



HIGH PERFORMANCE SCIENCE

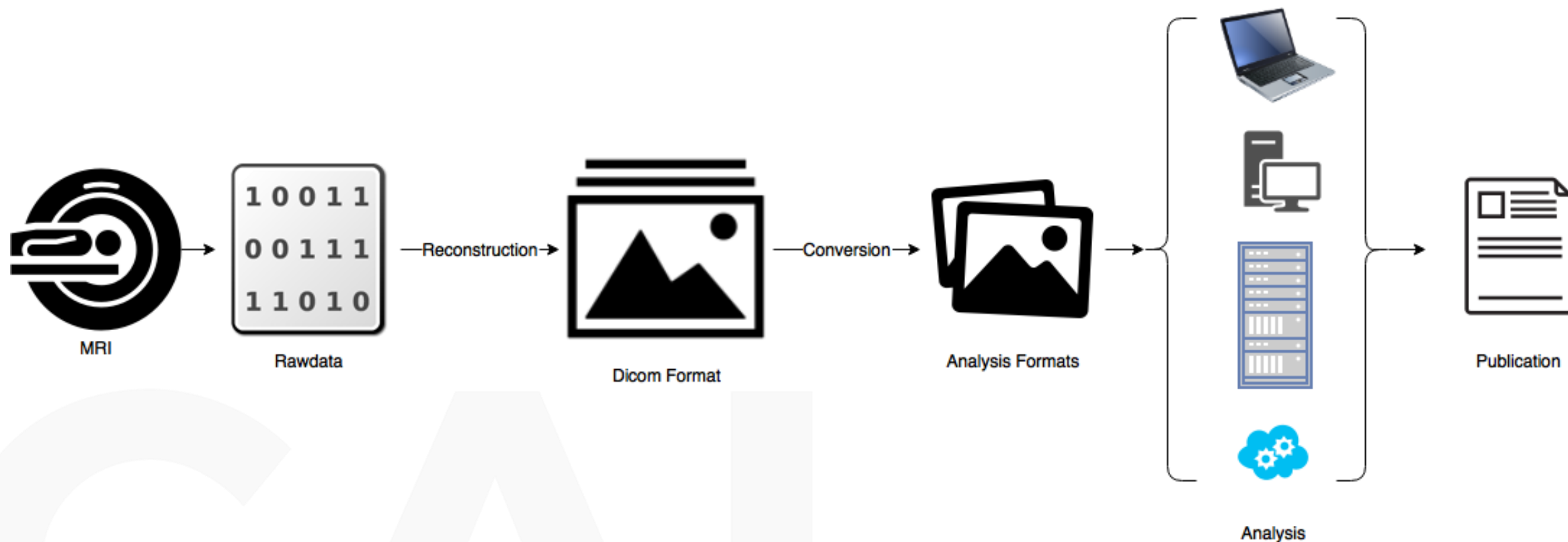
open and reproducible
neuroscience image
processing that scales?

WHOAMI

- Researcher in magnetic resonance imaging @Centre for Advanced Imaging, University of Queensland
- National Imaging Facility Fellow @NIF Brisbane Australia
- Creating images of the human brain



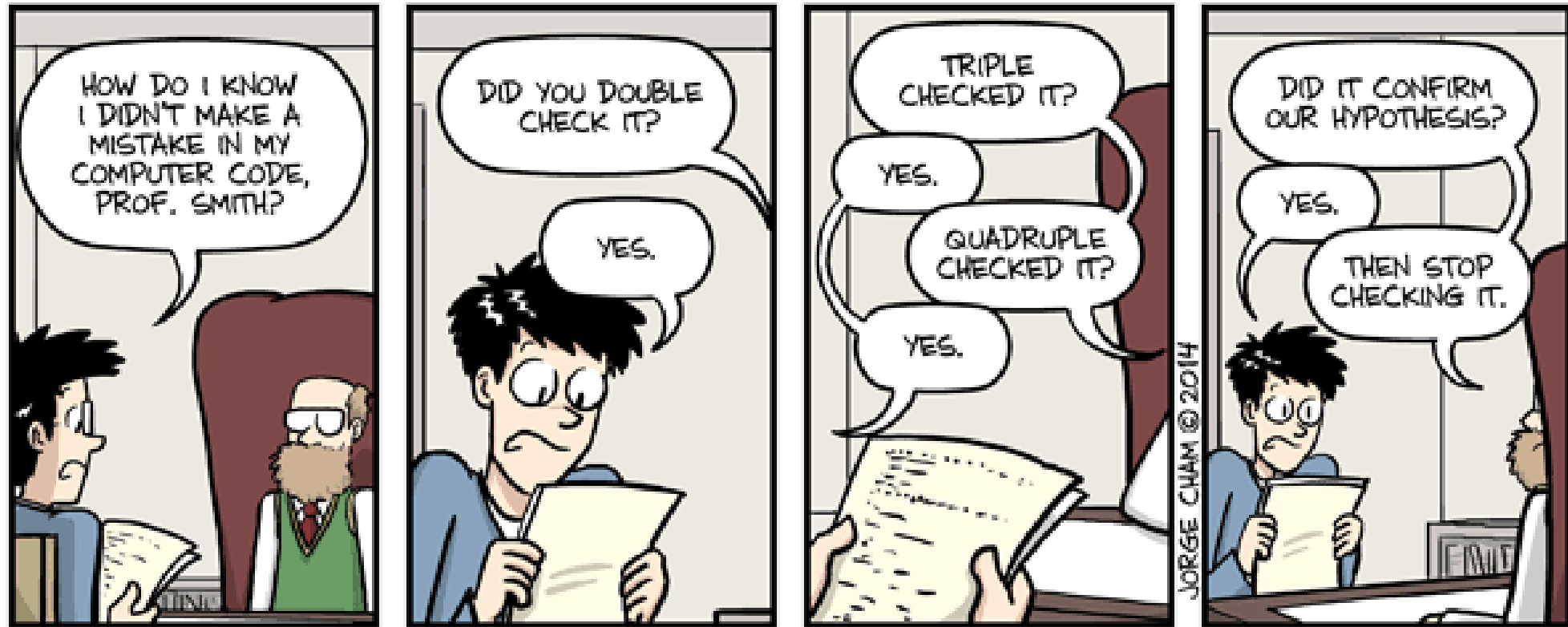
A LONG WAY TO PUBLICATION



WHAT CAN POSSIBLY GO WRONG?

- Scientists write software, but are not trained in software development
- There will be bugs – tricky to find if plausible results

SELECTIVE DEBUGGING



National
Imaging
Facility

WHAT CAN POSSIBLY GO WRONG?

- Scientists write software, but are not trained in software development
- There will be bugs – tricky to find if plausible results
- Large amounts of data
- Analyses run on different hardware and operating systems

GLIBC 2.5 VS 2.18

Reproducibility of neuroimaging analyses across operating systems

Tristan Glatard^{1,2}, Lindsay B. Lewis¹, Rafael Ferreira da Silva³, Reza Adalat¹, Natacha Beck¹, Claude Lepage¹, Pierre Rioux¹, Marc-Etienne Rousseau¹, Tarek Sherif¹, Ewa Deelman³, Najmeh Khalili-Mahani¹ and Alan C. Evans^{1}*

```
expf(1.540518522262573242187500000000)  
=4.6670093536376953125000
```

```
expf(1.540518522262573242187500000000)  
=4.6670098304748535156250
```

- glibc 2.5 vs 2.18 deliver different floating-point results
- leads to significant differences in long pipelines



WHAT CAN POSSIBLY GO WRONG?

- Scientists write software, but are not trained in software development
- There will be bugs – tricky to find if plausible results
- Large amounts of data
- Analyses run on different hardware and operating systems
- Difficult to share data and reproduce an analysis

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund^{a,b,c,1}, Thomas E. Nichols^{d,e}, and Hans Knutsson^{a,c}

