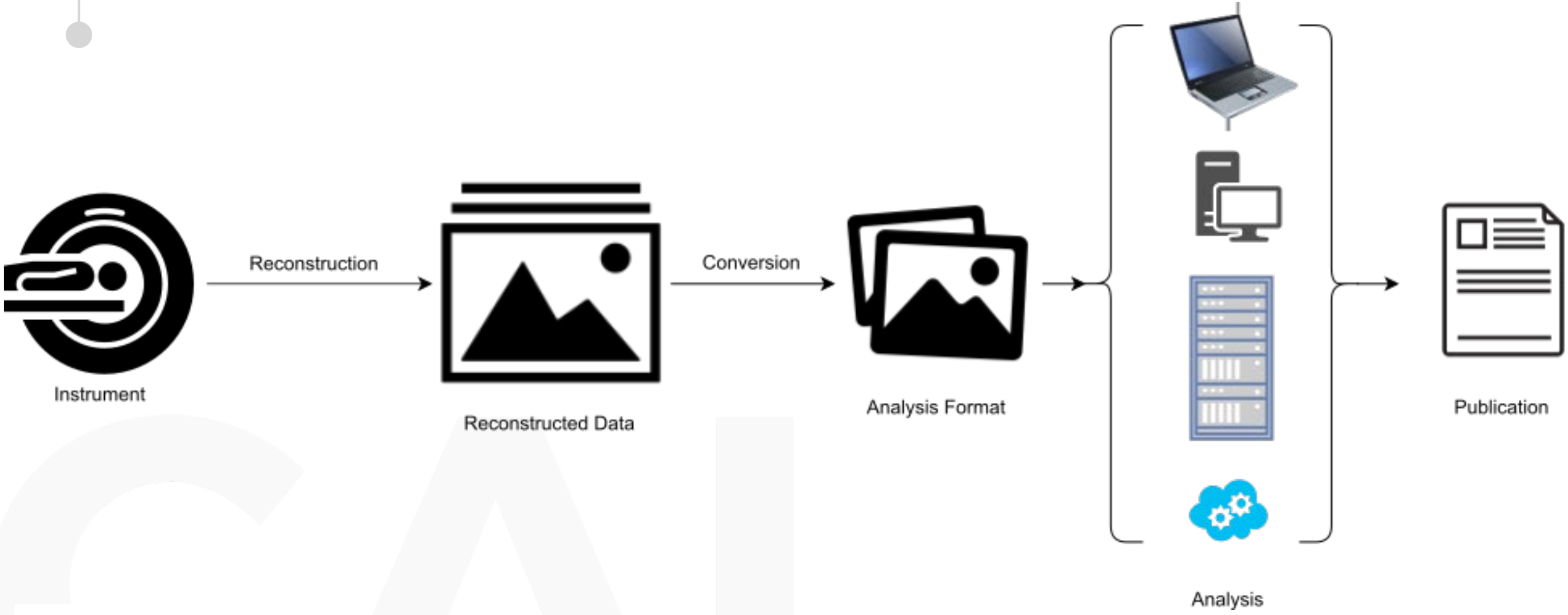
HIGH PERFORMANCE SCIENCE

Steffen Bollmann

open and reproducible neuroscience image processing that scales?

