Toolkits for open science: Data Structure and Containers

Dr. Joanes Grandjean

KEITH HARING and ANDY WARHOL, Andy Mouse Series

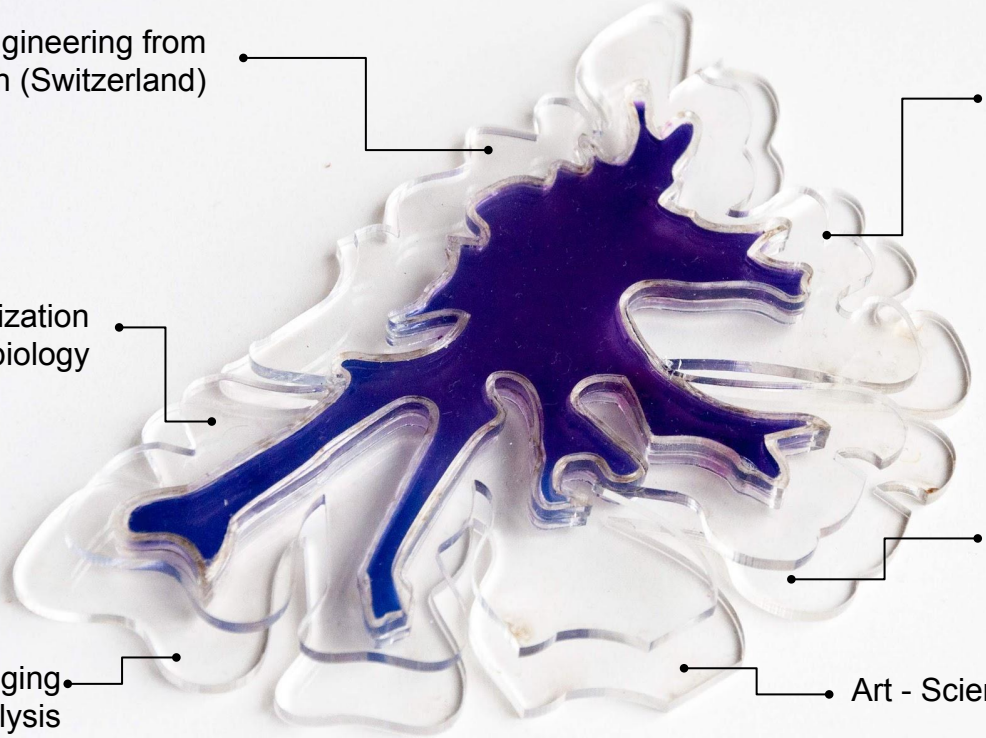
DONDERS INSTITUTE

Radboud University Radboudumc

PhD in biomedical engineering from ETH Zurich (Switzerland)

Investigates brain dysfunction in stress disorders

Msc in Biology, specialization in neurobiology



Functional MRI in animals

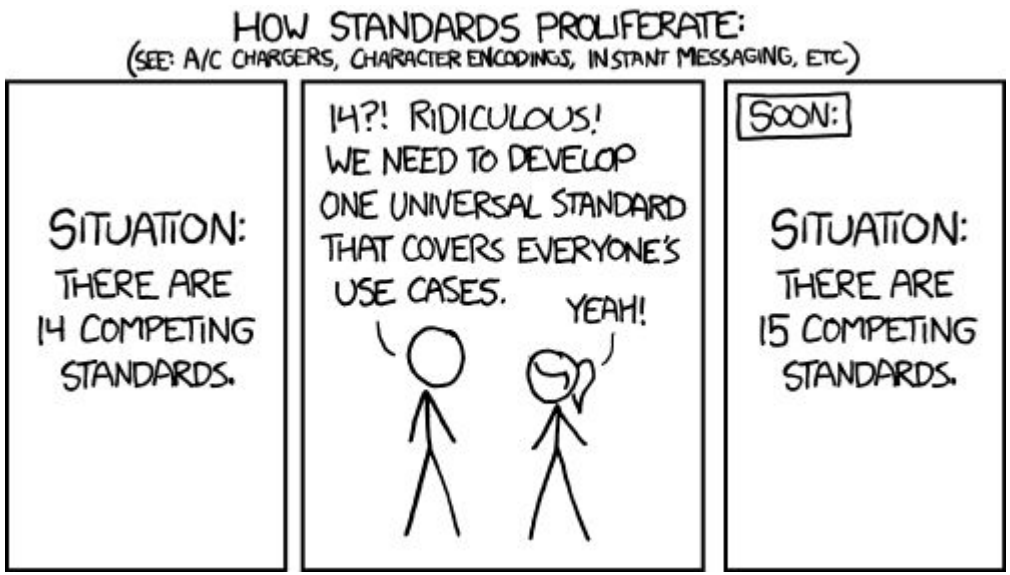
Interested in imaging technologies and analysis

