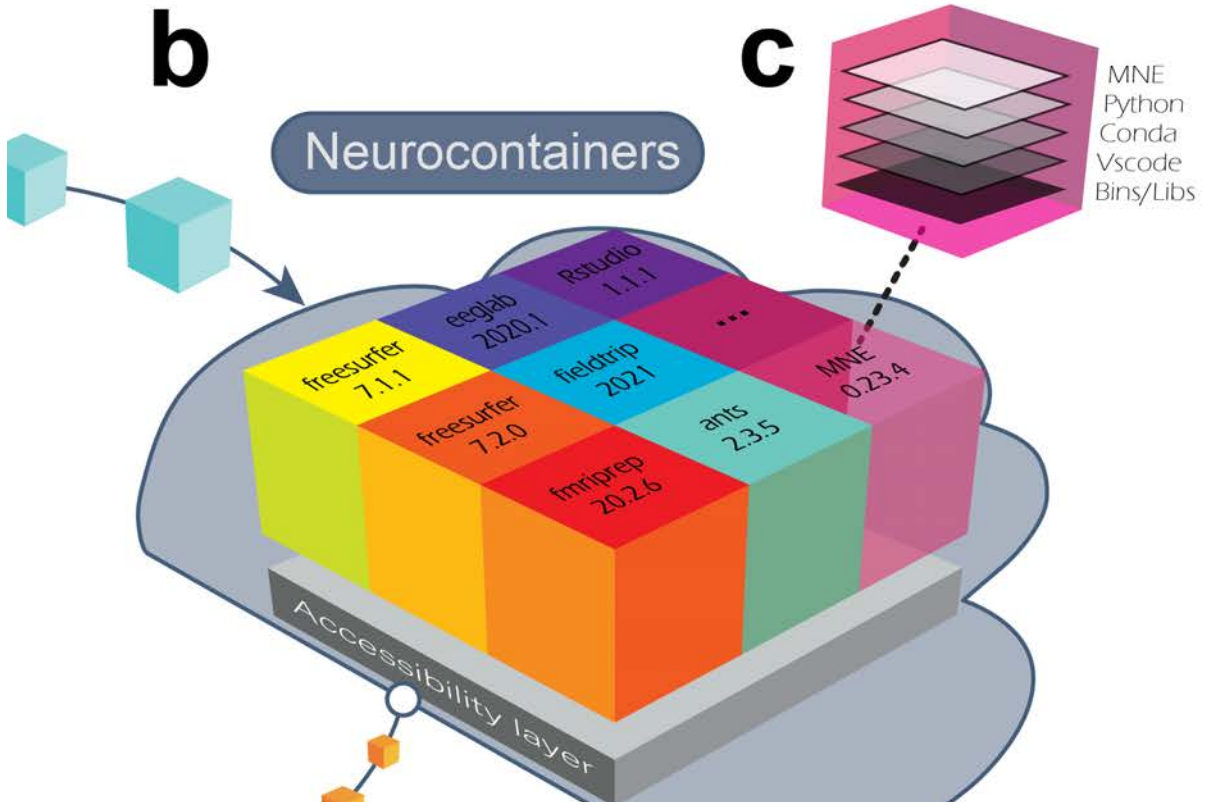
The NeuroDesk.org Open-Source Platform

Community builds and maintains software containers



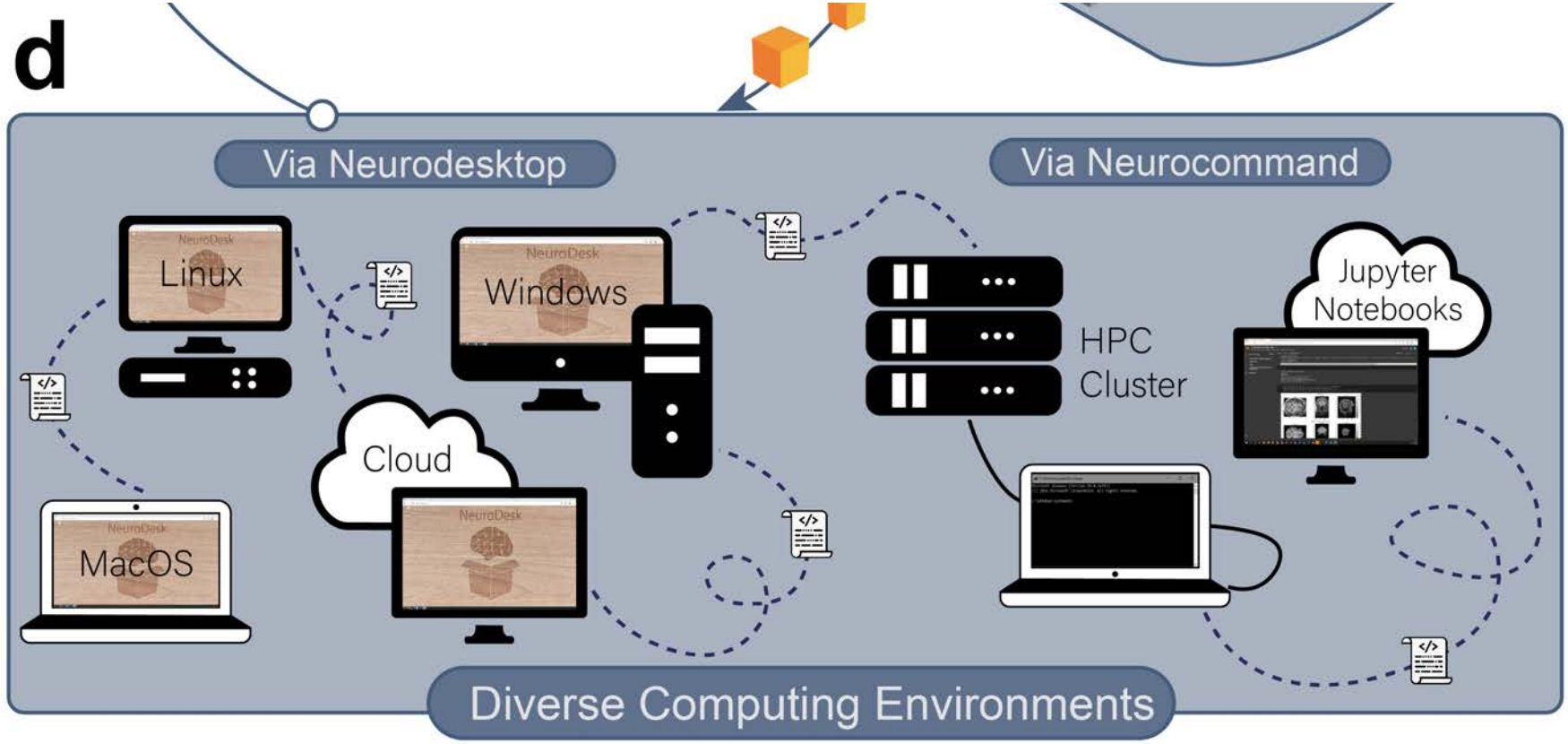
The NeuroDesk.org Open-Source Platform

Neurocontainers automatically builds and distributes a repository of software containers