FROM DATA TO PUBLICATION



National Imaging Facility

@sbollmann_MRI

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

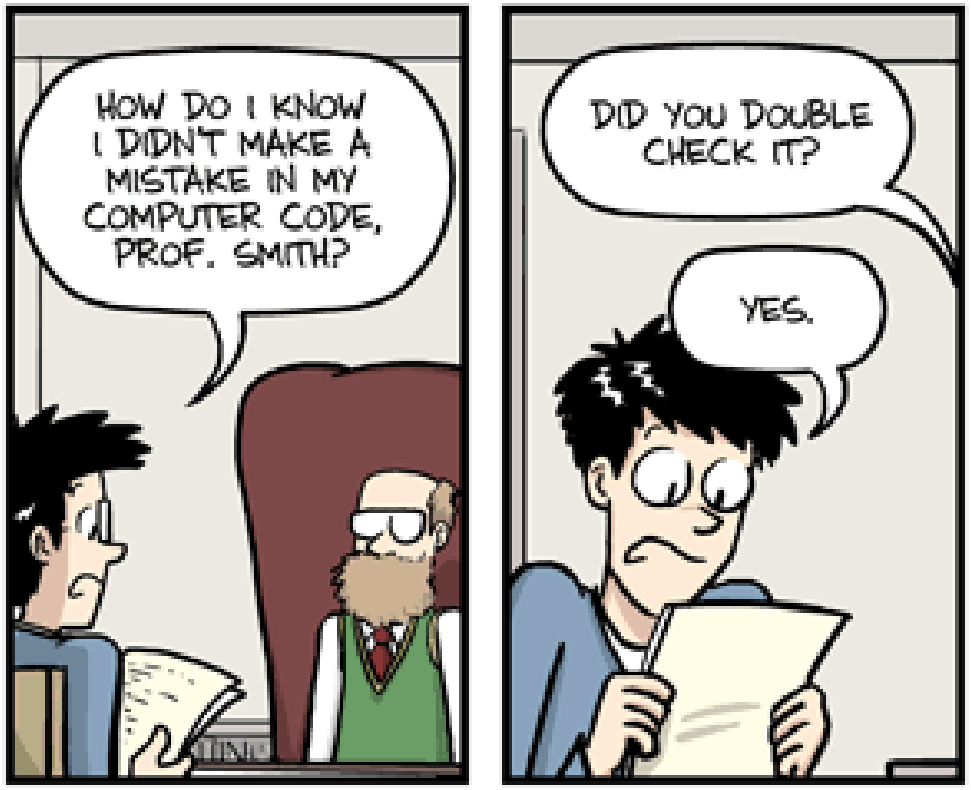
@sbollmann_MRI

Steffen.Bollmann@cai.uq.edu.au

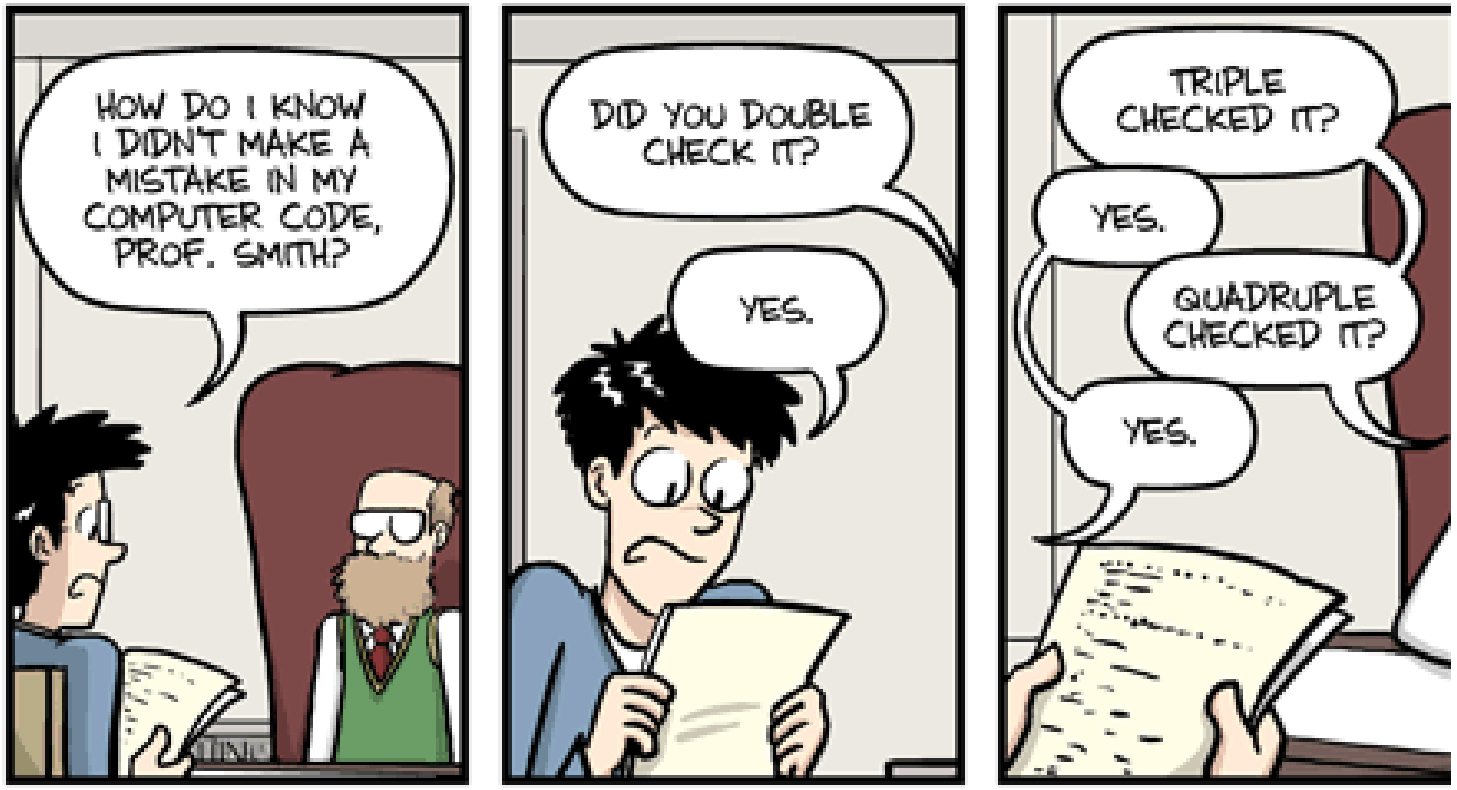
The University Of Queensland AUSTRALIA

Create change

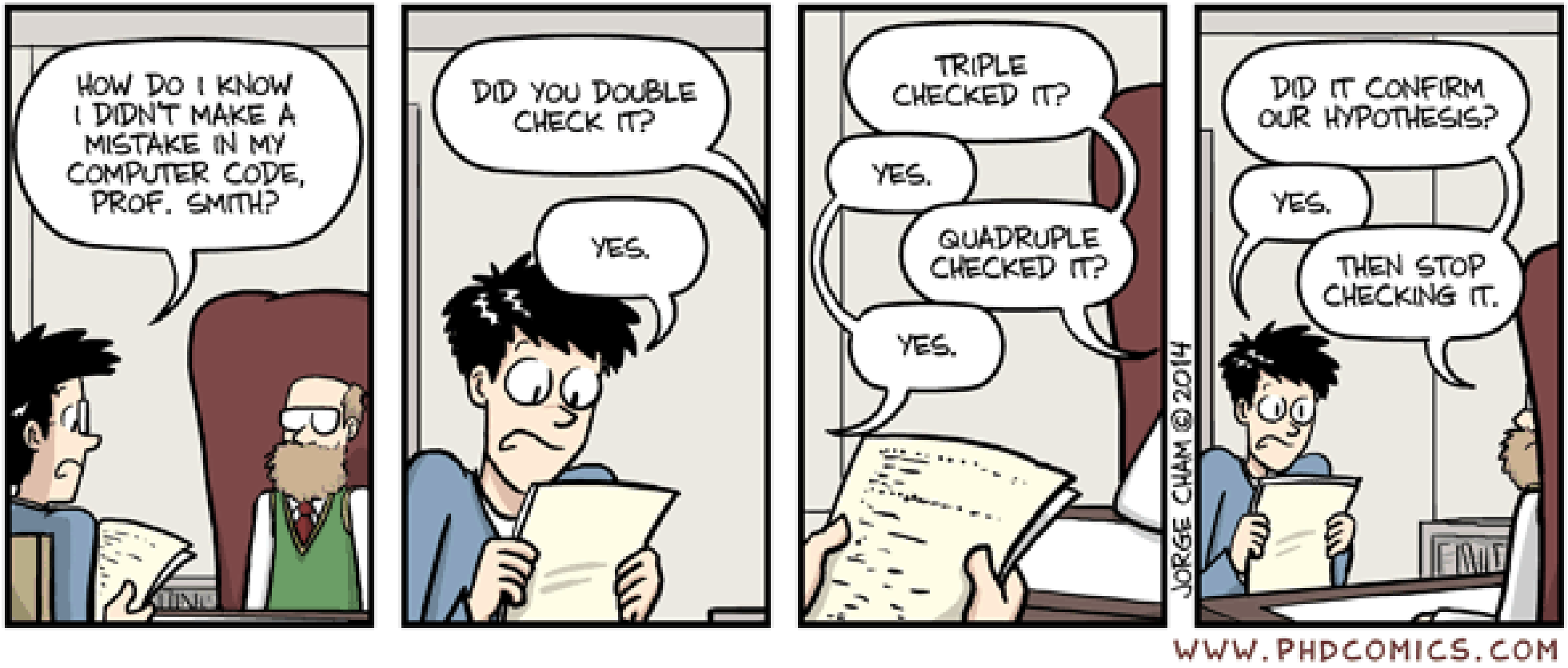
SELECTIVE DEBUGGING



SELECTIVE DEBUGGING



SELECTIVE DEBUGGING



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

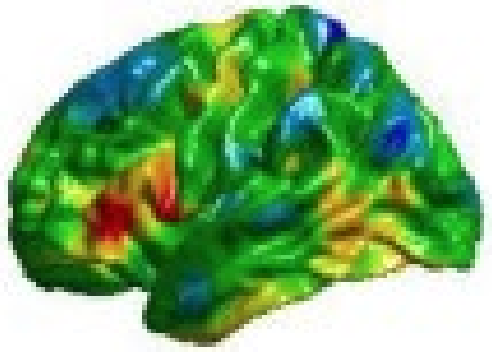
GLIBC 2.5 VS 2.18

Reproducibility of neuroimaging analyses across operating systems

Tristan Glatard[1,2], Lindsay B. Lewis[1], Rafael Ferreira da Silva[3], Reza Adalat[1], Natacha Beck[1], Claude Lepage[1], Pierre Rioux[1], Marc-Etienne Rousseau[1], Tarek Sherif[1], Ewa Deelman[3], Najmeh Khalili-Mahani[1] and Alan C. Evans[1]

```
expf(1.54051852226257324218750000000)
      =4.667009353637695312500000

expf(1.54051852226257324218750000000)
      =4.667009830474853515625000
```