Art - Science - Coffee

Clarke and Grandjean 2020

Problematic : How to optimally format data for sharing and re-use?

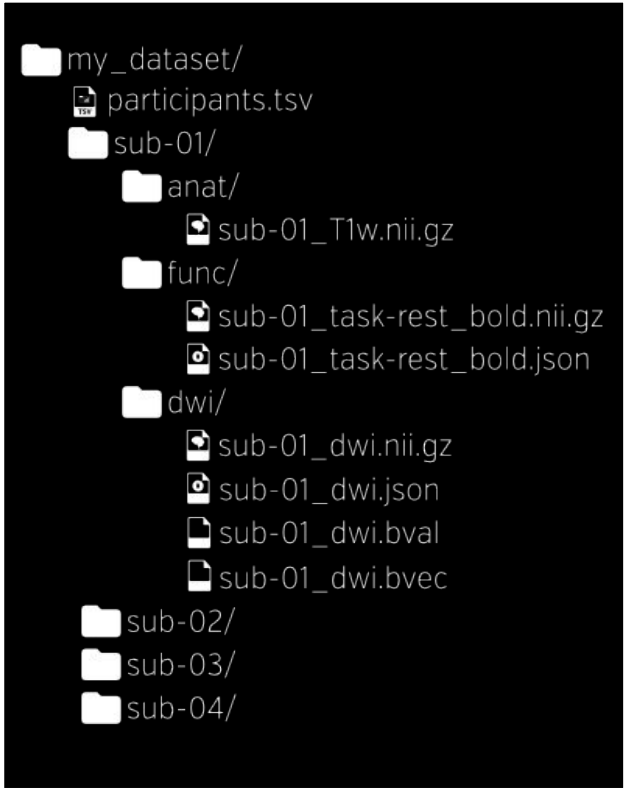
- Neuroimaging data comes with varying formats (e.g. NIFTI, analyze, MINC, DICOM)
- Meta-data is important to make sense of the imaging data (e.g. TR, TE, b-value)
- Machine readable file names
- **The Brain Imaging Data Structure brings is the solution!**



XKCD

The Brain Imaging Data Structure in a nutshell

- Data is organized by subject
- Optional session sub-folders
- Distinct folders for anatomical, functional, diffusion ….
- File names is self-documenting
- Accompanying metadata file in .json file
- Hyphens and underscore have dedicated use!