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

OUR JOURNEY BEGINS



<https://cai.centre.uq.edu.au/facilities/human-imaging/3t-magnetom-prisma>



National
Imaging
Facility

Steffen.Bollmann@cai.uq.edu.au

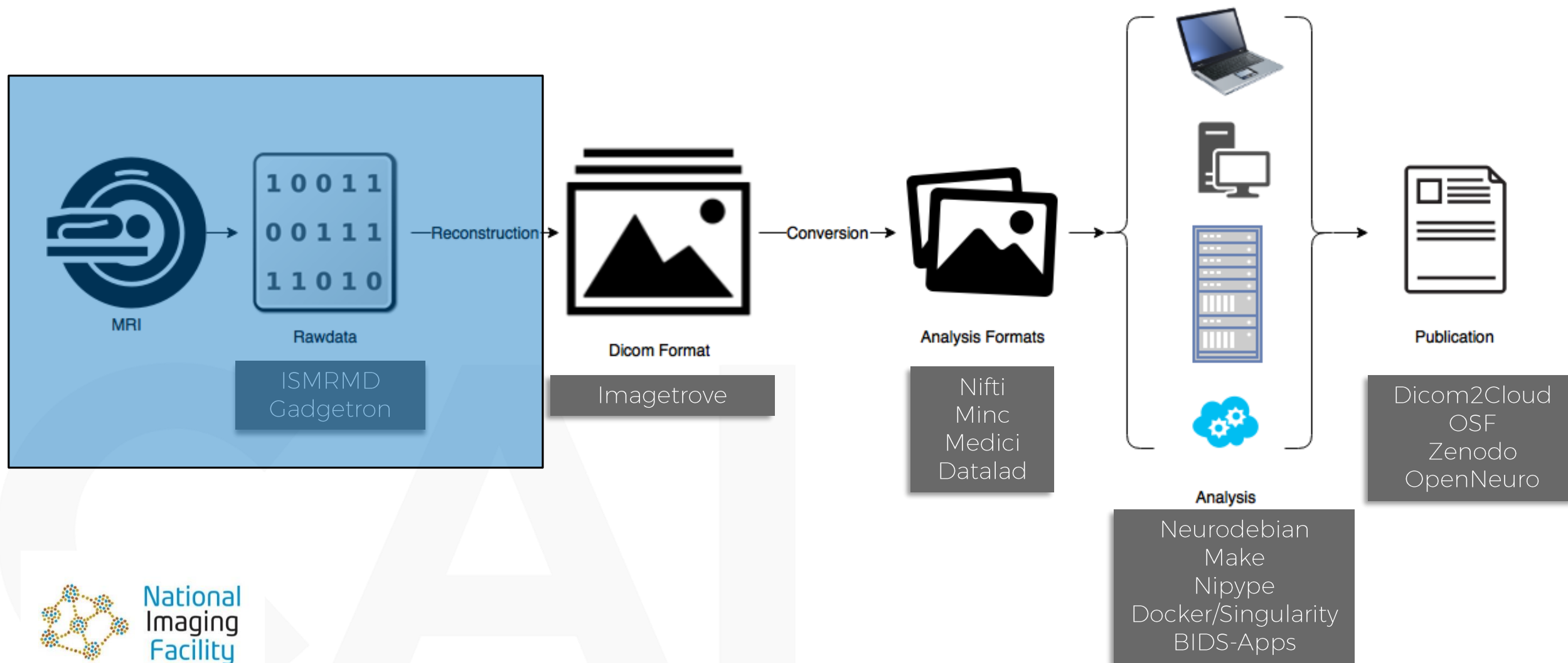


THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

Create change

Slide 9

A LONG WAY TO PUBLICATION



ISMRRMRD

ISMRM Raw Data Format: A Proposed Standard for MRI Raw Datasets

Souheil J. Inati,¹ Joseph D. Naegele,¹ Nicholas R. Zwart,² Vinai Roopchansingh,¹ Martin J. Lizak,³ David C. Hansen,⁴ Chia-Ying Liu,⁵ David Atkinson,⁶ Peter Kellman,⁷ Sebastian Kozerke,⁸ Hui Xue,⁷ Adrienne E. Campbell-Washburn,⁷ Thomas S. Sørensen,⁹ and Michael S. Hansen^{7*}

- raw data is vendor specific
- ISMRMRD: Open format based on HDF5 + routines for converting from any closed source vendor format

ISMRRMRD

- XML header
- Raw Data

GADGETRON

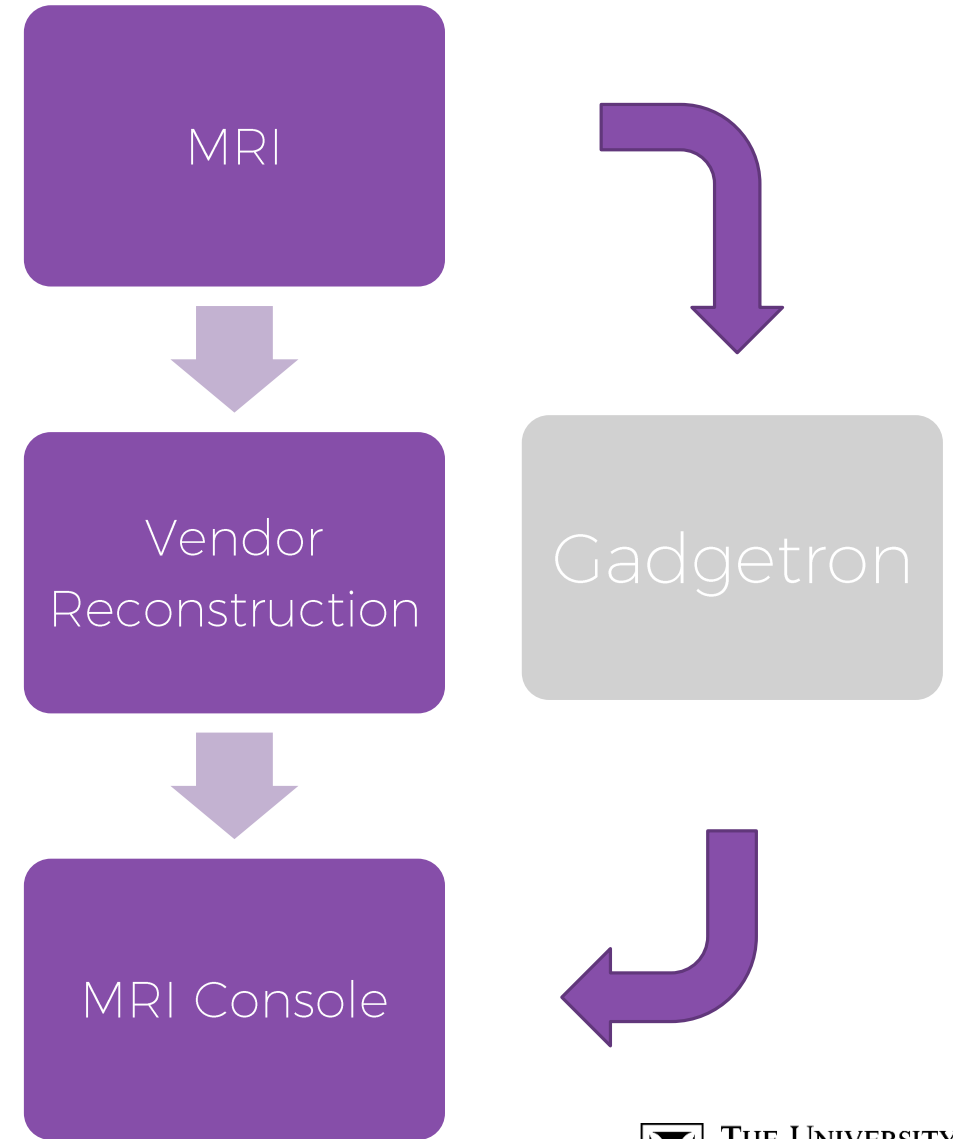
Gadgetron: An Open Source Framework for Medical Image Reconstruction

Michael Schacht Hansen^{1*} and Thomas Sangild Sørensen^{2,3}

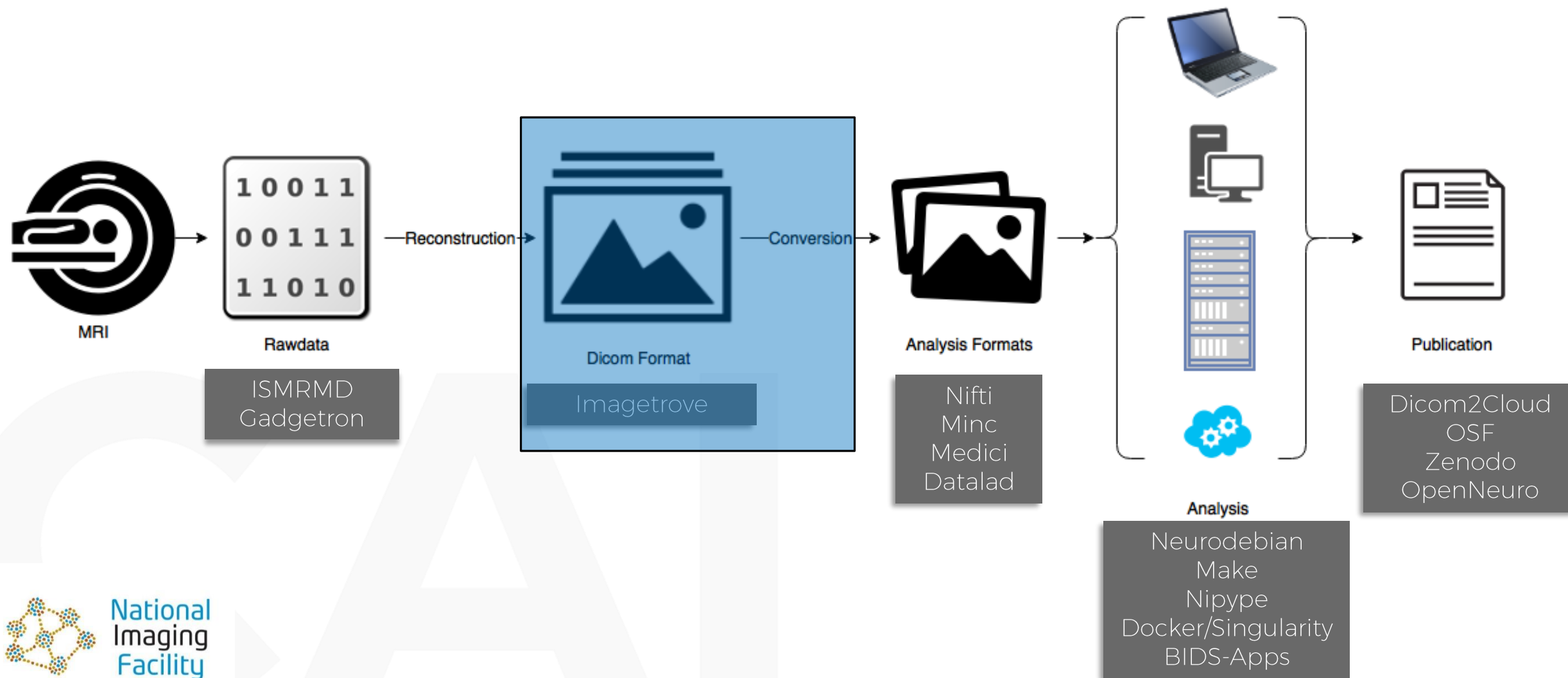
- Reconstructing raw data is vendor specific
- Gadgetron: Open source reconstruction system compatible with ISMRMRD