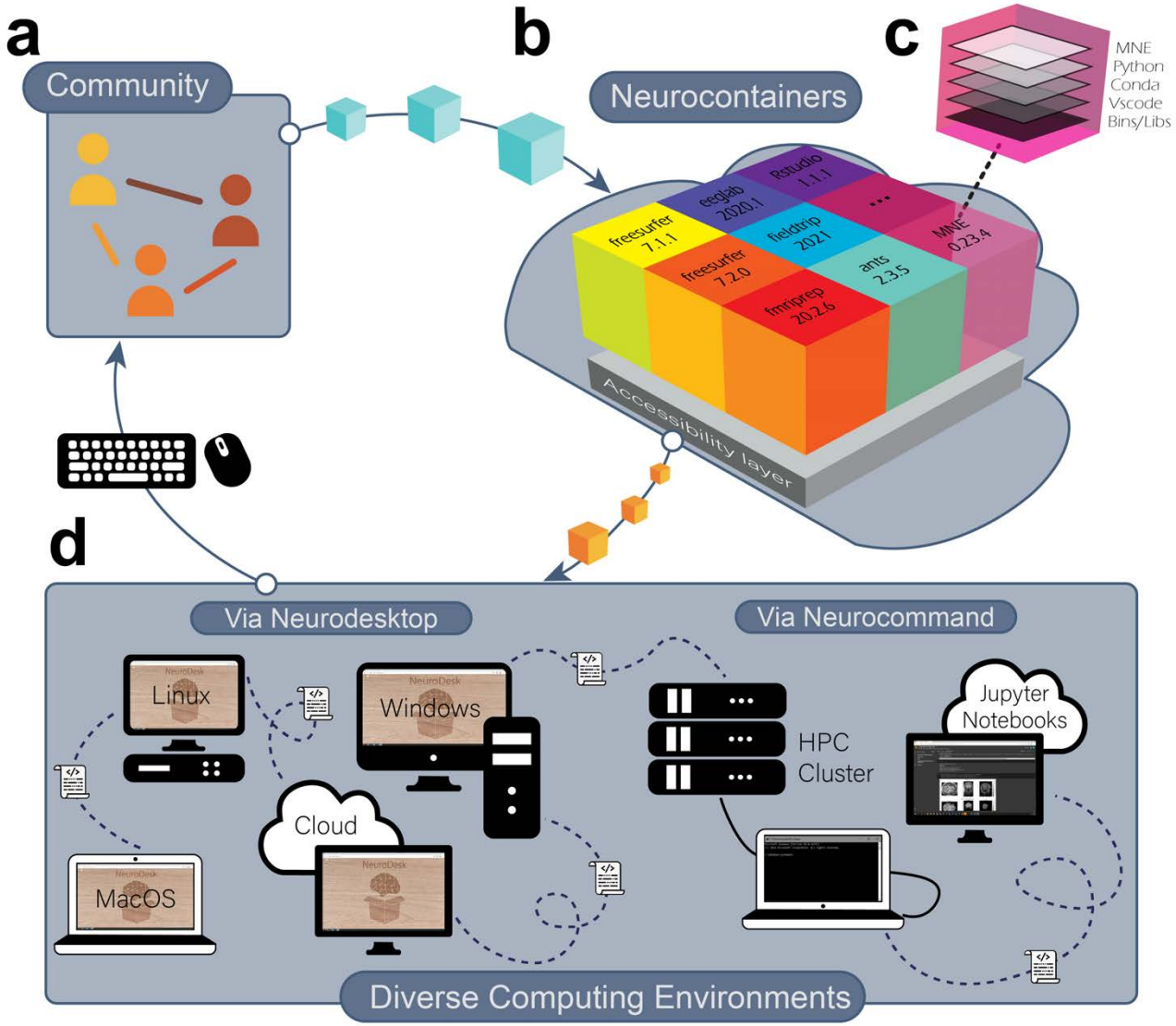


The NeuroDesk.org Open-Source Platform

Software containers are available for all compute environments:



The NeuroDesk.org Open-Source Platform



Challenges we tackle:

1. Research software is difficult to install (e.g. dependency conflicts, lack of packages/maintenance)
2. Variable compute environments and operating systems (e.g. HPCs, workstations, laptops, cloud …)
3. Large Datasets

Partners and Funders:

Australian Research Data Commons

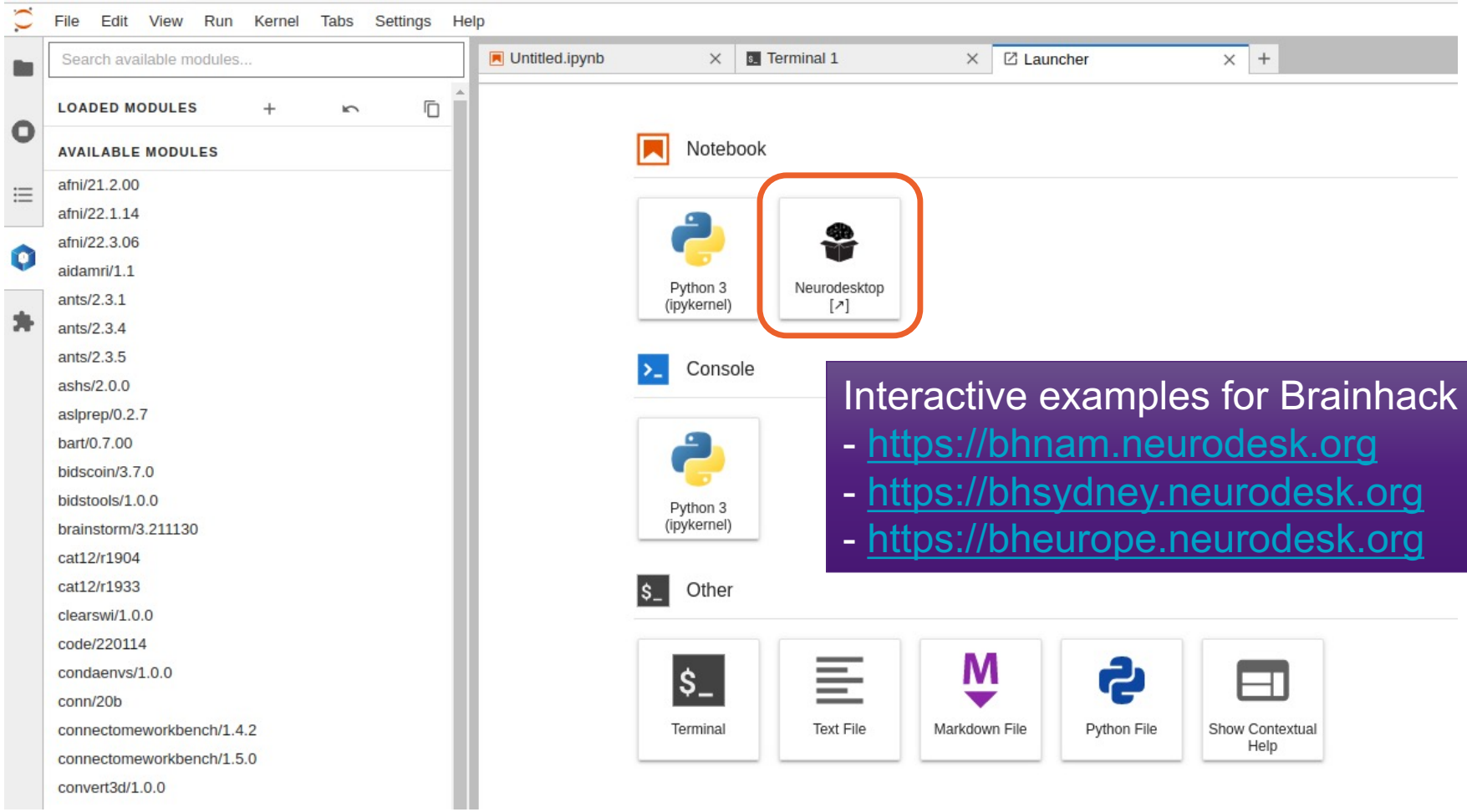
National Imaging Facility

ORACLE for Research

SWINBURNE UNIVERSITY OF TECHNOLOGY

THE UNIVERSITY OF QUEENSLAND AUSTRALIA CREATE CHANGE

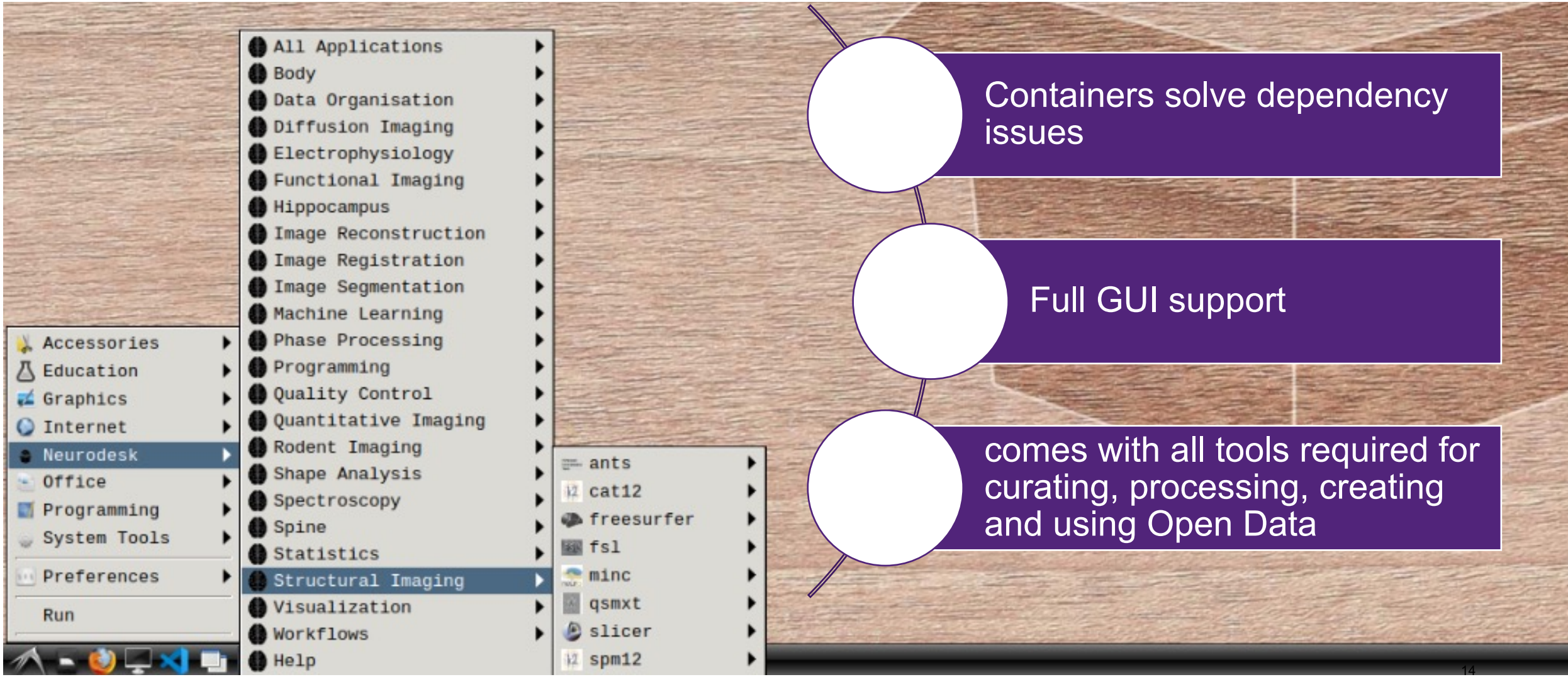
NeuroDesk can be integrated in Jupyter Lab