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

National Imaging Facility

@sbollmann_MRI

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*

National Imaging Facility

@sbollmann_MRI

The University Of Queensland AUSTRALIA

Create change

OUR JOURNEY BEGINS

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

National Imaging Facility

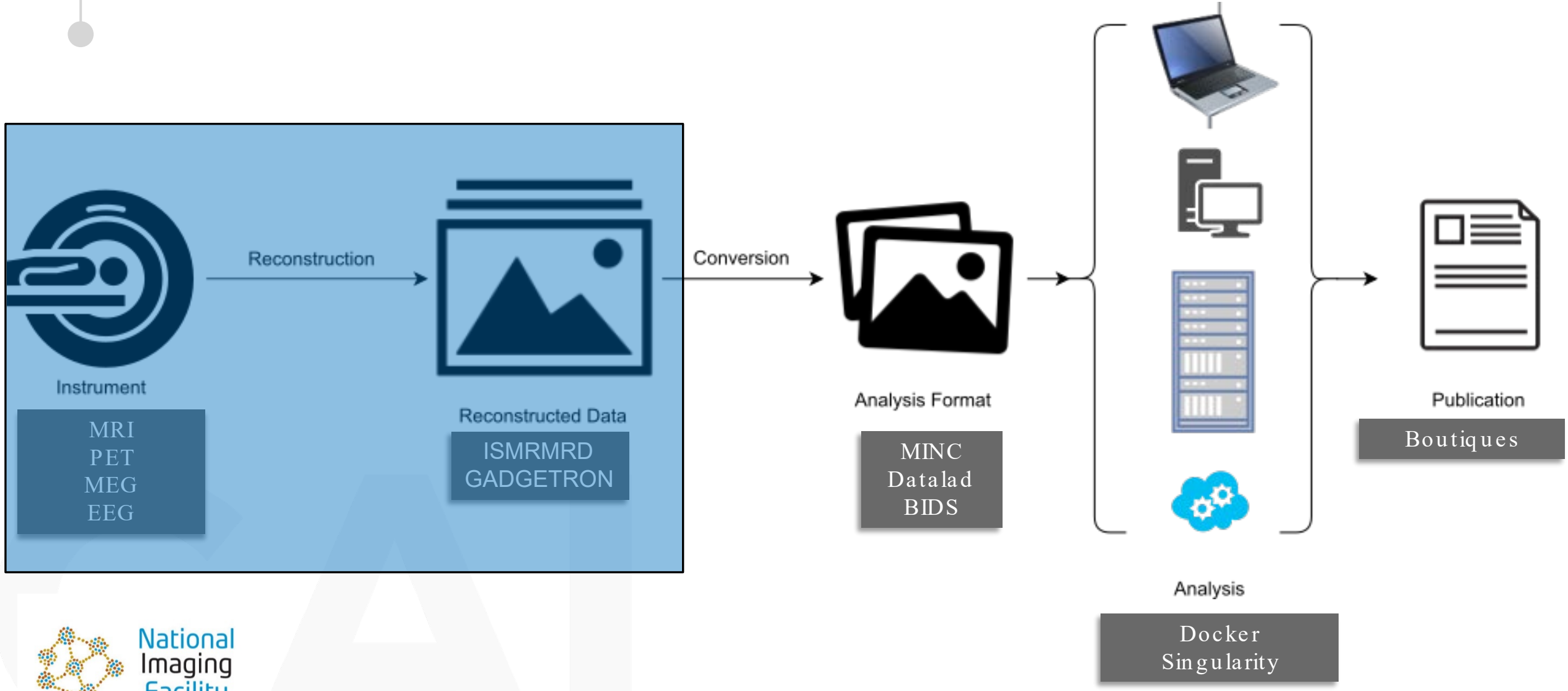
@sbollmann_MRI

The University Of Queensland AUSTRALIA

Create change

Steffen.Bollmann@cai.uq.edu.au

FROM DATA TO PUBLICATION



National Imaging Facility

@sbollmann_MRI

ISMRMRD & GADGETRON

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]*

Gadgetron: An Open Source Framework for Medical Image Reconstruction

Michael Schacht Hansen[1]* and Thomas Sangild Sørensen[2,3]