<https://github.com/gadgetron>

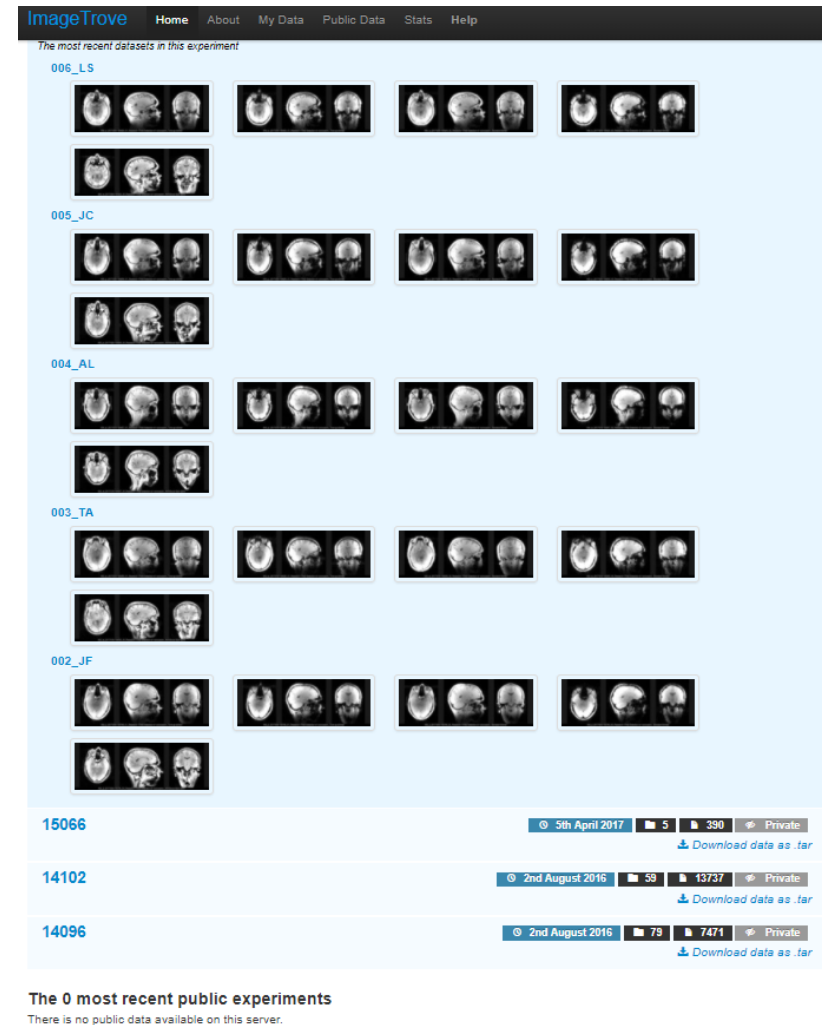


A LONG WAY TO PUBLICATION

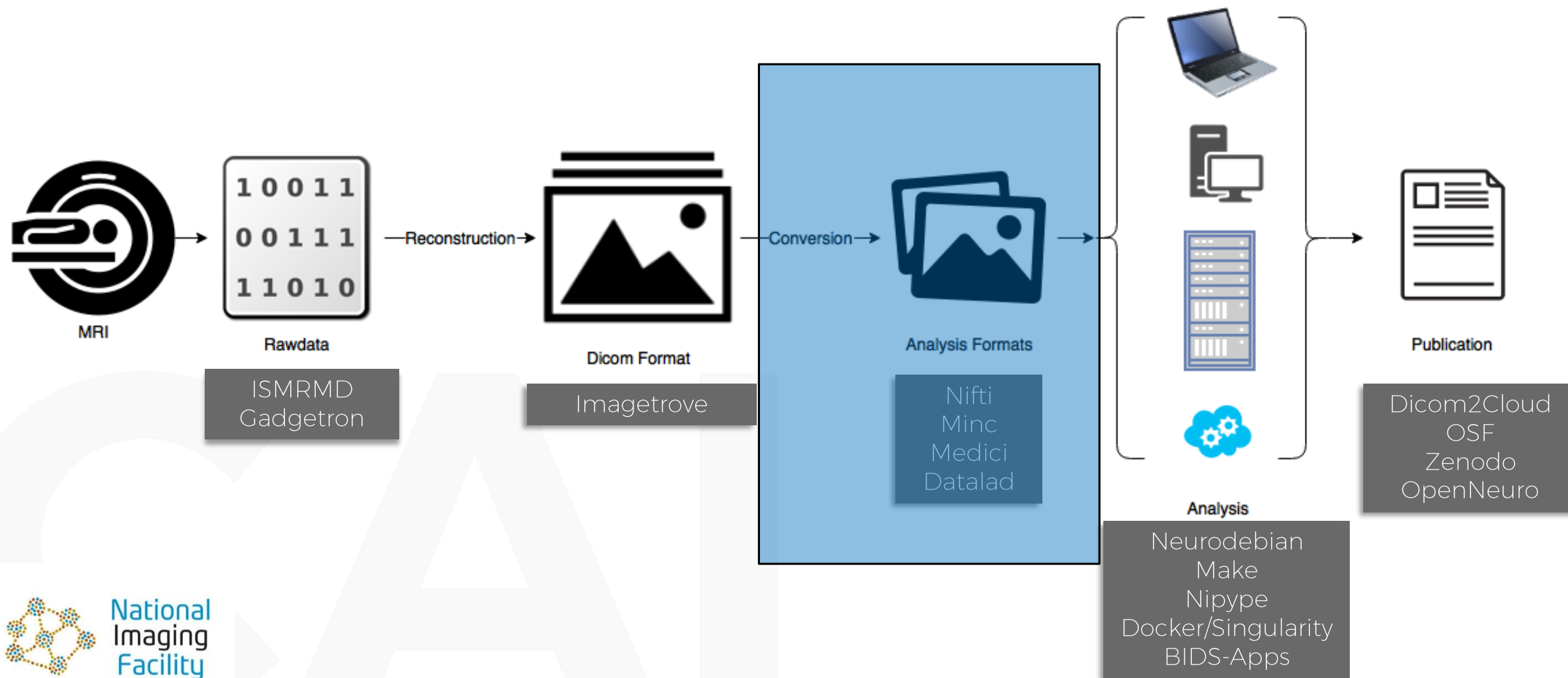


CURATING DICOM DATA

- imaging data is stored in DICOM format, which is difficult to archive, curate and process
- **Imagetrove and Trusted Data Repositories based on Mytardis**
 - interface for data upload
 - curation of data and conversion to analysis formats (nii, minc)
 - Quality Control and Assurance



A LONG WAY TO PUBLICATION

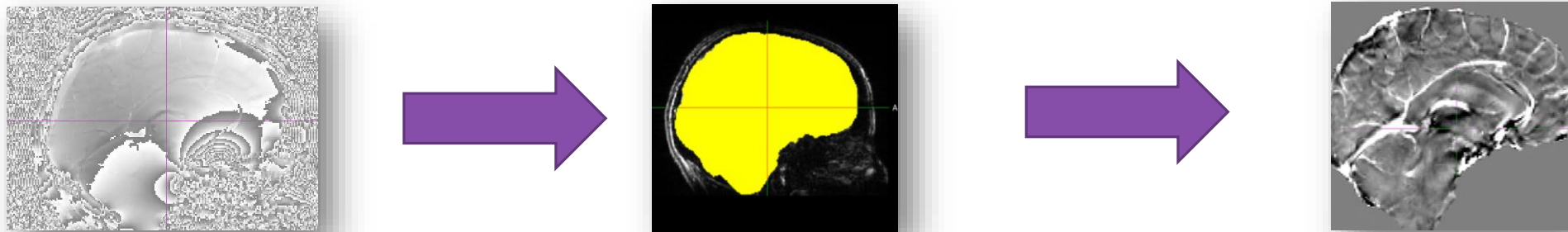


METROPOLITAN DATA CACHING INFRASTRUCTURE (MEDICI)



DATA PROVENANCE

- Data Provenance / Lineage = keeping track of what happens to data during an analysis



```
[fsluser@localhost data]$ bet mag.nii.gz mask -n -m -R -f 0.1 -g 0.0
```

```
[fsluser@localhost TGVQSM-master-011045626121baa8bfdd6633929974c732ae35e3]$  
tgv_qsm -p ../phase.nii.gz -m ../mask_mask.nii.gz -f 2.89 -t 0.02 -s -o qsm
```


MINC: BUILD-IN PROVENANCE

MINC 2.0: A Flexible Format for Multi-Modal Images

Robert D. Vincent¹, Peter Neelin², Najmeh Khalili-Mahani¹, Andrew L. Janke³, Vladimir S. Fonov¹, Steven M. Robbins¹, Leila Baghdadi⁴, Jason Lerch^{4,5}, John G. Sled^{4,5}, Reza Adalat¹, David MacDonald⁶, Alex P. Zijdenbos⁷, D. Louis Collins^{1,8} and Alan C. Evans^{1}*

```
$ minchistory FA.reg.mnc
```

```
--- History of B027915.01_DTI.B0.reg.clp.mnc ---
```

```
[01] Thu Nov 21 14:28:14 2002>>> mincaverage -clobber \  
    /usr/people/steve/data/stroke/B027915/01_DTI/B027915.01_DTI.frame001.mnc \  
    /usr/people/steve/data/stroke/B027915/01_DTI/B027915.01_DTI.frame002.mnc \  
    /usr/people/steve/data/stroke/B027915/01_DTI/B027915.01_DTI.frame003.mnc \  
    /usr/people/steve/data/stroke/B027915/01_DTI/B027915.01_DTI.frame008.mnc \  
    /usr/people/steve/data/stroke/B027915/01_DTI/B027915.01_DTI.B0.mnc
```

```
[02] Mon Nov 3 16:54:24 2003>>> mincresample -clobber -like \  
    /home/rotor/data/stroke/B027915/01_DTI/B027915.01_DTI.B0.mnc \  
    -transformation \  
    /home/rotor/data/stroke/B027915/xfms/B027915.01_DTI.midline-align.xfm \  
    /home/rotor/data/stroke/B027915/01_DTI/B027915.01_DTI.B0.mnc \  
    /home/rotor/data/stroke/B027915/reg/B027915.01_DTI.B0.reg.mnc
```

```
[03] Mon Nov 10 23:50:21 2003>>> minccalc -clobber -outfile value \  
    /home/rotor/data/stroke/B027915/01_DTI/B027915.01_DTI.B0.reg.mnc
```