Interactive examples for Brainhack 2022:
- https://bhnam.neurodesk.org
- https://bhsydney.neurodesk.org
- https://bheurope.neurodesk.org

Supported by:
ORACLE for Research

NeuroDesktop – A Linux desktop accessible via the browser



Publishing Open Data

Where to store the data?

Which tools exist?

Making our Data openly available

OSF

?

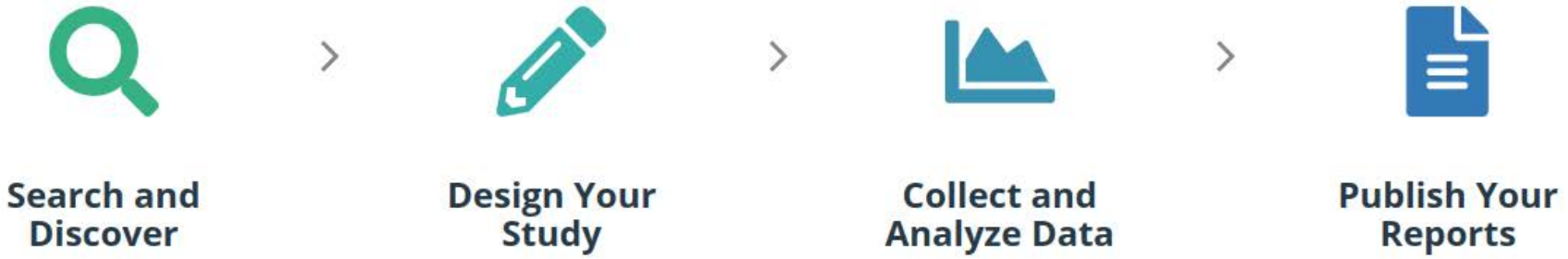
DATA AVAILABILITY STATEMENT

We facilitate the reproducibility of our study by providing an interactive version of our implementation on a publicly accessible cloud-based platform. The readers can explore the implementation of the model (neural network), train the model with different hyper-parameters and architectures, investigate the stability of the training process, and reproduce our results with the identical model used in this manuscript (https://github.com/sbollmannMRI/scout2B1, 320a6ab). We anonymized and stored the input data (localizer, SA2RAGE B_1^+) of 28 participants in OSF (OSF, Center for Open Science, Inc., Virginia, USA) accessible via https://osf.io/y5cq9/.

Data and links can be updated if bugs found or services move

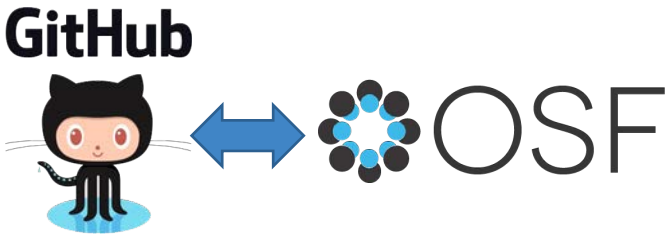
What is the Open Science Framework (OSF)?

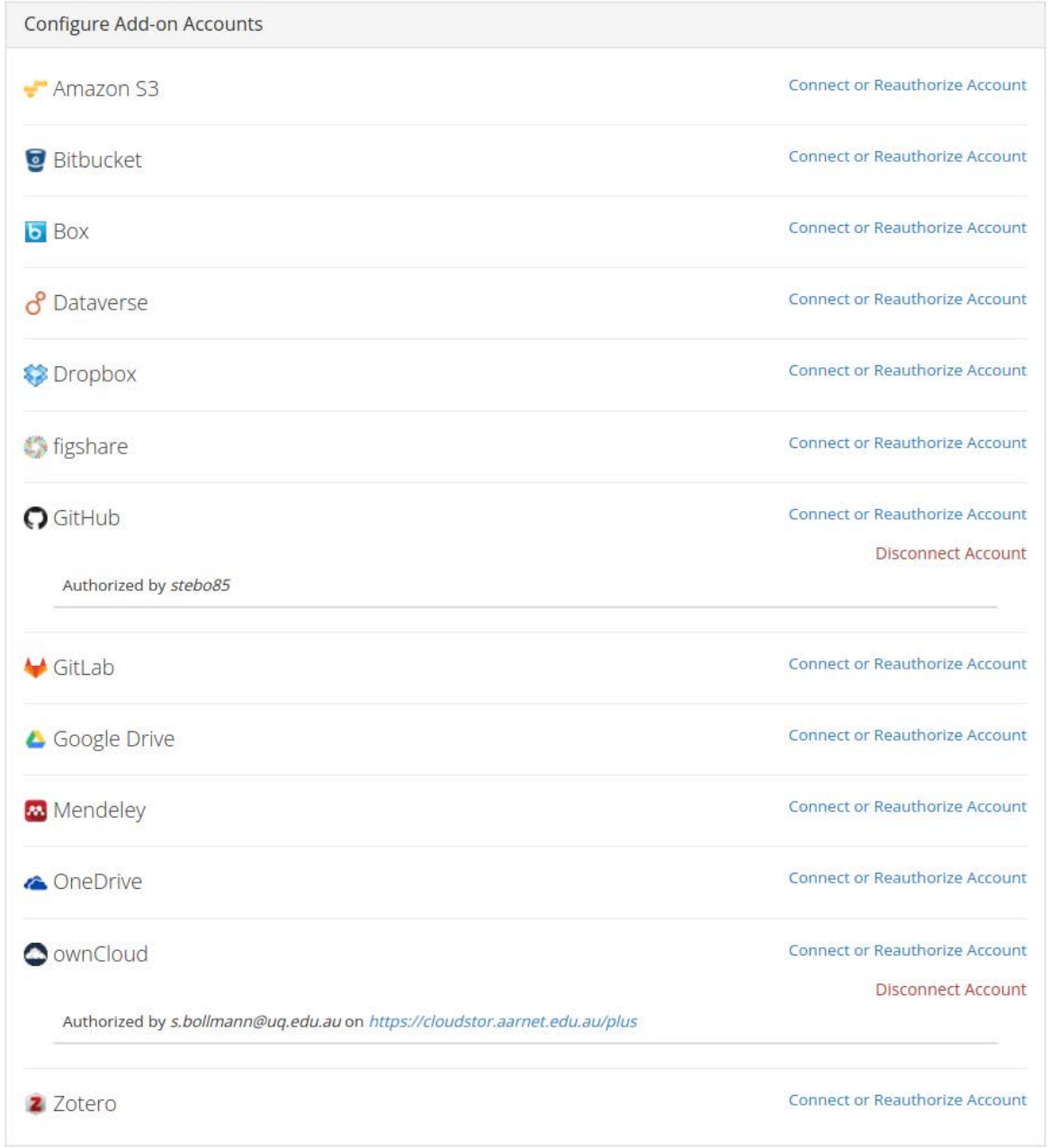
- Online platform that enables researchers to plan, collect, analyze and share their work
- Developed and maintained by the non-profit organization Center for Open Science (COS)
- Preservation fund to provide 50+ years read access to the hosted data