https://bids.neuroimaging.io/

Converting data to BIDS

- BIDScoin is developed at the Donders Institute
- Converts dicom data to BIDS

- Brkraw is developed at UNC Chapel Hill
- Converts raw Bruker data to BIDS

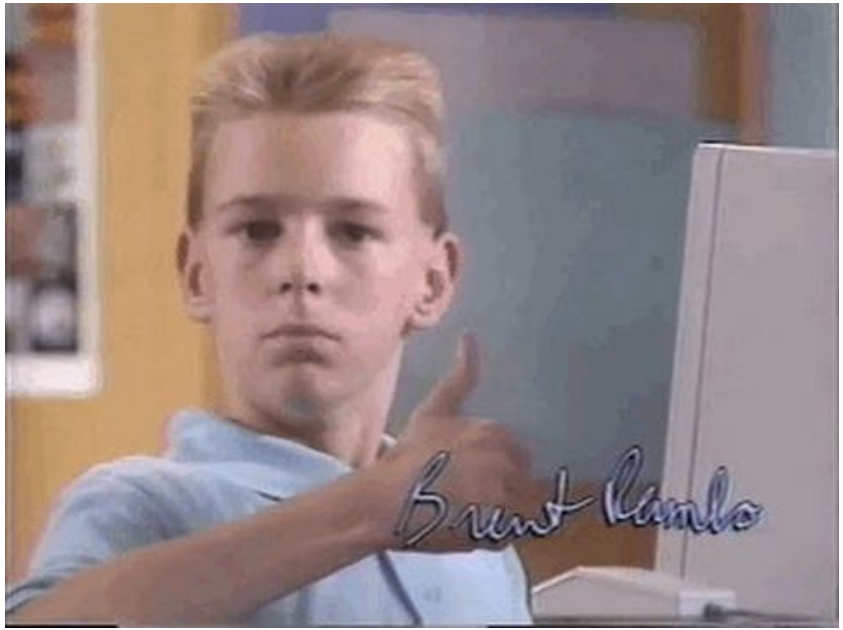
BRKRAW

https://bidscoin.readthedocs.io/en/latest/ https://github.com/BrkRaw/brkraw

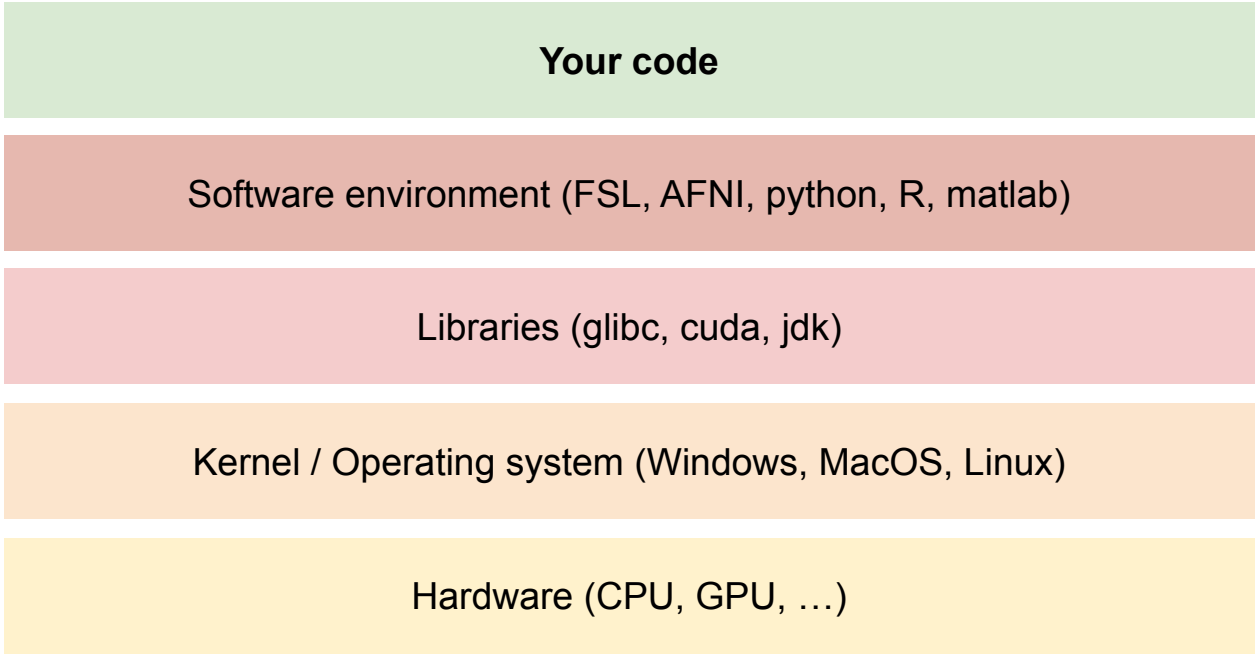
Problematic : How to optimally format analysis code for sharing and re-use?