GIT | GIT-ANNEX | DATALAD

Feature	Git	git-annex	DataLad
Version control (text, code)	✓	✓	✓
Version control (binary data)	(not advised)	✓	✓
Auto-crawling available resources		✓ RSS feeds	✓ flexible
Unified dataset handling			✓
• recursive operation on datasets			✓
• seamless operation across datasets boundaries			✓
• metadata support		✓ per-file	✓
• metadata aggregation			✓ flexible
Unified authentication interface			✓

<http://datalad.org/for/git-users>

BRAIN IMAGING DATA STRUCTURE

OPEN

SUBJECT CATEGORIES

- » Data publication and archiving
- » Research data

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

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

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

BRAIN IMAGING DATA STRUCTURE

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

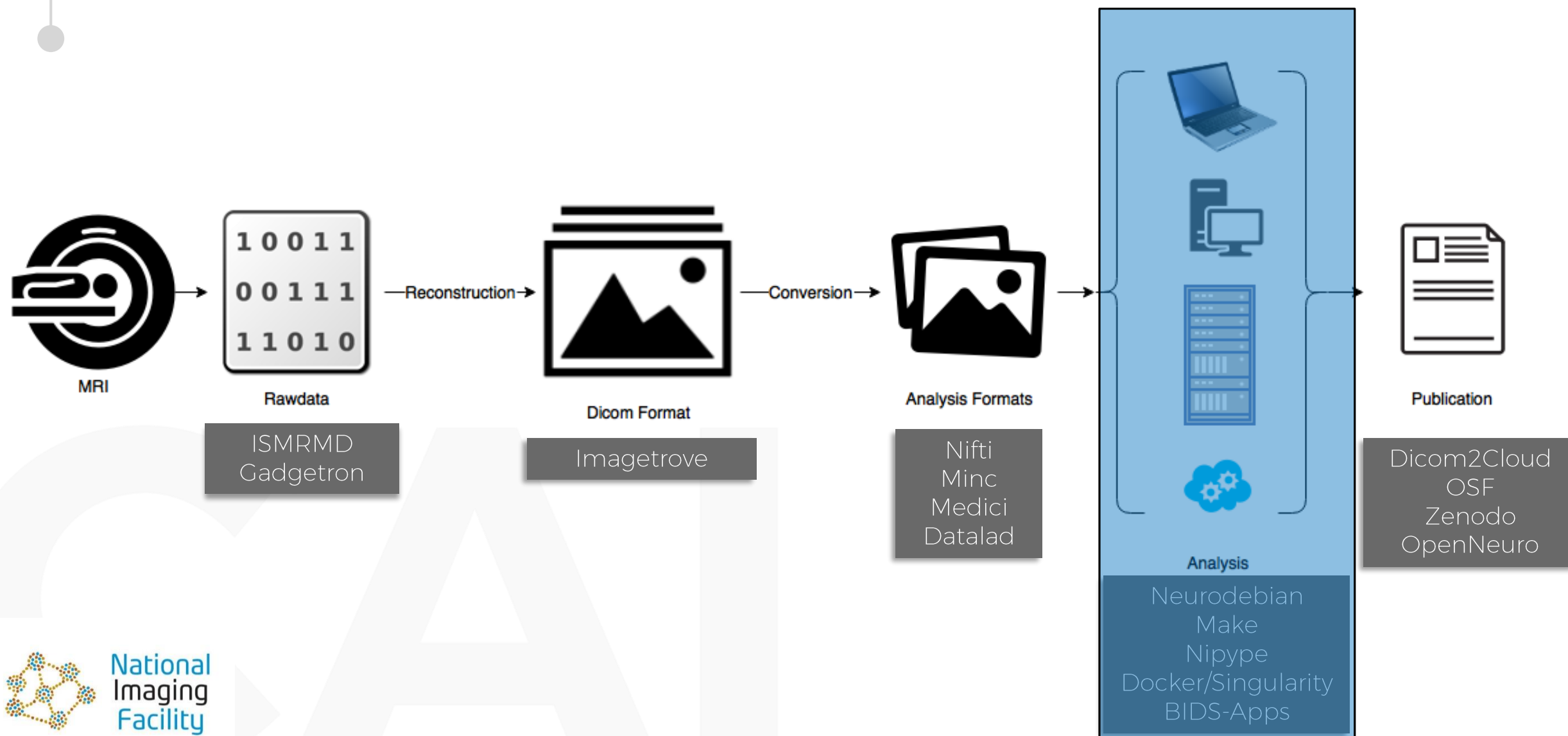
sub-01_dwi.bvec

sub-02/

sub-03/

sub-04/

A LONG WAY TO PUBLICATION



NEURODEBIAN

frontiers in
NEUROINFORMATICS

OPINION ARTICLE
published: 29 June 2012
doi: 10.3389/fninf.2012.00022



Open is not enough. Let's take the next step: an integrated, community-driven computing platform for neuroscience

Yaroslav O. Halchenko^{1,2,3 *†} and *Michael Hanke*^{4,5,3 *†}



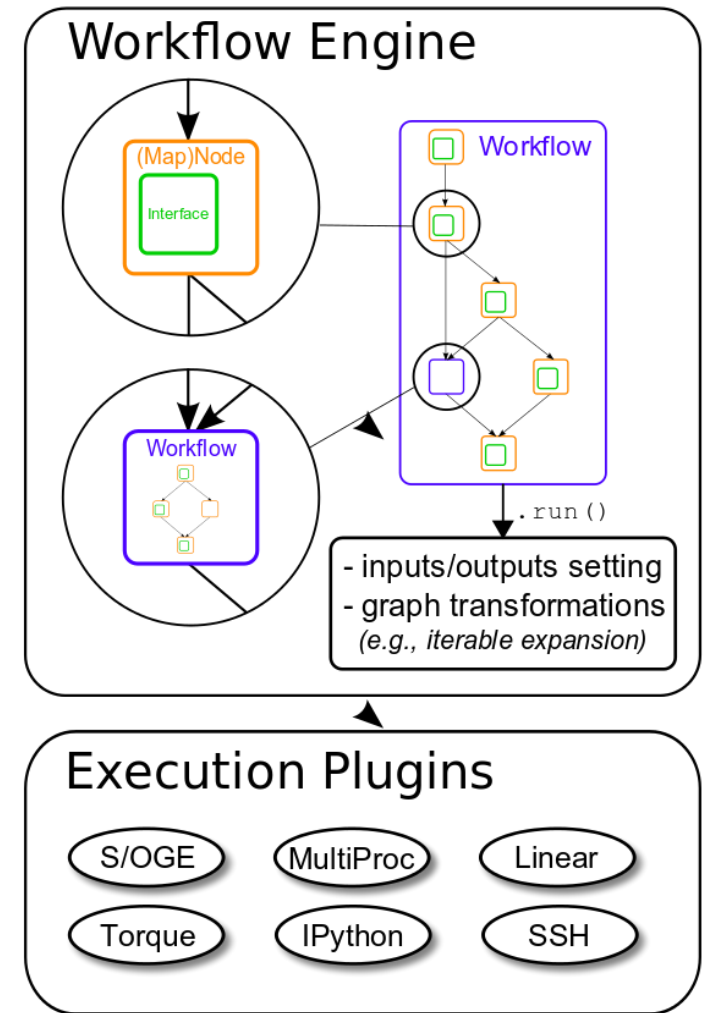
- Neurodebian curates software packages needed for neuroimaging (+ backport of new software)
- provides a virtual machine for running on Windows / Mac
- allows installing specific versions of packages for reproducing an analysis



<http://neuro.debian.net/>

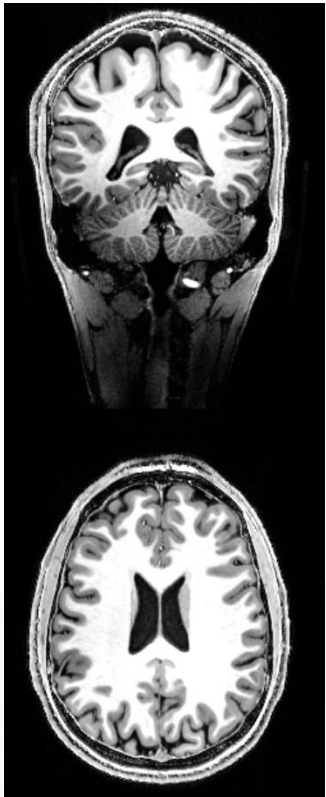
ANALYSIS PIPELINES

- Allow definition of analysis steps
 - Check if outputs exist and re-run based on dependencies
 - handle parallelization
-
- Bash, R, Matlab, Python
 - Make, Nipype, LONI pipeline, CBRAIN, OpenMOLE, Galaxy, Kepler ... (more than 100 ... <https://github.com/pditommaso/awesome-pipeline>)



<http://nipype.readthedocs.io/en/latest/>

MAKE FOR DATA ANALYSIS

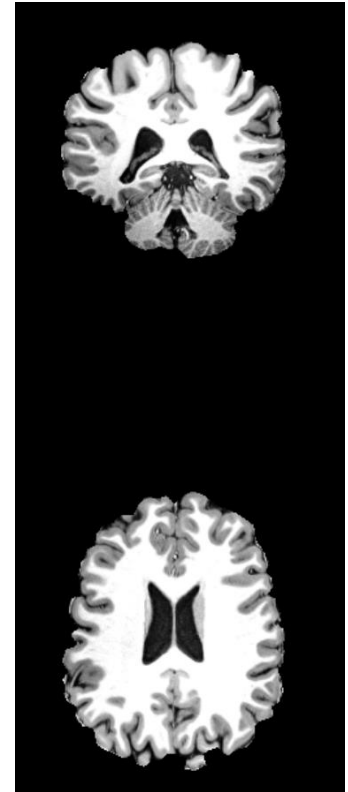


Using Make for Reproducible and Parallel Neuroimaging Workflow and Quality-Assurance

Mary K. Askren^{1*}, Trevor K. McAllister-Day¹, Natalie Koh¹, Zoé Mestre², Jennifer N. Dines¹, Benjamin A. Korman¹, Susan J. Melhorn³, Daniel J. Peterson¹, Matthew Peverill⁴, Xiaoyan Qin¹, Swati D. Rane¹, Melissa A. Reilly¹, Maya A. Reiter¹, Kelly A. Sambrook¹, Karl A. Woelfer¹, Thomas J. Grabowski^{1,5} and Tara M. Madhyastha¹

```
T1_brain.nii.gz: T1.nii.gz  
    bet T1.nii.gz T1_brain.nii.gz
```

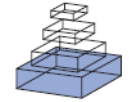
Example 1A | Makefile for skull stripping.