ISMRMRD

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

https://github.com/ismrmrd

Gadgetron

- Gadgetron : Open source reconstruction system compatible with ISMRMRD

https://github.com/gadgetron

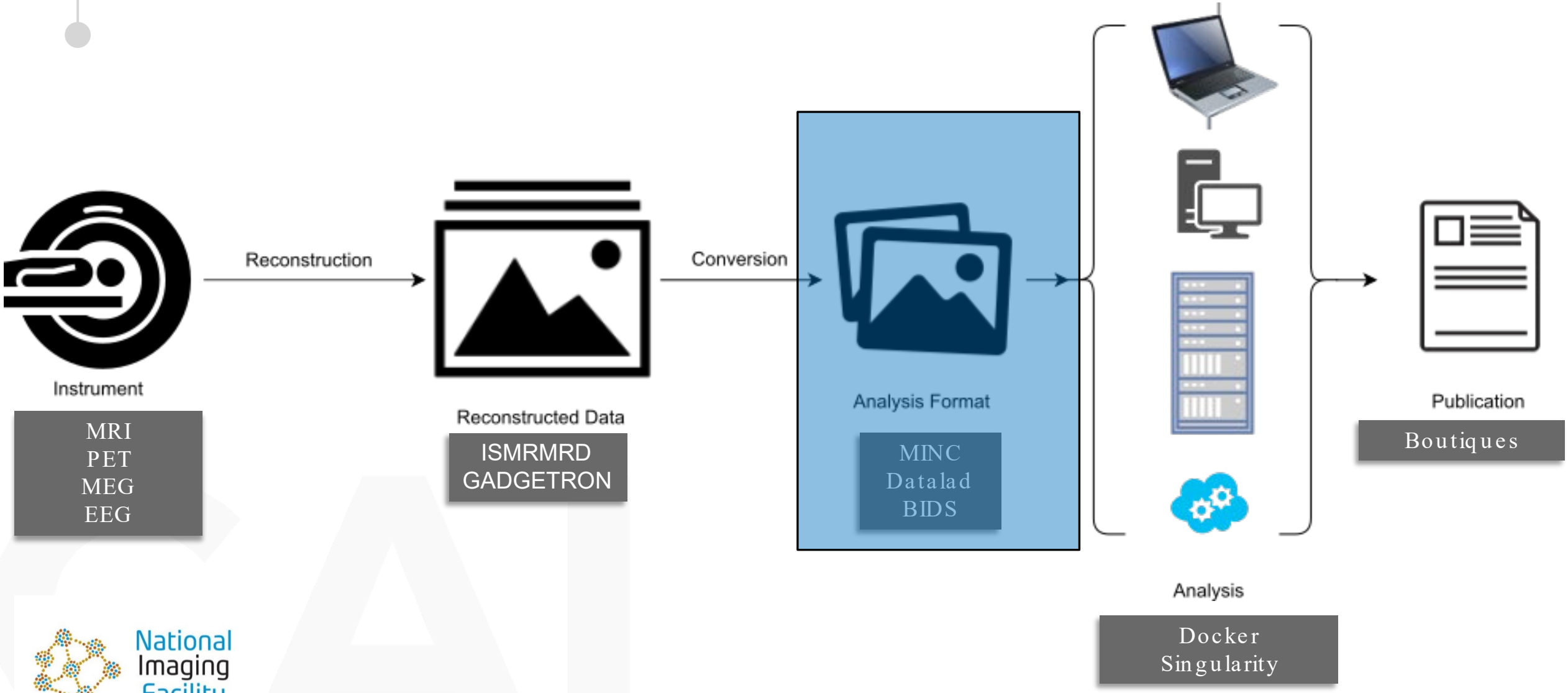
National Imaging Facility

@sbollmann_MRI

The University Of Queensland AUSTRALIA

Create change

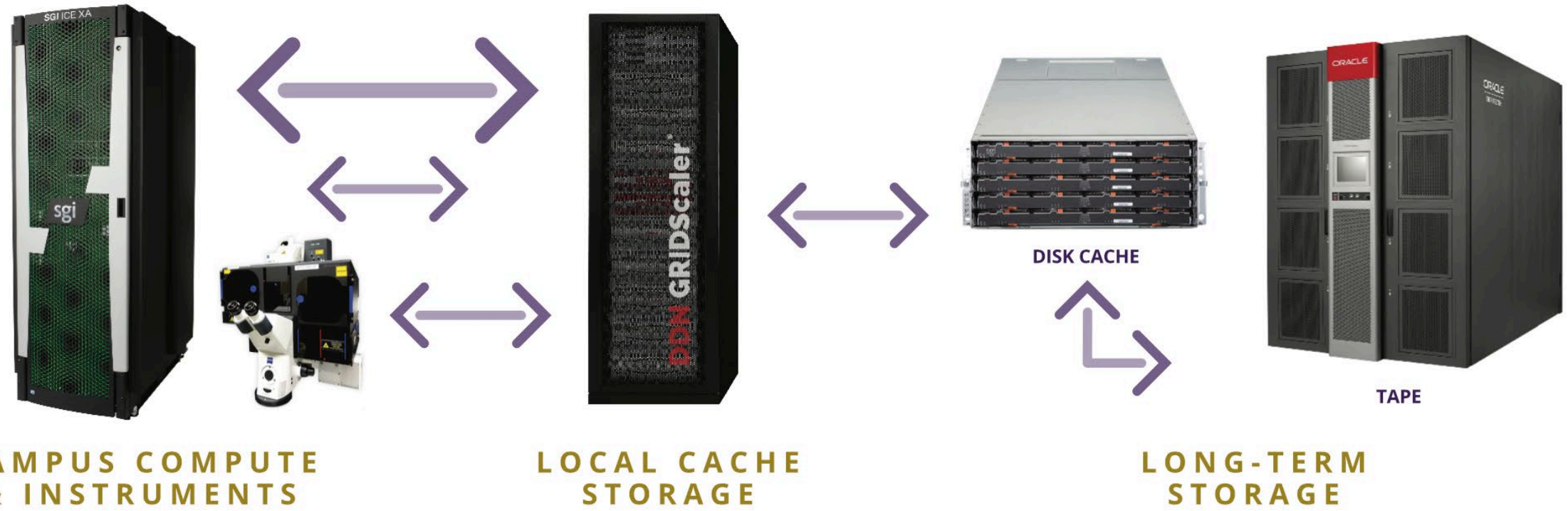
FROM DATA TO PUBLICATION



National Imaging Facility

@sbollmann_MRI

METROPOLITAN DATA CACHING INFRASTRUCTURE (MEDICI)



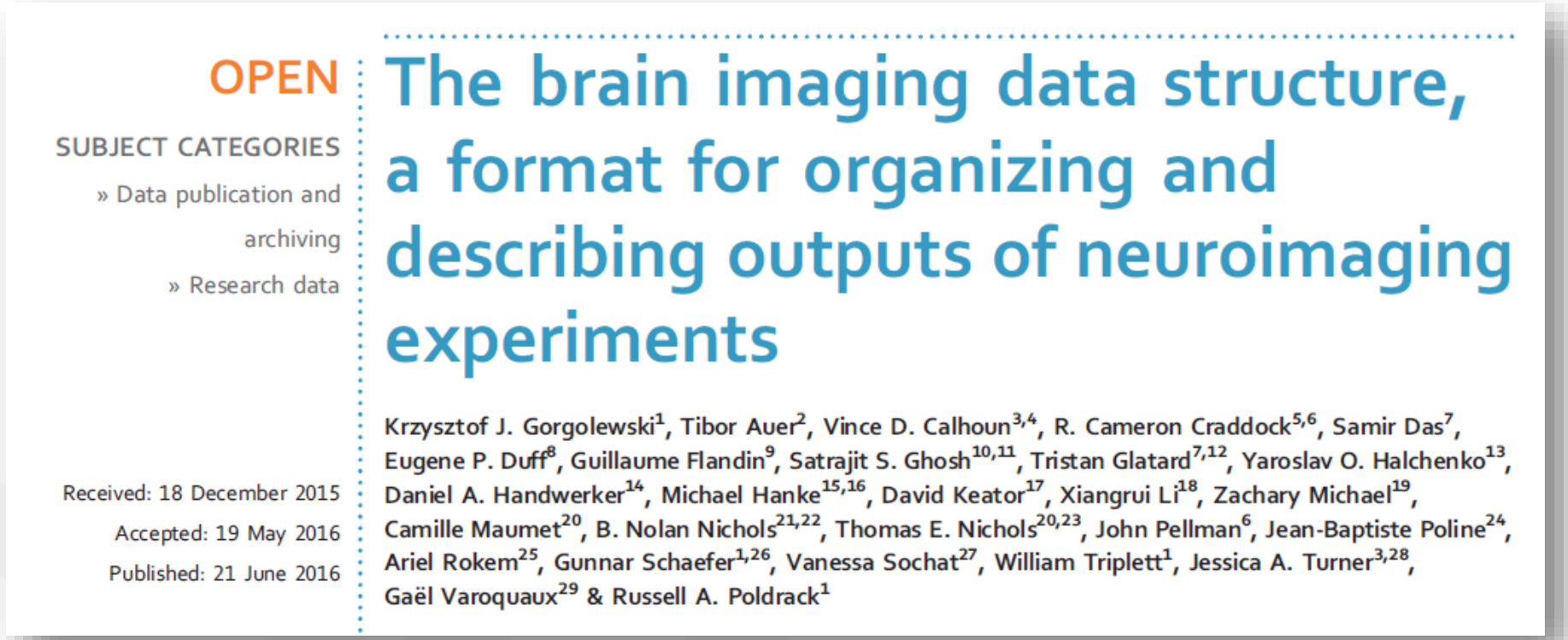
National Imaging Facility

https://rcc.uq.edu.au/data -storage

The University Of Queensland AUSTRALIA

Create change

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

Krzysztof J. Gorgolewski[1], Tibor Auer[2], Vince D. Calhoun[3,4], R. Cameron Craddock[5,6], Samir Das[7], Eugene P. Duff[8], Guillaume Flandin[9], Satrajit S. Ghosh[10,11], Tristan Glatard[7,12], Yaroslav O. Halchenko[13], Daniel A. Handwerker[14], Michael Hanke[15,16], David Keator[17], Xiangrui Li[18], Zachary Michael[19], Camille Maumet[20], B. Nolan Nichols[21,22], Thomas E. Nichols[20,23], John Pellman[6], Jean-Baptiste Poline[24], Ariel Rokem[25], Gunnar Schaefer[1,26], Vanessa Sochat[27], William Triplett[1], Jessica A. Turner[3,28], Gaël Varoquaux[29] & Russell A. Poldrack[1]