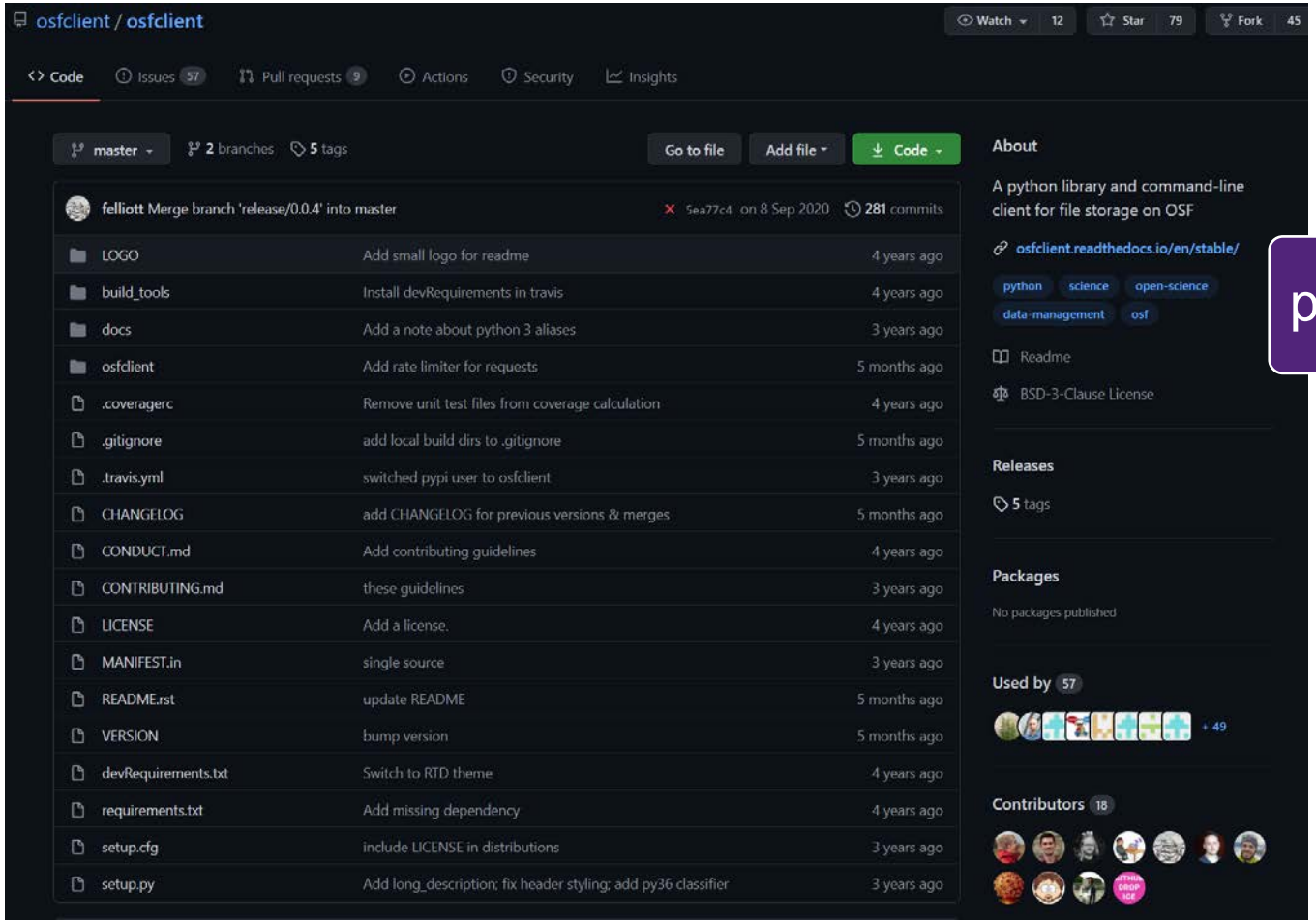
https://help.osf.io/article/546-getting-started-faq-s

Connecting Services





Setting up OSF command line client & uploading data



pip install osfclient

osf init

osf upload –r . osfstorage/data

https://osfclient.readthedocs.io/en/latest/

Scout2B1

116.8MB | Public | 0 | ...

Contributors: Steffen Bollmann, Shahrokh Abbasi Rad

Date created: 2020-09-11 02:13 PM | Last Updated: 2021-04-28 03:19 PM

Identifier: DOI 10.17605/OSF.IO/Y5CQ9

Category: Project

Description: *This data and code are available to reproduce the results of the paper Abbasi-Rad, Shahrokh, Kieran O'Brien, Samuel Kelly, Viktor Vegh, Anders Rodell, Yasvir Tesiram, Jin Jin, Markus Barth, and Steffen Bollmann. 'Improving FLAIR SAR Efficiency at 7T by Adaptive Tailoring of Adiabatic Pulse Power through Deep Learning Estimation'. Magnetic Resonance in Medicine n/a, no. n/a (2020). https://doi.org/10.1002/mrm.28590.*

(preprint: Abbasi-Rad, S., O'Brien, K., Kelly, S., Vegh, V., Rodell, A., Tesiram, Y., Jin, J., Barth, M., Bollmann, S., 2019. Improving FLAIR SAR efficiency at 7T by adaptive tailoring of adiabatic pulse power using deep convolutional neural networks. arXiv:1911.08118 [physics].)

Wiki

>>> The interactive computational notebook can be found here <<<

Files

Filter

Name	Modified
Scout2B1	
OSF Storage (Australia - Sydney)	
.osfcli.config	2020-09-12 01:55 PM
B1Map	
checkpoints	

Citation

Recent Activity

Steffen Bollmann deauthorized the GitHub addon for Scout2B1
2021-04-28 03:19 PM

Scout2B1 registered
2021-01-12 03:02 PM

Registration of Scout2B1 was approved
2021-01-12 03:02 PM

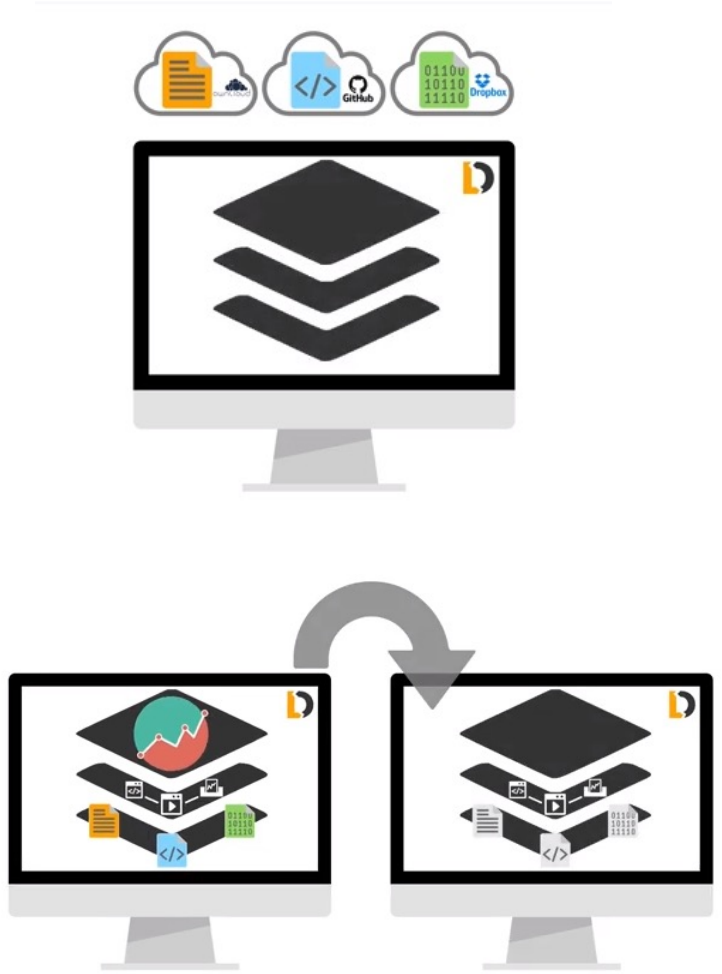
Steffen Bollmann initiated a registration of Scout2B1
2021-01-10 10:21 AM

Steffen Bollmann created external identifier(s) doi:10.17605/OSF.IO/Y5CQ9 on Scout2B1
2021-01-10 09:30 AM

Steffen Bollmann edited description of Scout2B1
2020-12-15 02:23 PM

What is DataLad?