NIPYPE

frontiers in
NEUROINFORMATICS

METHODS ARTICLE
published: 22 August 2011
doi: 10.3389/fninf.2011.00013



Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python

Krzysztof Gorgolewski^{1}, Christopher D. Burns², Cindee Madison², Dav Clark³, Yaroslav O. Halchenko⁴, Michael L. Waskom^{5,6}, Satrajit S. Ghosh⁷*

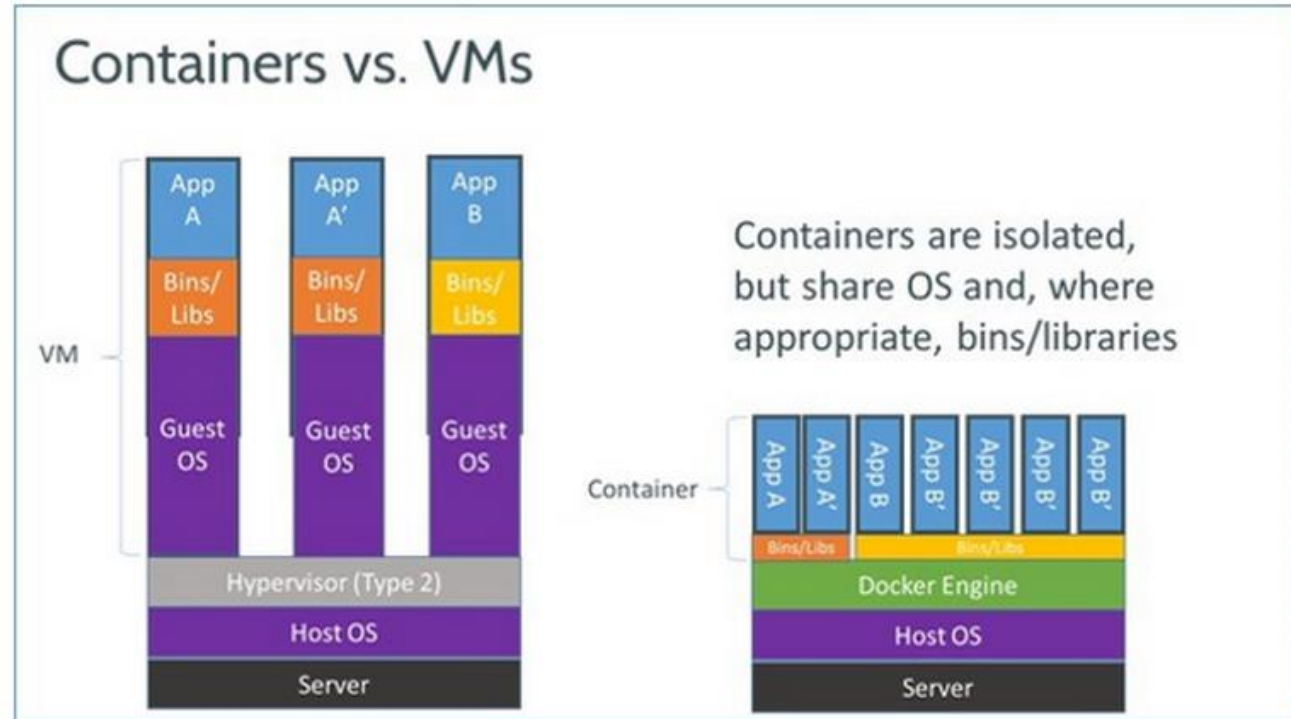
```
import nipype.interfaces.fsl as fsl
mybet = fsl.BET(in_file='T1.nii.gz',
                out_file='T1_brain.nii.gz')
result = mybet.run()
```

Example 1B | Nipype command for skull stripping.

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

DOCKER

- Containers are lighter and allow easier sharing of analysis pipelines
- Neurodocker and ReproZip help with building and deploying imaging pipelines
- But: Security on shared HPC systems is a problem



<http://www.zdnet.com/article/what-is-docker-and-why-is-it-so-darn-popular/>

<https://github.com/kaczmarj/neurodocker>

ROOTLESS CONTAINERS?



LINUX.CONF.AU
16-20 JANUARY 2017 HOBART
THE FUTURE OF OPEN SOURCE



Aleksa Sarai Rootless Containers with runC

<https://www.cyphar.com/blog/post/20160627-rootless-containers-with-runc>



National
Imaging
Facility



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

Create change

SINGULARITY

RESEARCH ARTICLE

Singularity: Scientific containers for mobility of compute

Gregory M. Kurtzer¹, Vanessa Sochat^{2*}, Michael W. Bauer^{1,3,4}

Interactive Development

```
sudo singularity build --sandbox tmpdir/ Singularity
```

```
sudo singularity build --writable container.img Singularity
```

BUILD ENVIRONMENT

Build from Recipe

```
sudo singularity build container.img Singularity
```

Build from Singularity

```
sudo singularity build container.img shub://vsoch/hello-world
```

Build from Docker

```
sudo singularity build container.img docker://ubuntu
```

Container Execution

```
singularity run container.img  
singularity shell container.img  
singularity exec container.img ...
```

Reproducible Sharing

```
singularity pull shub://...  
singularity pull docker://... *
```

PRODUCTION ENVIRONMENT

BIDS APPS

RESEARCH ARTICLE

BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

Krzysztof J. Gorgolewski^{1*}, Fidel Alfaro-Almagro², Tibor Auer³, Pierre Bellec^{4,5}, Mihai Capotă⁶, M. Mallar Chakravarty^{7,8}, Nathan W. Churchill⁹, Alexander Li Cohen¹⁰, R. Cameron Craddock^{11,12}, Gabriel A. Devenyi^{7,8}, Anders Eklund^{13,14,15}, Oscar Esteban¹, Guillaume Flandin¹⁶, Satrajit S. Ghosh^{17,18}, J. Swaroop Guntupalli¹⁹, Mark Jenkinson², Anisha Keshavan²⁰, Gregory Kiar^{21,22}, Franziskus Liem²³, Pradeep Reddy Raamana^{24,25}, David Raffelt²⁶, Christopher J. Steele^{7,8}, Pierre-Olivier Quirion¹⁵, Robert E. Smith²⁶, Stephen C. Strother^{24,25}, Gaël Varoquaux²⁷, Yida Wang⁶, Tal Yarkoni²⁸, Russell A. Poldrack¹

BRAIN IMAGING DATA STRUCTURE

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/



National
Imaging
Facility

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



Create change

BIDS APPS

RESEARCH ARTICLE

BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

Krzysztof J. Gorgolewski^{1*}, Fidel Alfaro-Almagro², Tibor Auer³, Pierre Bellec^{4,5}, Mihai Capotă⁶, M. Mallar Chakravarty^{7,8}, Nathan W. Churchill⁹, Alexander Li Cohen¹⁰, R. Cameron Craddock^{11,12}, Gabriel A. Devenyi^{7,8}, Anders Eklund^{13,14,15}, Oscar Esteban¹, Guillaume Flandin¹⁶, Satrajit S. Ghosh^{17,18}, J. Swaroop Guntupalli¹⁹, Mark Jenkinson², Anisha Keshavan²⁰, Gregory Kiar^{21,22}, Franziskus Liem²³, Pradeep Reddy Raamana^{24,25}, David Raffelt²⁶, Christopher J. Steele^{7,8}, Pierre-Olivier Quirion¹⁵, Robert E. Smith²⁶, Stephen C. Strother^{24,25}, Gaël Varoquaux²⁷, Yida Wang⁶, Tal Yarkoni²⁸, Russell A. Poldrack¹