- Analysis depends on several toolboxes (e.g FSL, AFNI).
- Each toolbox are available with various versions
- Each toolbox have their own dependencies
- Each toolbox can be installed different ways
- Analysis are performed on different operating systems
- Installing new software is a risk for your system

- **How to safely re-run a procedure the exact same way, irrespective of the machine??**



Portable and reproducible software environment



Portable and reproducible software environment

Your code

Software environment (python, R)

Portable and reproducible software environment

sessionInfo.txt 2.56 KiB

```
R version 3.5.1 (2018-07-02)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: CentOS Linux 7 (Core)

Matrix products: default
BLAS: /opt/R/3.5.1/lib64/R/lib/libRblas.so
LAPACK: /opt/R/3.5.1/lib64/R/lib/libRlapack.so

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C               LC_TIME=en_US.UTF-8
 [4] LC_COLLATE=en_US.UTF-8     LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C                  LC_ADDRESS=C
[10] LC_TELEPHONE=C             LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] effectsize_0.4.4-1 parameters_0.13.0  multcomp_1.4-10    TH.data_1.0-10
 [5] MASS_7.3-51.1      survival_2.44-1.1  mvtnorm_1.0-11     extrafont_0.17
 [9] reshape2_1.4.3     oro.nifti_0.11.0   nat_1.8.14         rgl_0.100.50
[13] ggplot2_3.1.0      stringr_1.4.0
```

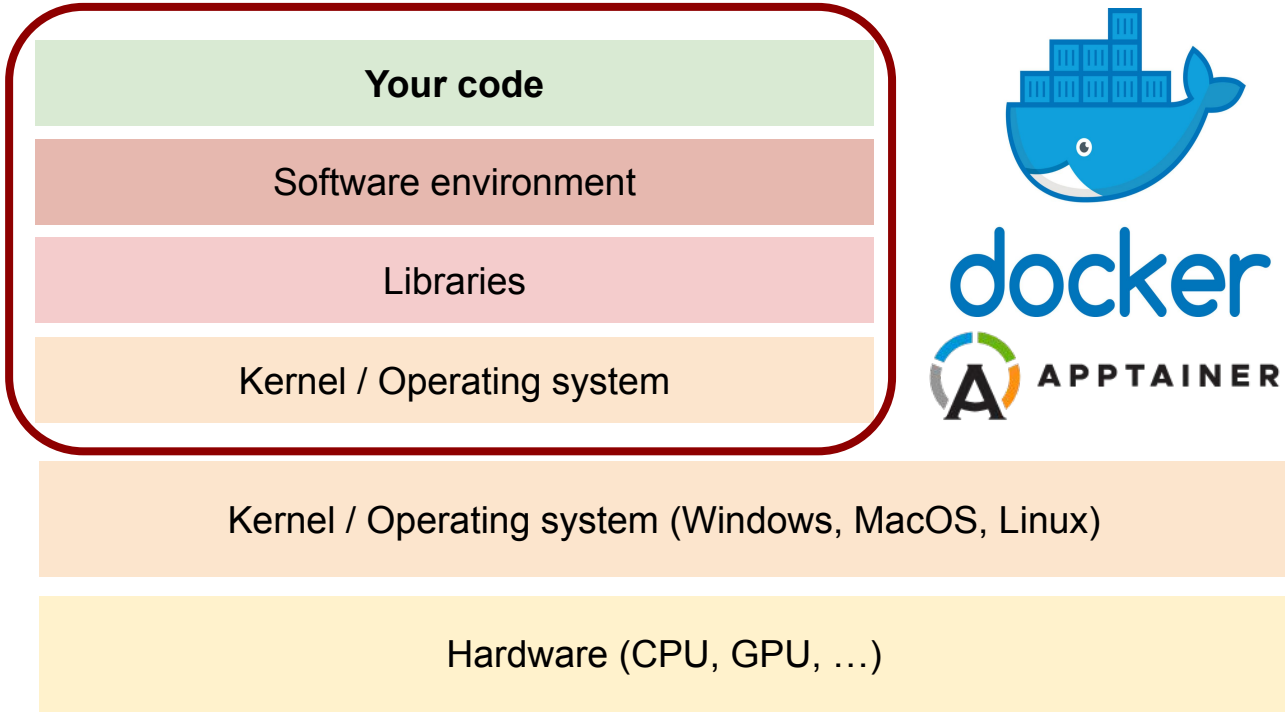
https://gitlab.socsci.ru.nl/preclinical-neuroimaging/insula