- free and open source distributed data management system
  - Keeps track of data
  - Creates structure
  - Ensures reproducibility
  - Supports collaboration
  - Integrates with widely used data infrastructure -> including the OSF :)



https://www.datalad.org/

Uploading data using the DataLad OSF extension

Step 1: creating a DataLad dataset

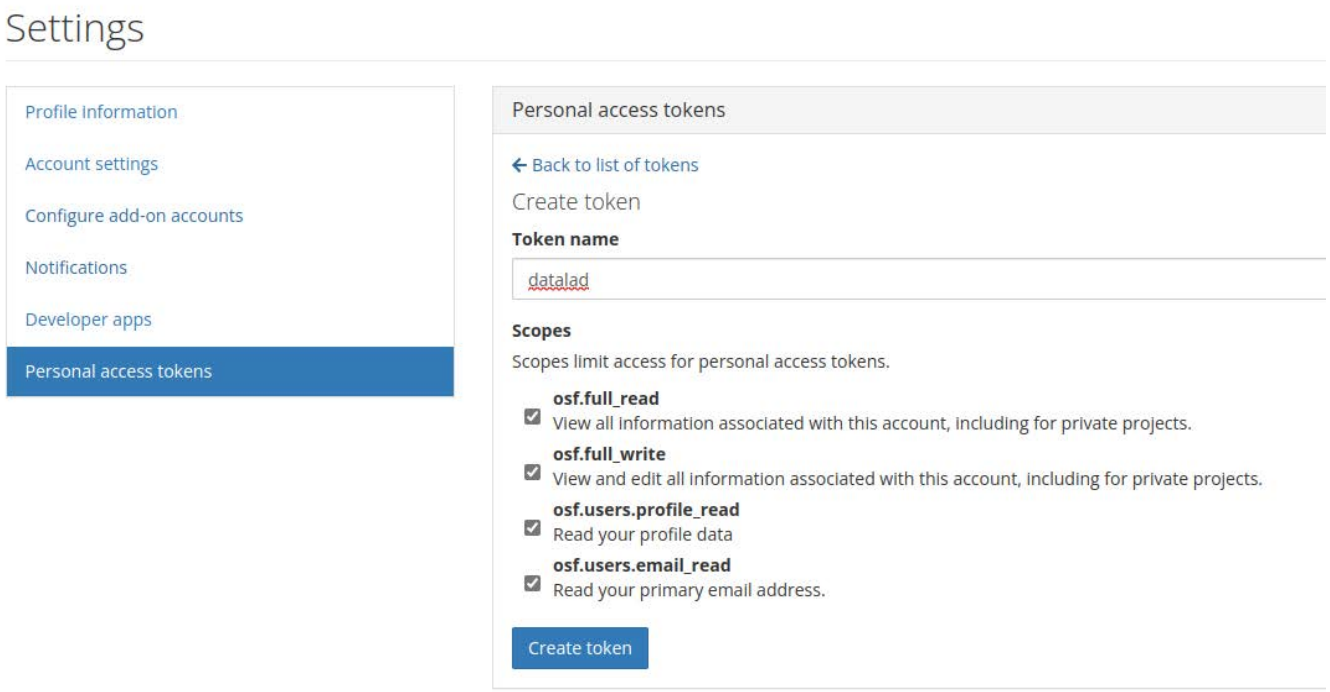
```
jovyan@neurodesktop:~$ datalad create my_dataset

jovyan@neurodesktop:~/my_dataset$ datalad save -m "added"
add(ok): B1Map/rB1MapinScout_1.nii.gz (file)
add(ok): B1Map/rB1MapinScout_10.nii.gz (file)
add(ok): B1Map/rB1MapinScout_11.nii.gz (file)
add(ok): B1Map/rB1MapinScout_12.nii.gz (file)
add(ok): B1Map/rB1MapinScout_13.nii.gz (file)
add(ok): B1Map/rB1MapinScout_14.nii.gz (file)
add(ok): B1Map/rB1MapinScout_15.nii.gz (file)
add(ok): B1Map/rB1MapinScout_16.nii.gz (file)
add(ok): B1Map/rB1MapinScout_17.nii.gz (file)
add(ok): B1Map/rB1MapinScout_18.nii.gz (file)
```

http://docs.datalad.org/projects/osf/en/latest/

Uploading data using the DataLad OSF extension

Step 2: Creating an OSF token & authenticating




```
jovyan@neurodesktop:~$ datalad osf-credentials
```

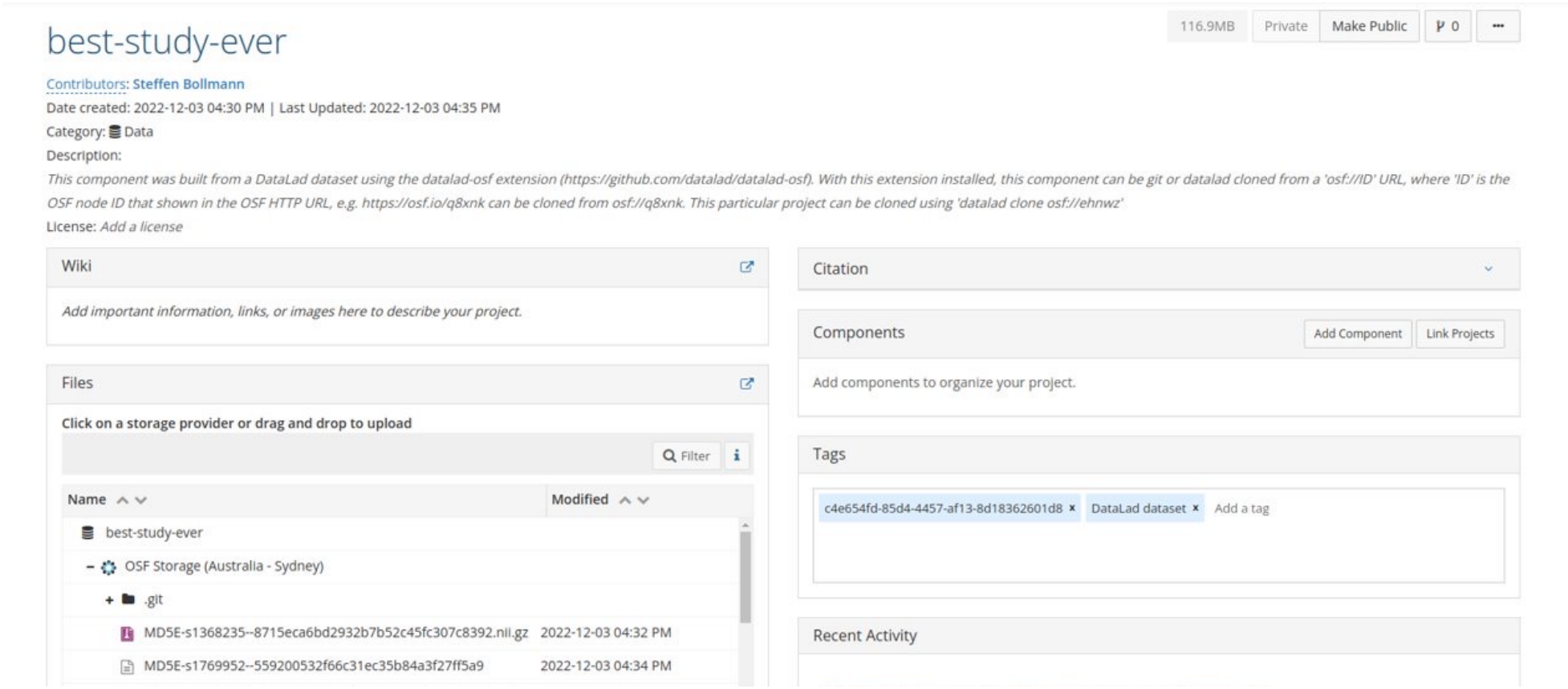
http://docs.datalad.org/projects/osf/en/latest/

Uploading data using the DataLad OSF extension

Step 3: Creating and pushing a DataLad sibling

```
jovyan@neurodesktop:~/my_dataset$ datalad create-sibling-osf --title best-study-ever -s osf

jovyan@neurodesktop:~/my_dataset$ datalad push --to osf
```



Accessing Open Data

How do we efficiently access the data?

Do we have to download everything at once?

Data access within Jupyter notebook.

DATA AVAILABILITY STATEMENT

We facilitate the reproducibility of our study by providing an interactive version of our implementation on a publicly accessible cloud-based platform. The readers can explore the implementation of the model (neural network), train the model with different hyper-parameters and architectures, investigate the stability of the training process, and reproduce our results with the identical model used in this manuscript (https://github.com/sbollmannMRI/scout2B1, 320a6ab). We anonymized and stored the input data (localizer, SA2RAGE B_1^+) of 28 participants in OSF (OSF, Center for Open Science, Inc., Virginia, USA) accessible via https://osf.io/y5cq9/.

Scout2B1

116.8MB | Public | 0 | ...