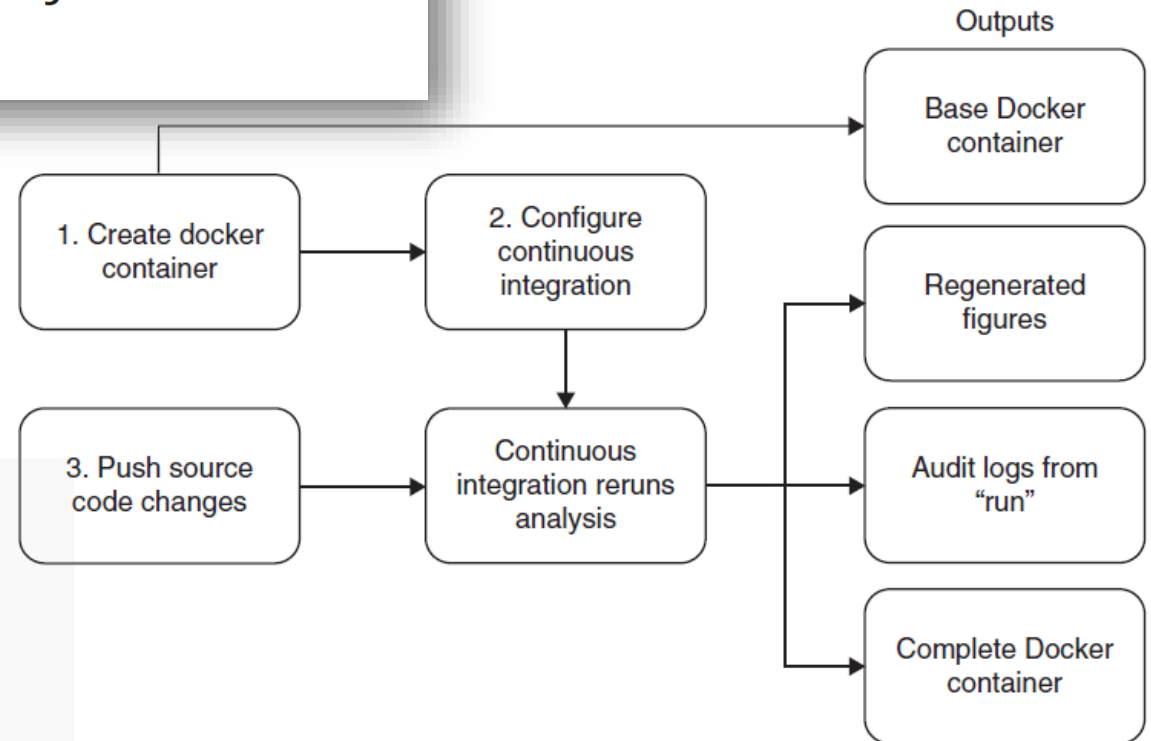
- neuroimaging tools and workflows in docker containers
- e.g. FreeSurfer, SPM, MRtrix3, AFNI, ANTS, HCP Pipelines

CONTINUOUS ANALYSIS?

Reproducibility of computational workflows is automated using continuous analysis

Brett K Beaulieu-Jones¹ & Casey S Greene²

- continuous integration and delivery is great for software development
- can we apply the same systems to a data-analysis pipeline?



https://github.com/greenelab/continuous_analysis

CONTINUOUS ANALYSIS?

- No manual steps
- version control
- utilizing established projects – not re-inventing the wheel
- testing of code – e.g. assertions, smoke tests, (unit testing)

```
ages = [32, 31, 23, 24, 22, 19, 99]
mean_age = np.mean(ages)
print(mean_age)
```

```
35.5714285714
```

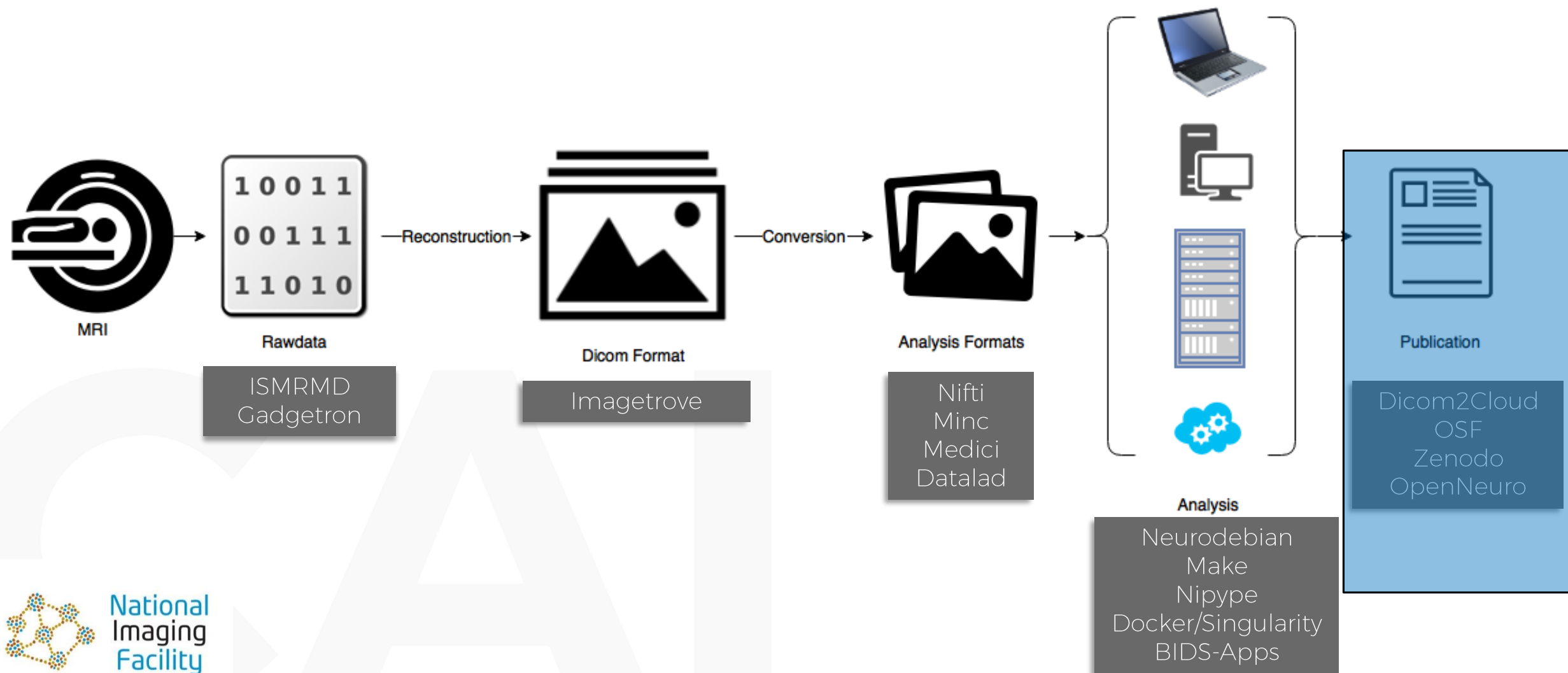
```
ages = [32, 31, 23, 24, 22, 19, 99]
for age in ages:
    assert age>18 and age<65
mean_age = np.mean(ages)
print(mean_age)
```

```
Traceback (most recent call last):
  File "age.py", line 6, in <module>
    assert age>18 and age<65
AssertionError
```

```
ages = [32, 31, 23, 24, 22, 19, 29]
for age in ages:
    assert age>18 and age<65
mean_age = np.mean(ages)
print(mean_age)
```