Portable and reproducible software environment

docker

APPTAINER

Portable and reproducible software environment



Getting started with Docker / Apptainer, the definition file

ubuntu_nvim_quatro > apptainer.def

```
1   BootStrap: docker
2   From: ubuntu:latest
3
4   %environment
5           PATH=/bin:/sbin:/usr/bin:/usr/sbin:/usr/local/bin:/usr/local/sbin
6           LANG=C.UTF-8 LC_ALL=C.UTF-8
7
8   %post
9
10          apt-get install -y --no-install-recommends cmake git r-base r-base-dev r-cran-tidyverse
11
```

<- Use Linux Ubuntu provided by docker as the basis image

<- Define environment, e.g. language, time zone, etc.

<- Give instructions to install programs, e.g. R and tidyverse packages

My own apptainer recipe

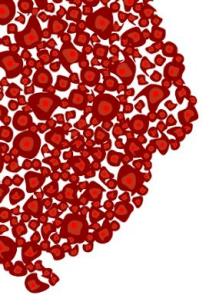
Easy Docker / Apptainer containers with Neurodocker

Welcome to Neurodocker!

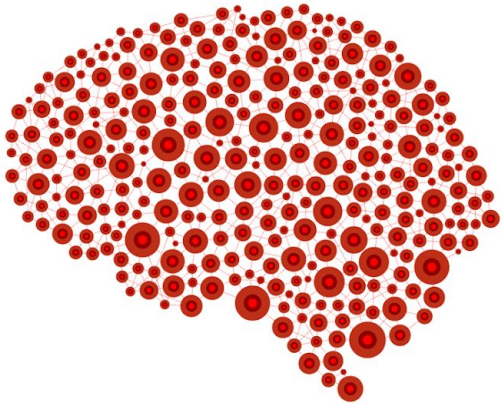
Neurodocker is a command-line program that generates custom Dockerfiles and Singularity recipes for neuroimaging and minifies existing containers. Its purpose is to make it easier for scientists (and others) to easily create reproducible computational environments.

```
neurodocker generate singularity --pkg-manager apt \
    --base-image neurodebian:buster \
    --ants version=2.3.4 \
    --miniconda version=latest conda_install="nipype notebook" \
    --user nonroot
```

https://github.com/ReproNim/neurodocker

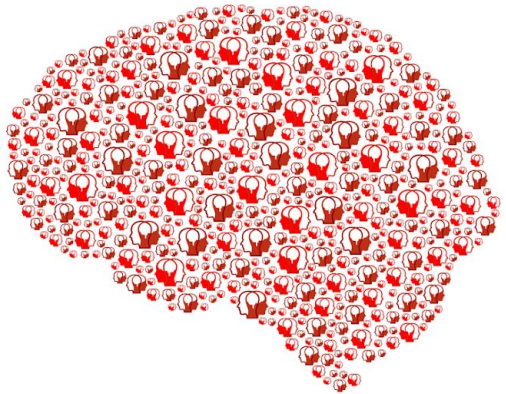
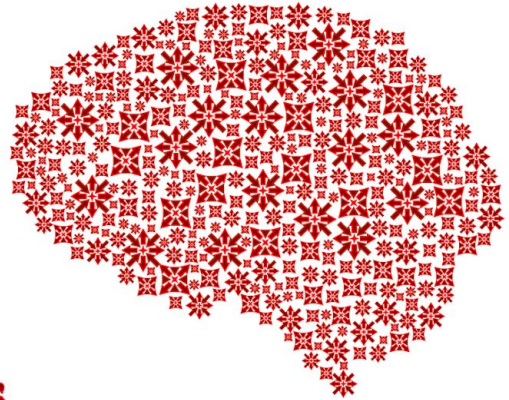


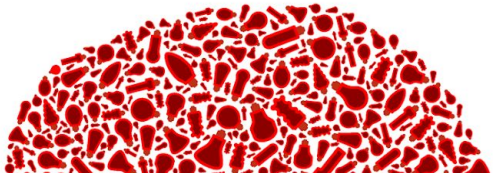

DONDERS INSTITUTE



www.ru.nl/donders

grandjeanlab.github.io



https://github.com/grandjeanlab

@grandjeanlab

Radboud University Radboudumc