Contributors: Steffen Bollmann, Shahrokh Abbasi Rad

Date created: 2020-09-11 02:13 PM | Last Updated: 2021-04-28 03:19 PM

Identifier: DOI 10.17605/OSF.IO/Y5CQ9

Category: Project

Description: *This data and code are available to reproduce the results of the paper Abbasi-Rad, Shahrokh, Kieran O'Brien, Samuel Kelly, Viktor Vegh, Anders Rodell, Yasvir Tesiram, Jin Jin, Markus Barth, and Steffen Bollmann. 'Improving FLAIR SAR Efficiency at 7T by Adaptive Tailoring of Adiabatic Pulse Power through Deep Learning Estimation'. Magnetic Resonance in Medicine n/a, no. n/a (2020). https://doi.org/10.1002/mrm.28590.*

(preprint: Abbasi-Rad, S., O'Brien, K., Kelly, S., Vegh, V., Rodell, A., Tesiram, Y., Jin, J., Barth, M., Bollmann, S., 2019. Improving FLAIR SAR efficiency at 7T by adaptive tailoring of adiabatic pulse power using deep convolutional neural networks. arXiv:1911.08118 [physics].)

Wiki

>>> The interactive computational notebook can be found here <<<

Files

Filter

Name	Modified
Scout2B1	
OSF Storage (Australia - Sydney)	
.osfcli.config	2020-09-12 01:55 PM
B1Map	
checkpoints	

Citation

Recent Activity

Steffen Bollmann deauthorized the GitHub addon for Scout2B1
2021-04-28 03:19 PM

Scout2B1 registered
2021-01-12 03:02 PM

Registration of Scout2B1 was approved
2021-01-12 03:02 PM

Steffen Bollmann initiated a registration of Scout2B1
2021-01-10 10:21 AM

Steffen Bollmann created external identifier(s) doi:10.17605/OSF.IO/Y5CQ9 on Scout2B1
2021-01-10 09:30 AM

Steffen Bollmann edited description of Scout2B1
2020-12-15 02:23 PM

Scout2B1.ipynb ☆

File Edit View Insert Runtime Tools Help Changes will not be saved

Share S

Table of contents

- Setup for running on Google Colab
- Setup for Running on your own GPU
- Check if GPU is enabled
- Connect google drive
- Download Data
- Import Packages
- Function Definitions
  - Utility Functions
    - Function: convert_to
    - Function: cut_one_example
    - Function: generate_file_list
    - Function: data_input_fn
  - Model Architecture
    - Function: conv_arci
  - Main Functions
    - Function: make_data
    - Function: train_model
    - Function: predict_model
- Step 01: Preparing Data for Training
- Step 02: Training the model

+ Code + Text Copy to Drive Connect Editing

Download Data

The data for training our model is stored in OSF (https://osf.io/y5cq9/).
The data will be downloaded from OSF to the Disk allocated by the colab on its virtual machine in the path below:
/content/data/osfstorage

Scout: Containing 28 localizer scans
B1Map: Containing 28 SA2RAGE B1 map scans of the same group of patients which are already resliced into localizer space and masked.
UnseenData: Containing one pair of scout and B1 map which is not used during the training process. This data will be used for prediction.

The purpose is to use the 28 (scout, B1map) pairs to train a UNet CNN structure so that the network learns how to predict a B1 map from a scout scan.
In the next step, we will use the trained network on the unseen data to see if the CNN works on data not used for training.

```
[ ]  # Install the client code of Open Sience Framework (OSF)
     !pip install osfclient
```

```
Requirement already satisfied: osfclient in /usr/local/lib/python3.7/dist-packages (0.0.5)
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from osfclient) (1.15.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from osfclient) (4.41.1)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from osfclient) (2.23.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->osfclient) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->os
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->osfclient) (2020.12.5)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->osfclient) (2.10)
```

Using OSF data in a Jupyter Notebook

Installing the osfclient is straight forward:

```
[1]: !pip install osfclient

Requirement already satisfied: osfclient in /opt/conda/lib/python3.10/site-packages (0.0.5)
Requirement already satisfied: tqdm in /opt/conda/lib/python3.10/site-packages (from osfclient) (4.64.1)
Requirement already satisfied: requests in /opt/conda/lib/python3.10/site-packages (from osfclient) (2.28.1)
Requirement already satisfied: six in /opt/conda/lib/python3.10/site-packages (from osfclient) (1.16.0)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-packages (from requests->osfclient) (2022.9.24)
Requirement already satisfied: charset-normalizer<3,>=2 in /opt/conda/lib/python3.10/site-packages (from requests->osfclient) (2.1.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /opt/conda/lib/python3.10/site-packages (from requests->osfclient) (1.26.13)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages (from requests->osfclient) (3.4)
```

https://mri.sbollmann.net/index.php/2020/05/27/google-colab-osf/

Using OSF data in a Jupyter Notebook

OSF project ID

```
[*]: !osf -p y5cq9 clone .
```

```
0files [00:00, ?files/s]
100%|| 663k/663k [00:00<00:00, 15.3Mbytes/s]
1files [00:08,  8.25s/files]
100%|| 610k/610k [00:00<00:00, 24.7Mbytes/s]
2files [00:10,  4.62s/files]
100%|| 698k/698k [00:00<00:00, 57.6Mbytes/s]
3files [00:12,  3.38s/files]
100%|| 615k/615k [00:00<00:00, 86.8Mbytes/s]
4files [00:14,  2.79s/files]
```

```
[3]: !ls osfstorage
```

```
B1Map  Scout  UnseenData  checkpoints
```

https://mri.sbollmann.net/index.php/2020/05/27/google-colab-osf/

Using osf data with DataLad

```
jovyan@neurodesktop:~$ datalad clone osf://ehnwz
```

```
jovyan@neurodesktop:~/ehnwz$ ll
total 36
drwxr-sr-x.  8 jovyan users 4096 Dec  3 06:39 .
drwxrwsr-x. 28 root   users 4096 Dec  3 22:11 ..
drwxr-sr-x.  2 jovyan users 4096 Dec  3 06:39 .datalad
drwxr-sr-x. 10 jovyan users 4096 Dec  3 21:31 .git
-rw-r--r--.  1 jovyan users   55 Dec  3 06:39 .gitattributes
drwxr-sr-x.  2 jovyan users 4096 Dec  3 06:39 B1Map
drwxr-sr-x.  2 jovyan users 4096 Dec  3 06:39 Scout
drwxr-sr-x.  2 jovyan users 4096 Dec  3 06:39 UnseenData
drwxr-sr-x.  3 jovyan users 4096 Dec  3 06:39 checkpoints
```

```
jovyan@neurodesktop:~/ehnwz$ datalad get B1Map/
```

Great feature of DataLad: It doesn't download everything at once, but we can pick which files we want

DataLad provides access to a variety of open data sources

OpenNeuro

CBRAIN

brainlife.io

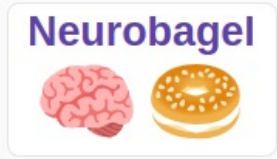
CONP

DANDI

Courtois Neuromod

OSF

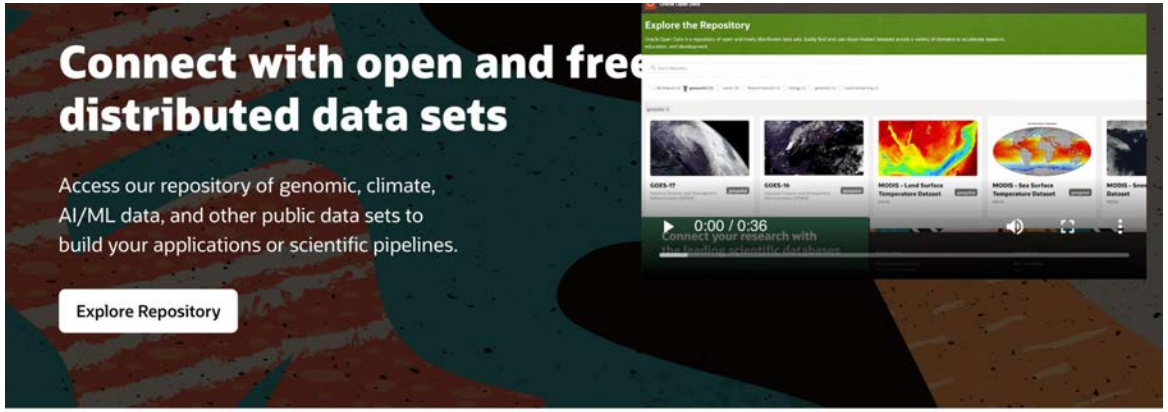
Dataverse



Neurobagel

https://www.datalad.org/in-the-wild.html

Oracle Open Data – Free Open Data hosting



Why Oracle Open Data?