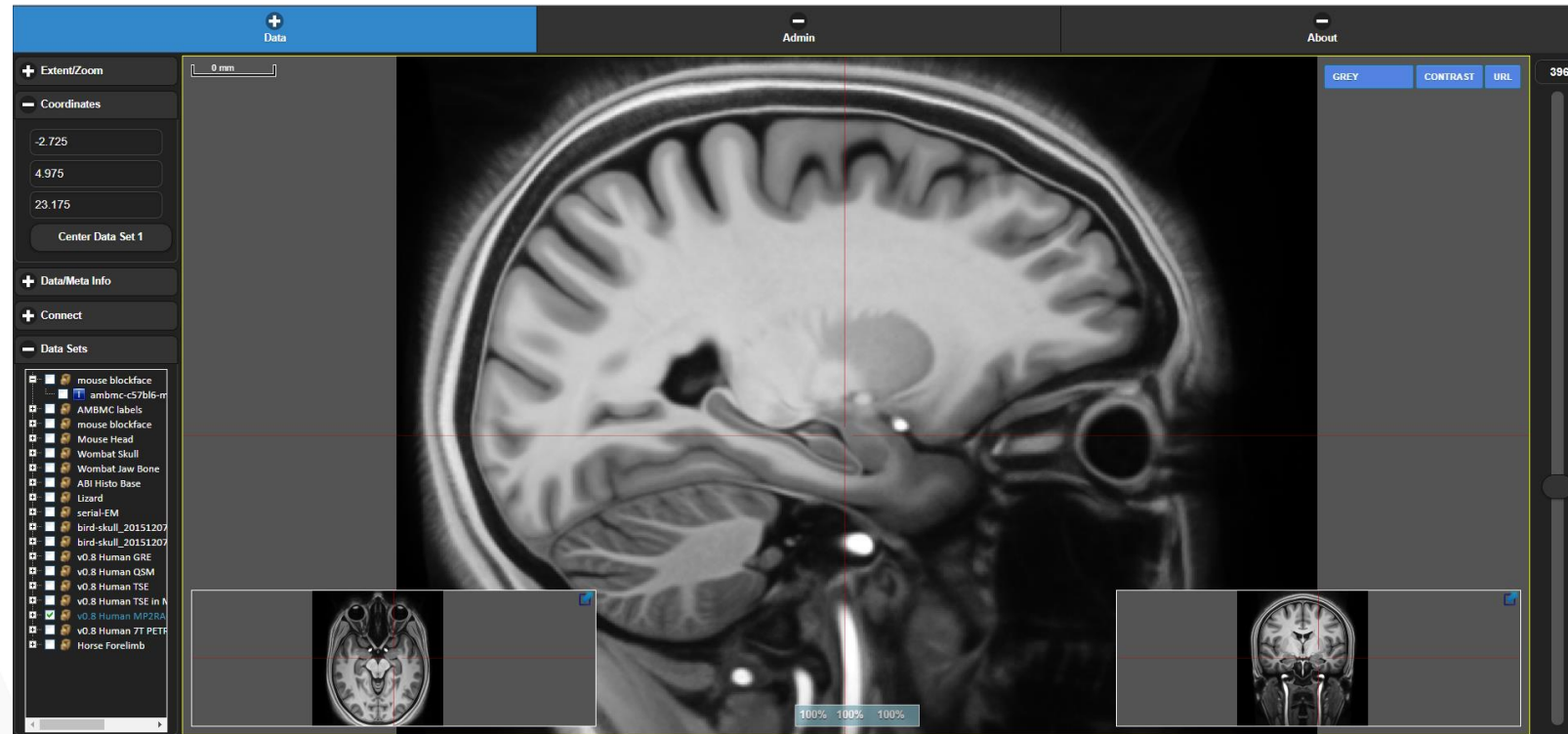
```
25.7142857143
```

A LONG WAY TO PUBLICATION



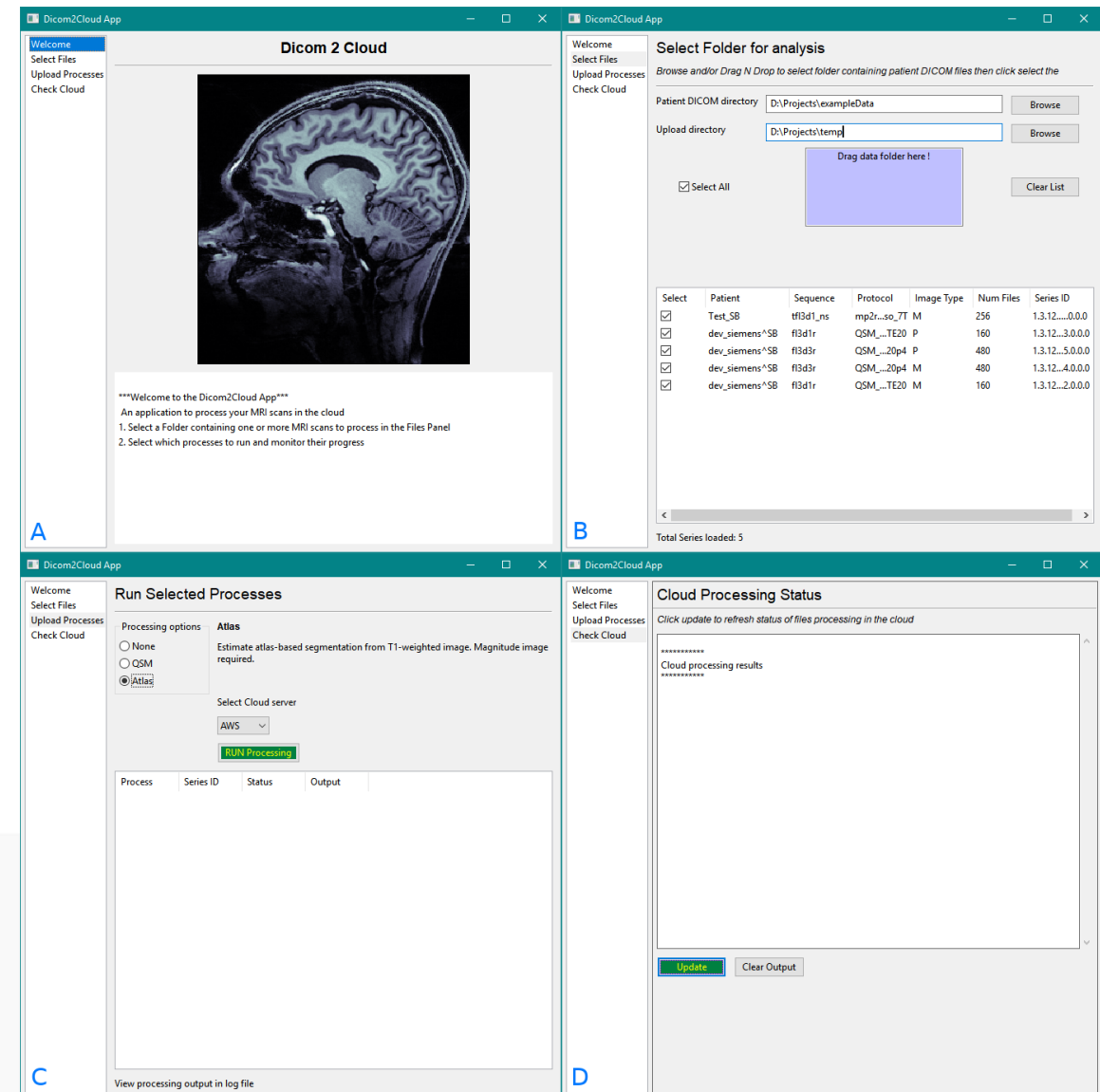
TISSUESTACK

- an image viewer in the browser
- can display multi-terabyte images



DICOM2CLOUD

- how can we distribute the pipelines we developed to a clinical partner?
- challenges: de-identification, simplicity of use, platform-independence
- prototype during health-hack Brisbane 2017



MYBINDER.ORG



- how can we share and run a full analysis for a publication on the web?

Build and launch a repository

GitHub repo or URL

GitHub repository name or link

Git branch, tag, or commit

master

Path to a notebook file (optional)

Path to a notebook file (optional) File

launch

Copy the URL below and share your Binder with others:

Fill in the fields to see a URL for sharing your Binder.

Copy the text below, then paste into your README to show a binder badge: launch binder

OPENNEURO.ORG

- share datasets in BIDS format
- run analyses for free
 - MRIQC
 - Freesurfer
 - fMRIPrep
 - NDMG



OpenNEURO^{BETA}

A free and open platform for analyzing
and sharing neuroimaging data



Sign in with Google



Sign in with ORCID



National
Imaging
Facility

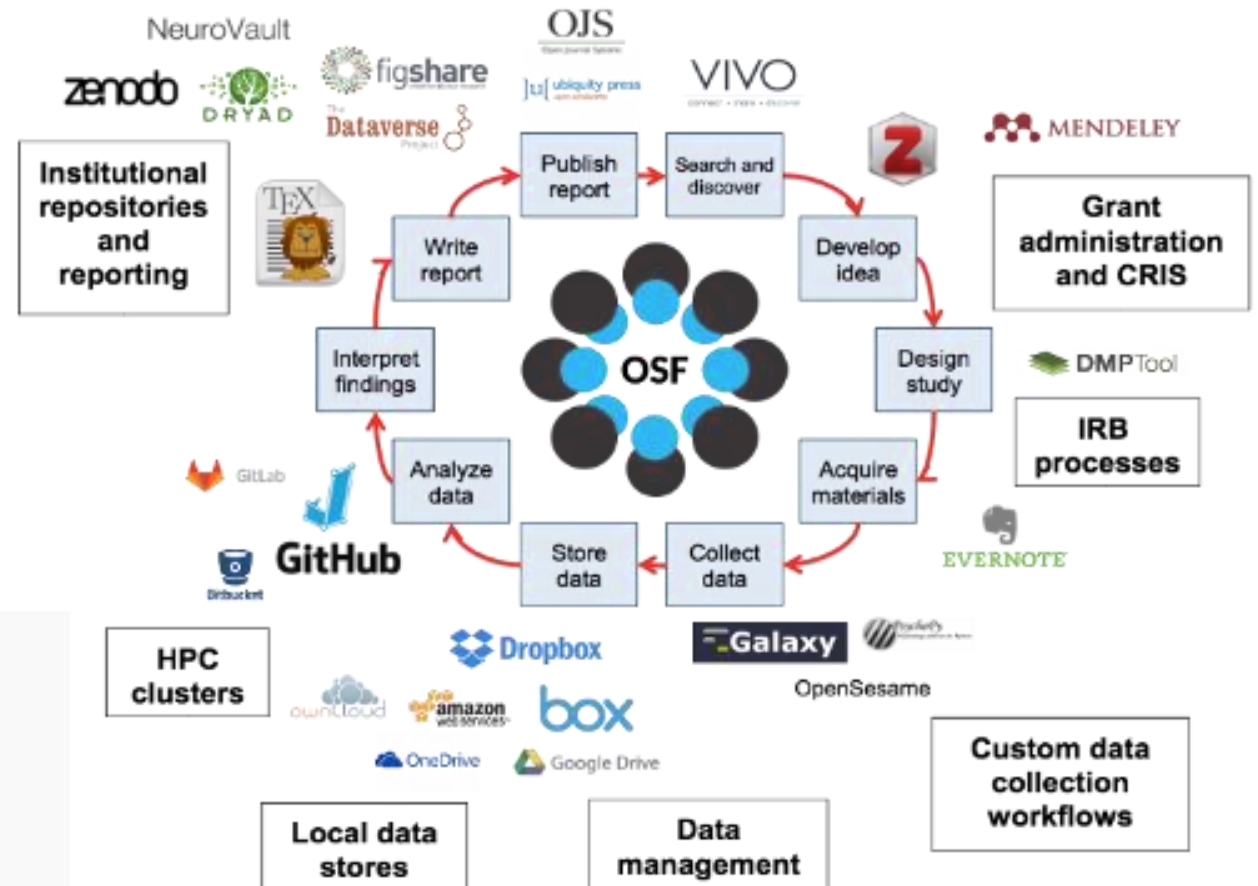


THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

Create change

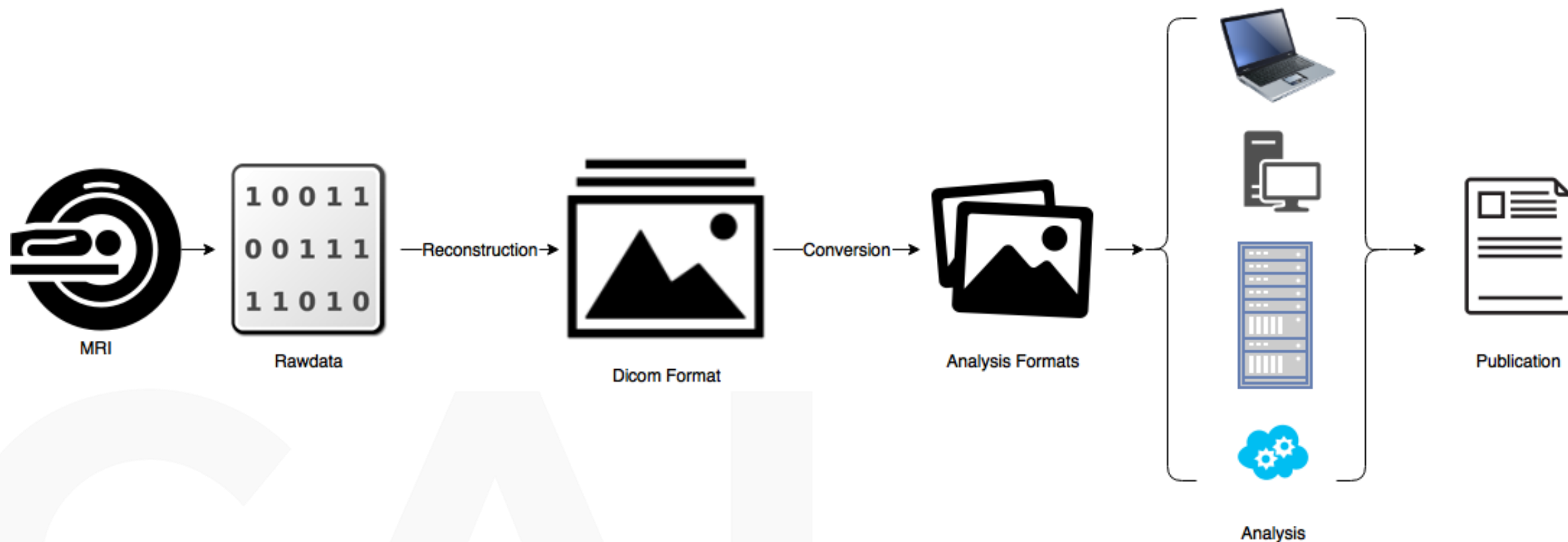
OPEN SCIENCE FRAMEWORK

- document and archive study designs, materials, and data
- preservation fund for 50 years read access to data
- a command line client:
 - <https://github.com/osfclient/osfclient>



<http://slidedeck.io/VickySteeves/OSF-Workshop>

A LONG WAY TO PUBLICATION



SUMMARY

- Many open-source tools exist to support the scientific process
- scientists are not software developers (and shouldn't be): But we use and write tools to implement our ideas
- how can we make more tools easier to access?
- how can we build seamless data provenance into file systems?
- how can we run software effortlessly across multiple platforms?
- how can we make it more desirable to contribute to existing projects instead of re-inventing?

Thank you

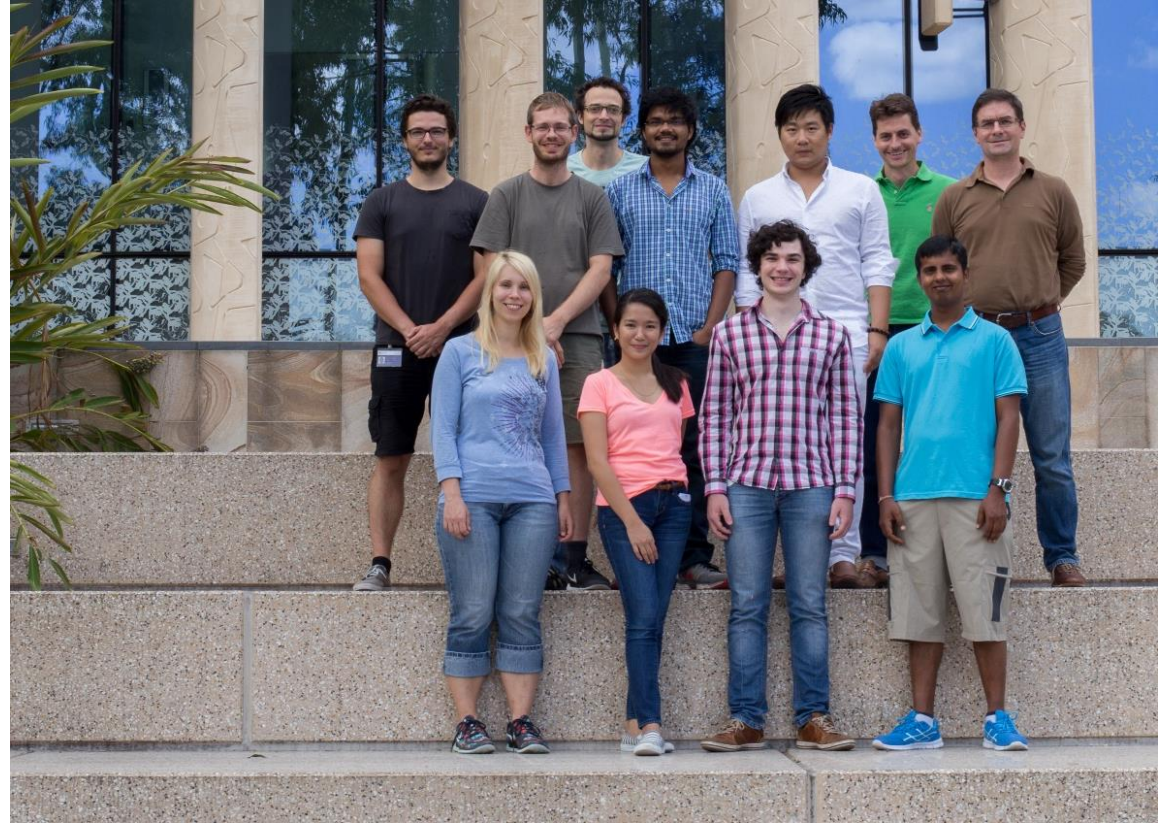


Centre for Advanced Imaging

www.cai.uq.edu.au/bollmann



steffen.bollmann@cai.uq.edu.au



ABOUT THE CENTRE

The Centre for Advanced Imaging (CAI) is a strategic initiative of The University of Queensland and reflects the growth in biotechnology and biomedical research requiring spectroscopic and imaging research capabilities.

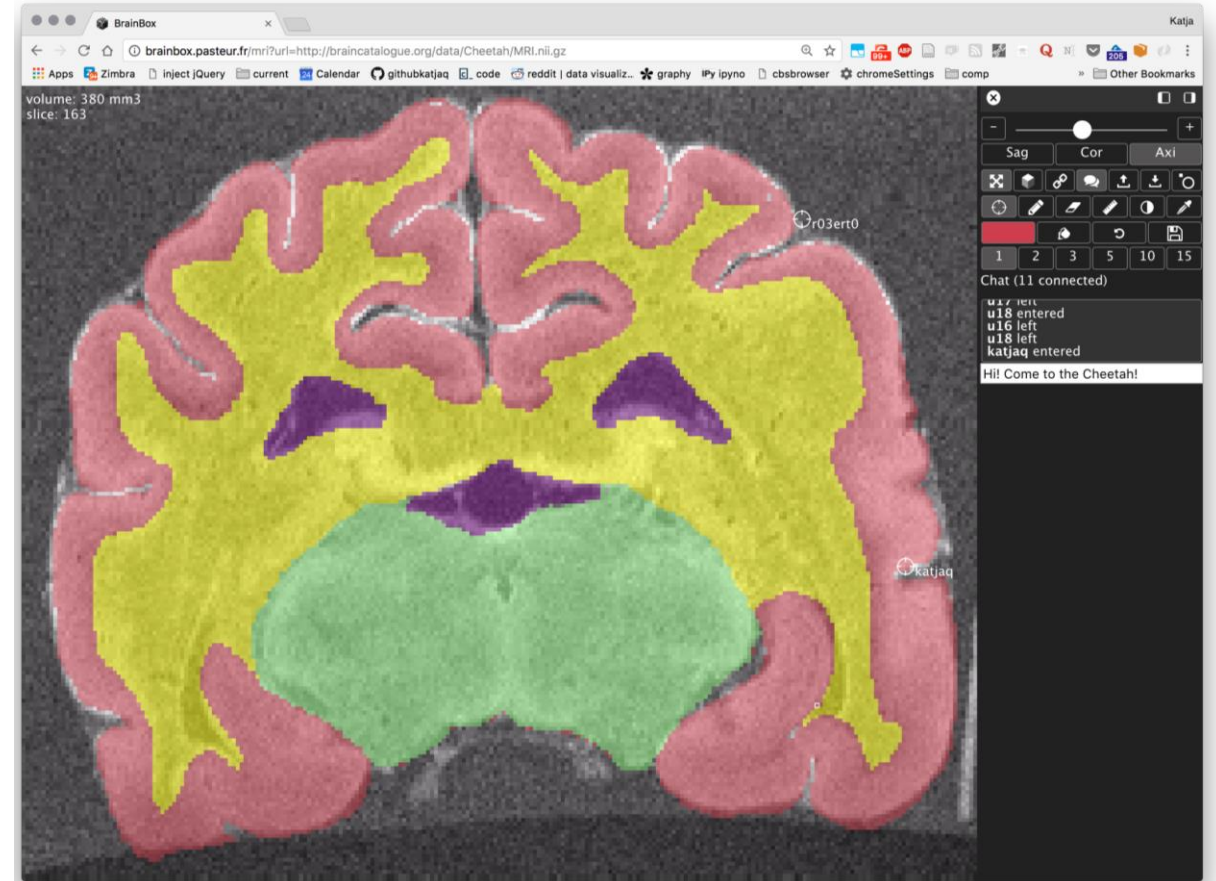
The CAI is an integrated research facility harnessing the spatial resolution of ultra-high-field magnetic resonance imaging (MRI), and the sensitivity of positron emission tomography (PET) in detecting molecular targets and the geometrical and electronic structural characterisation of molecules, using high-resolution nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) spectroscopy.

CAI researchers work on innovations in spectroscopic and imaging technology, imaging biomarker development and in biomedical research disciplines. They often work in collaboration with clinical research sites and other local, national, and international research institutes and companies.



BRAINBOX

- web application for the collaborative curation of neuroimaging data
- share, visualise and annotate MRI brain data



ZENODO.ORG



- How can I get a digital object identifier (DOI)?

1 Flip the switch

Select the repository you want to preserve, and toggle the switch below to turn on automatic preservation of your software.



2 Create a release

Go to GitHub and [create a release](#). Zenodo will automatically download a .zip-ball of each new release and register a DOI.

3 Get the badge

After your first release, a DOI badge that you can include in GitHub README will appear next to your repository below.