Findable and accessible data

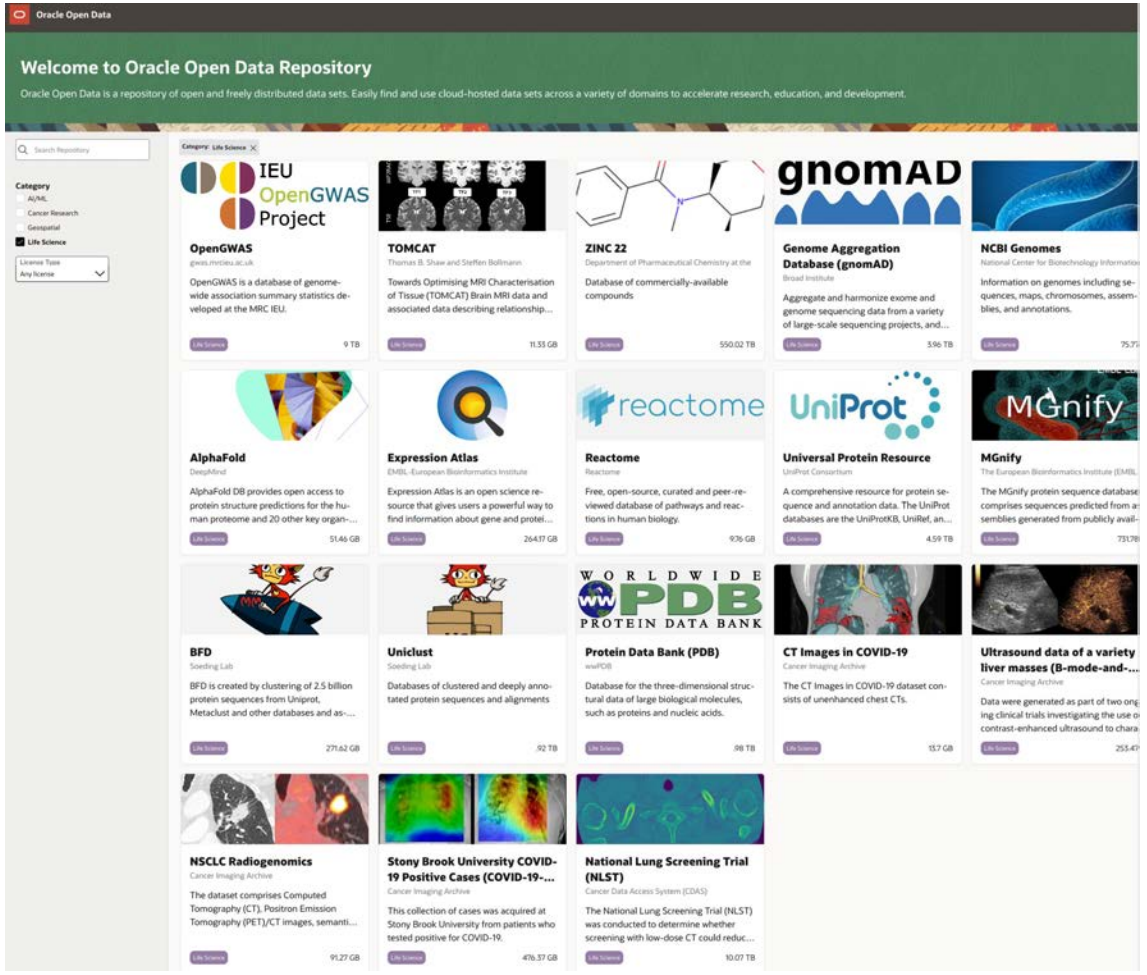
Curated, managed, and ready to use data sets from trusted institutions like NASA, DeepMind, Stanford.

Easy to use

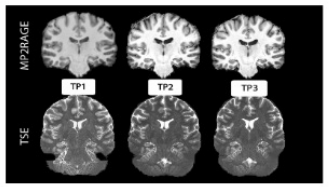
A platform that is easy to navigate, search, and ability to download files anywhere.

Low Cost, On-Premise & Cloud Re

Each data set consists of code and tooling usage examples for consumption and reproducibility.



https://docs.oracle.com/en/programs/research/oracle-open-data/



TOMCAT

Seven healthy participants were scanned using a Siemens Magnetom 7 Tesla (T) whole-body research MRI scanner (Siemens Healthcare, Erlangen, Germany). The first scan session was acquired in 2016 (time point one), the second and third session in 2019 (time point two and three, respectively) with the third session acquired 45 min following the second as a scan-rescan condition. The following scans were acquired for all time points: structural T1 weighted (T1w) MP2RAGE, high in-plane resolution Turbo-Spin Echo (TSE) dedicated for hippocampus subfield segmentation.

About Data Code

About this Data Set

Owner
Thomas B. Shaw and Steffen Bollmann

Owner Contact Information
t.shaw@uq.edu.au (Thomas B. Shaw)
s.bollmann@uq.edu.au (Steffen Bollmann)

Category
Life Science

Files
170

Size
11.33 GB

License Type
BSD 3-Clause "New"/"Revised" License

Tags
Life Science | Hippocampus | MRI/Brain

Oracle Open Data

Bulk downloads

Data Format

Download sample files

https://objectstorage.us-ashburn-1.oraclecloud.com/n/idrvm4tkz2a8/b/TOMCAT/o/example_runSingleSubjectTemplate.tar.gz

List files

HTTP CURL WGET

List files

```
curl https://objectstorage.us-ashburn-1.oraclecloud.com/n/idrvm4tkz2a8/b/TOMCAT/o/
```

List files with prefix

```
curl https://objectstorage.us-ashburn-1.oraclecloud.com/n/idrvm4tkz2a8/b/TOMCAT/o/?prefix=
```

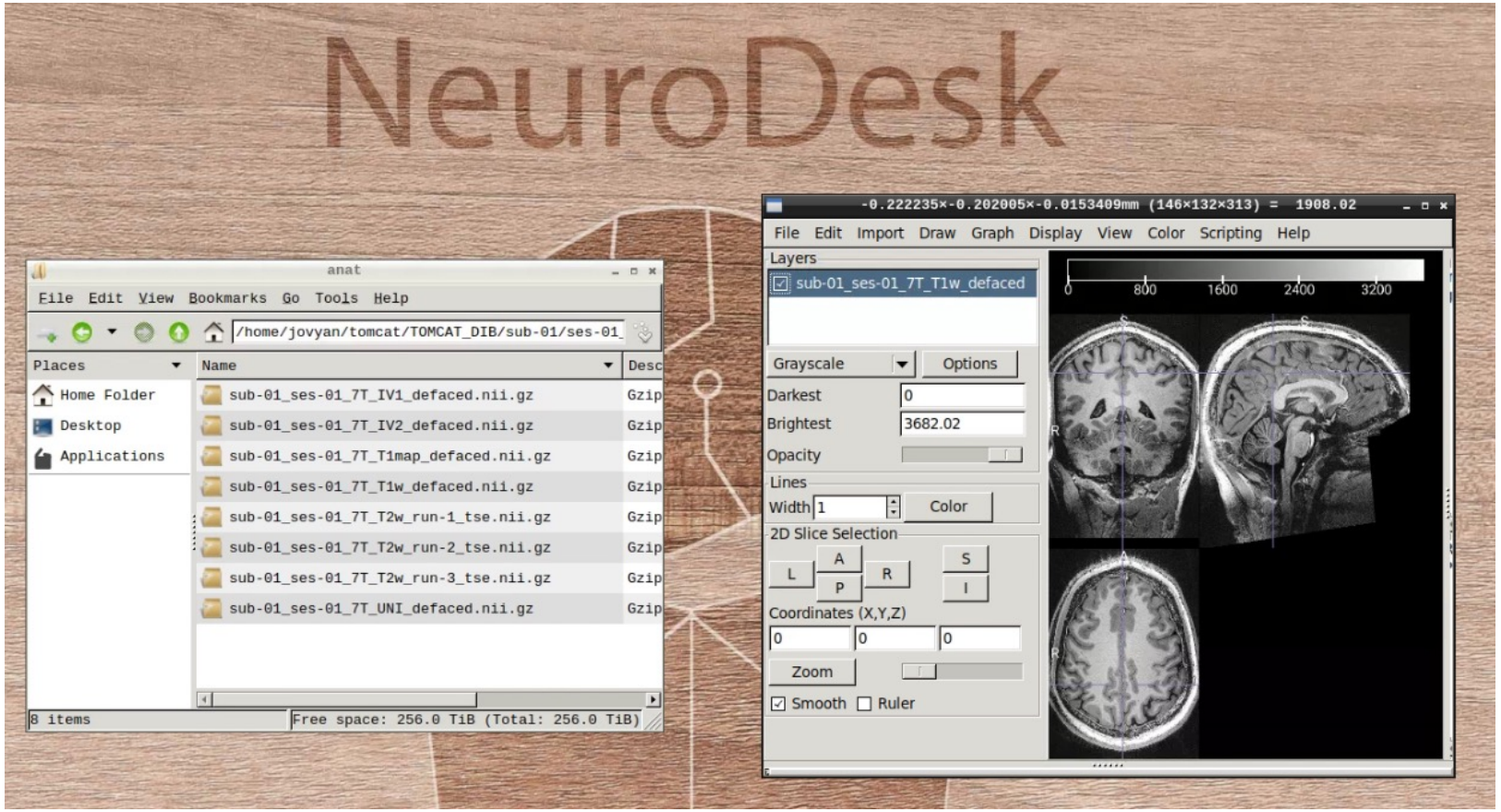
Browse and download

Download | Search keywords in all columns | Go | Filters | Rows 15

Name	Size
example_run.tar.gz	2.59 GB
example_runSingleSubjectTemplate.tar.gz	1.16 GB
TOMCAT_DIB/sub-01/ses-01_7T/anat/sub-01_ses-01_7T_IV1_defaced.nii.gz	44.14 MB
TOMCAT_DIB/sub-01/ses-01_7T/anat/sub-01_ses-01_7T_IV2_defaced.nii.gz	50.82 MB
TOMCAT_DIB/sub-01/ses-01_7T/anat/sub-01_ses-01_7T_T1w_defaced.nii.gz	69.32 MB

Oracle Open Data mounts on NeuroDesk

```
jovyan@neurodesktop:~$ s3fs TOMCAT ~/tomcat/ -o allow_other -o endpoint=us-ashburn-1 -o
url=https://idrvm4tkz2a8.compat.objectstorage.us-ashburn-1.oraclecloud.com/ -onomultipar
t -o use_path_request_style
```



Re-executable papers linking Open Code and Data

How close are we to a re-executable paper?

Can we use neuroimaging software inside Jupyter notebooks?

Interactive papers

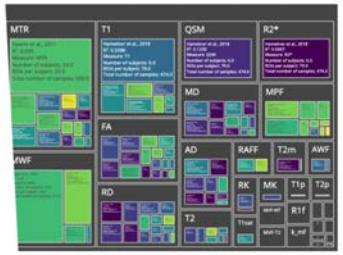
CONP PCNO

Supported by the Canadian Open Neuroscience Platform (CONP).

The quest for measuring myelin with MRI - An interactive meta-analysis

This study explores an important aspect of quantitative magnetic resonance imaging (qMRI): validation. Focusing specifically on myelin measures, we show the results of our meta-analysis comparing quantitative MRI with histology.

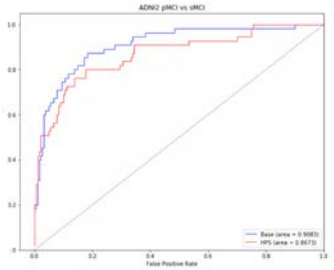
NeuroLibre Book GitHub Code



A highly predictive signature (HPS) of Alzheimer's disease dementia from cognitive and structural brain features

A jupyter notebook containing analyses that give a highly predictive signature (HPS) of Alzheimer's disease dementia from cognitive and structural features using simulated data.

NeuroLibre Book GitHub Code

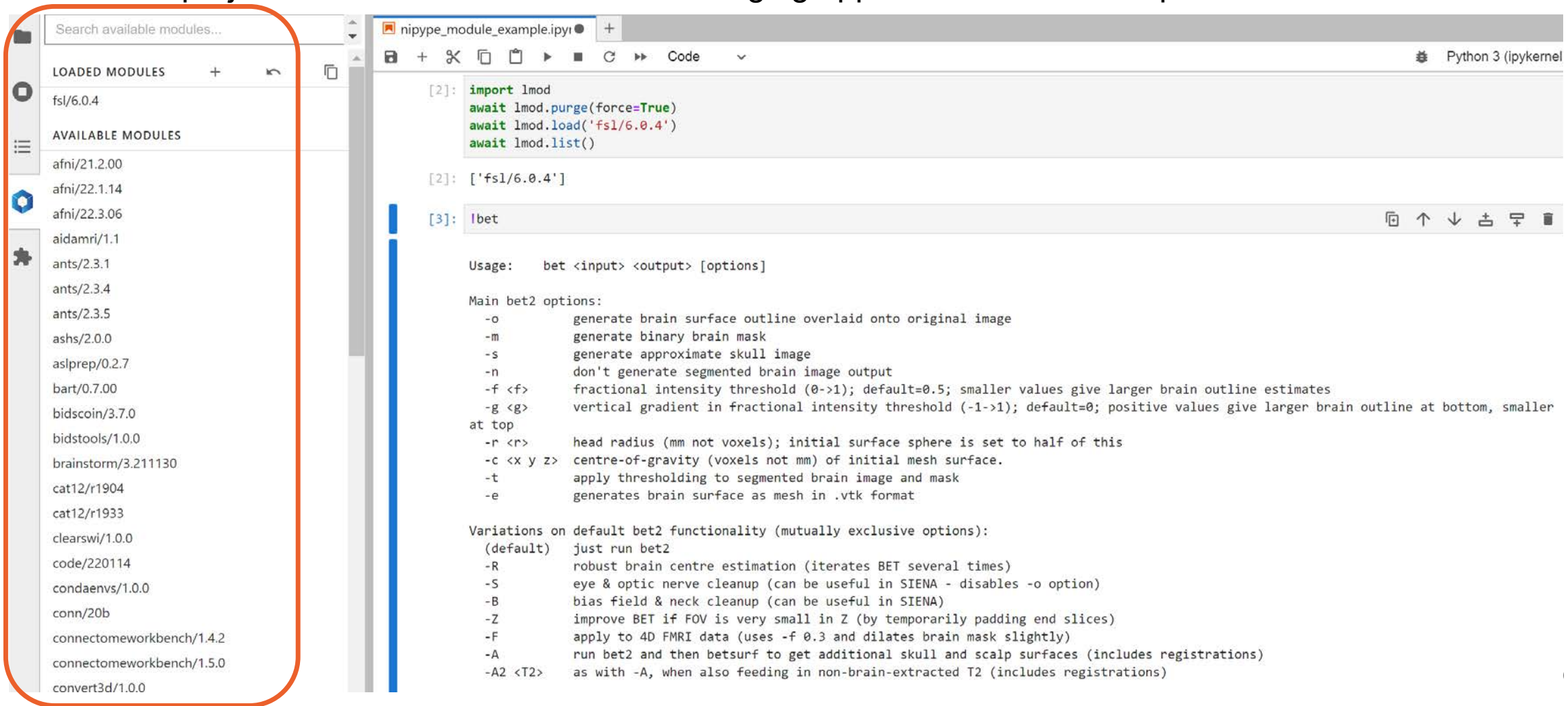


https://www.neurolibre.com/

https://elifesciences.org/labs/dc5acbde/welcome-to-a-new-era-of-reproducible-publishing

NeuroDesk applications & Jupyter notebooks

NeuroDesk project enables the use of all neuroimaging applications inside computational notebooks:



Interactively running in browser – no setup needed

Platform can be changed: link in OSF

Provide source code in an easy accessible way

This commit was used for the paper, but bug fixes possible

Data and links can be updated if bugs found or services move

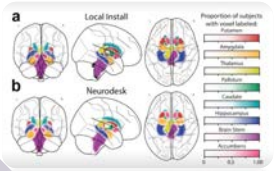
DATA AVAILABILITY STATEMENT

We facilitate the reproducibility of our study by providing an interactive version of our implementation on a publicly accessible cloud-based platform. The readers can explore the implementation of the model (neural network), train the model with different hyper-parameters and architectures, investigate the stability of the training process, and reproduce our results with the identical model used in this manuscript (https://github.com/sbollmannMRI/scout2B1, 320a6ab). We anonymized and stored the input data (localizer, SA2RAGE B_1^+) of 28 participants in OSF (OSF, Center for Open Science, Inc., Virginia, USA) accessible via https://osf.io/y5cq9/.

Thank you

- s.bollmann@uq.edu.au
- https://mri.sbollmann.net
- https://neurodesk.org
- @Sbollmann_MRI@masto.ai
- @NeuroDesk@masto.ai
- github.com/sbollmannmri
- github.com/NeuroDesk

⬇ More presentations from our group @MRItogether ⬇



Ashley Stewart - QSMxT: An Open Pipeline for Automated Quantitative Susceptibility Mapping

Dao Thanh Thuy - Investigating the reproducibility of the NeuroDesk Platform

Fernanda Ribeiro - An open-source framework for predicting brain functional maps with geometric deep learning

Korbinian Eckstein - Open-Source MRI Tools for Research (ROMEO, CLEAR-SWI and MCPC-3D-S)