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

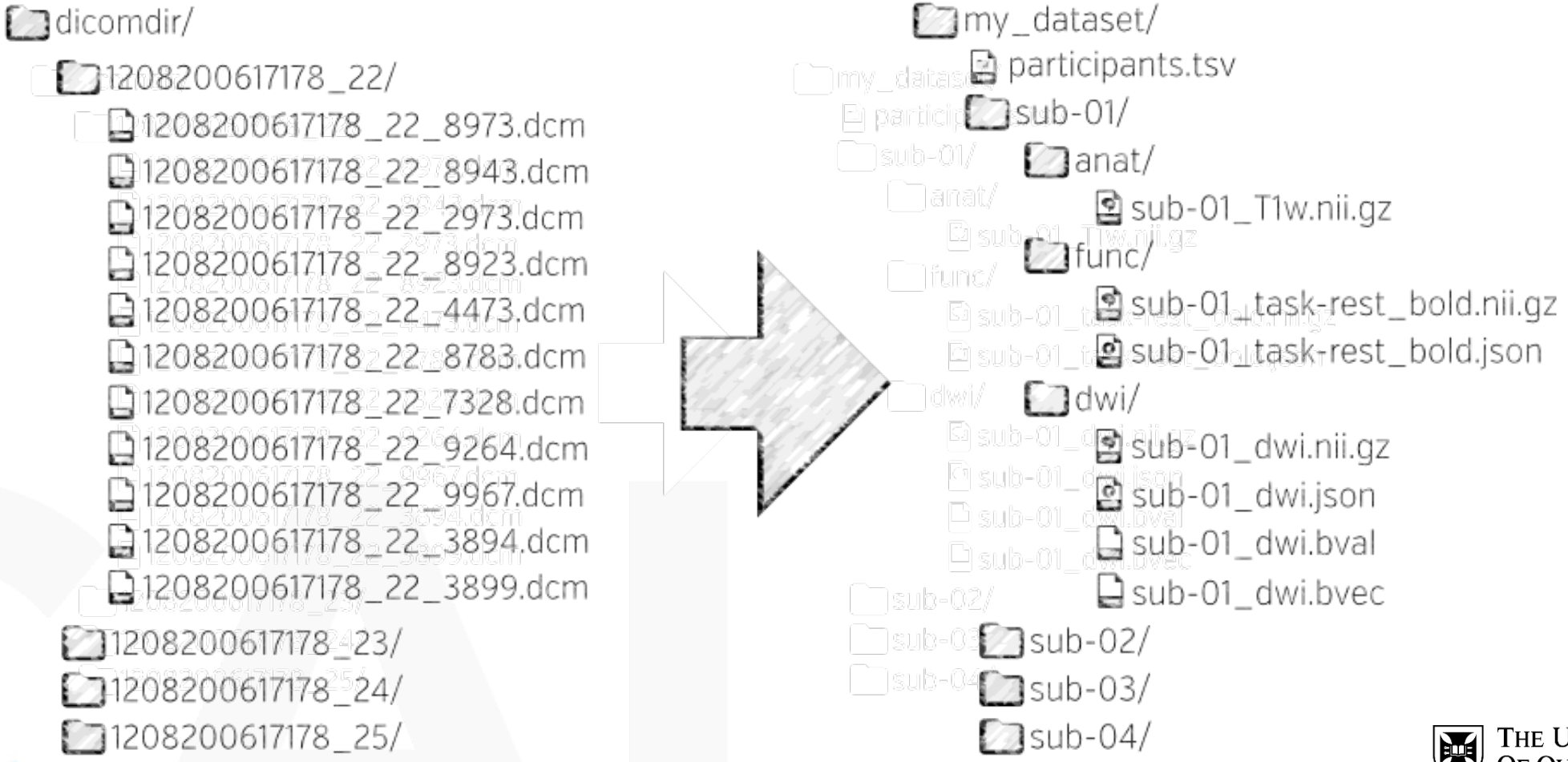
National Imaging Facility

@sbollmann_MRI

The University Of Queensland AUSTRALIA

Create change

BRAIN IMAGING DATA STRUCTURE



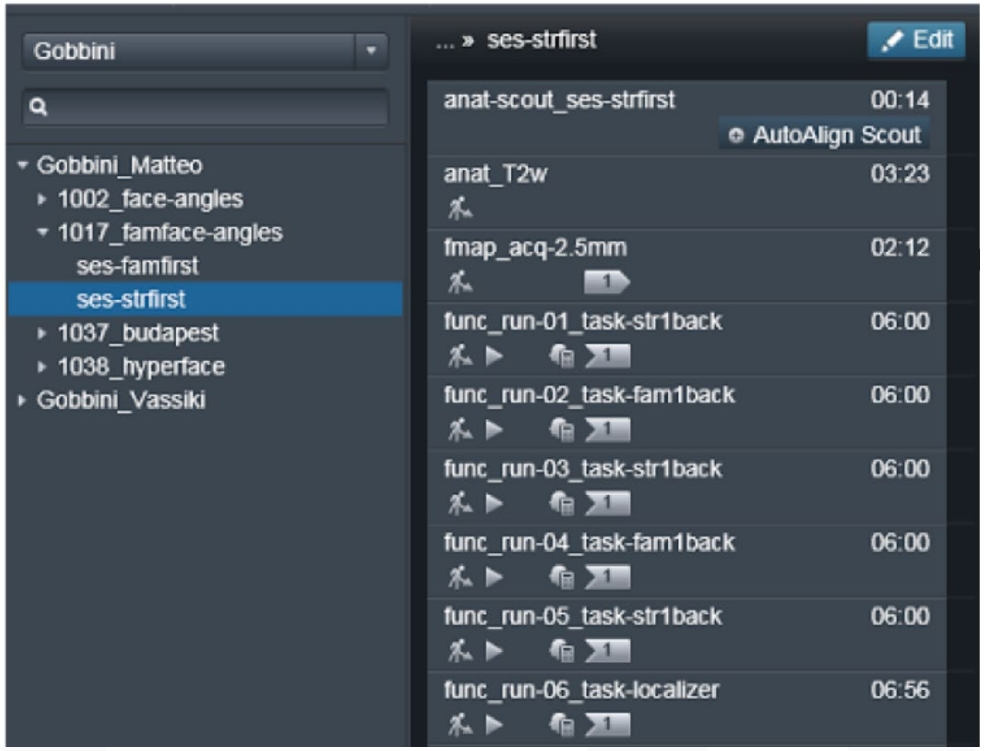
National Imaging Facility

@sbollmann_MRI

http://bids.neuroimaging.io/

REPROIN

- setting up protocols on scanner to enable automatic conversion to bids

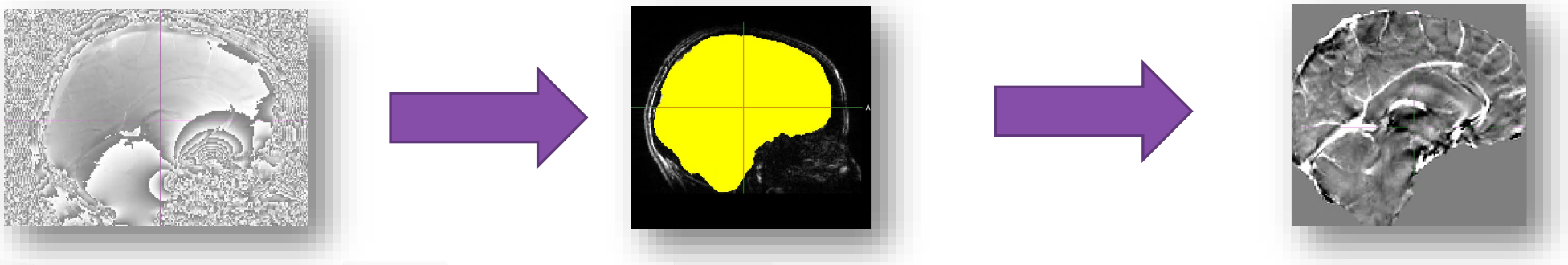



```
anat
  sub-sid000005_ses-strfirst_T2w.json
  sub-sid000005_ses-strfirst_T2w.nii.gz
fmap
  sub-sid000005_ses-strfirst_acq-25mm_magnitude1.json
  sub-sid000005_ses-strfirst_acq-25mm_magnitude1.nii.gz
  ...
func
  sub-sid000005_ses-strfirst_task-fam1back_run-02_bold.json
  sub-sid000005_ses-strfirst_task-fam1back_run-02_bold.nii.gz
  sub-sid000005_ses-strfirst_task-fam1back_run-02_events.tsv
  ...
sub-sid000005_ses-strfirst_scans.tsv
```

https://github.com/ReproNim/reproin

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
```

National Imaging Facility

@sbollmann_MRI

The University Of Queensland AUSTRALIA

Create change

MINC: BUILD -IN PROVENANCE

MINC 2.0: A Flexible Format for Multi-Modal Images

Robert D. Vincent[1], Peter Neelin[2], Najmeh Khalili-Mahani[1], Andrew L. Janke[3], Vladimir S. Fonov[1], Steven M. Robbins[1], Leila Baghdadi[4], Jason Lerch[4,5], John G. Sled[4,5], Reza Adalat[1], David MacDonald[6], Alex P. Zijdenbos[7], 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 \
```

https://bic -mni.github.io/

National Imaging Facility

@sbollmann_MRI

Steffen.Bollmann@cai.uq.edu.au

The University Of Queensland AUSTRALIA

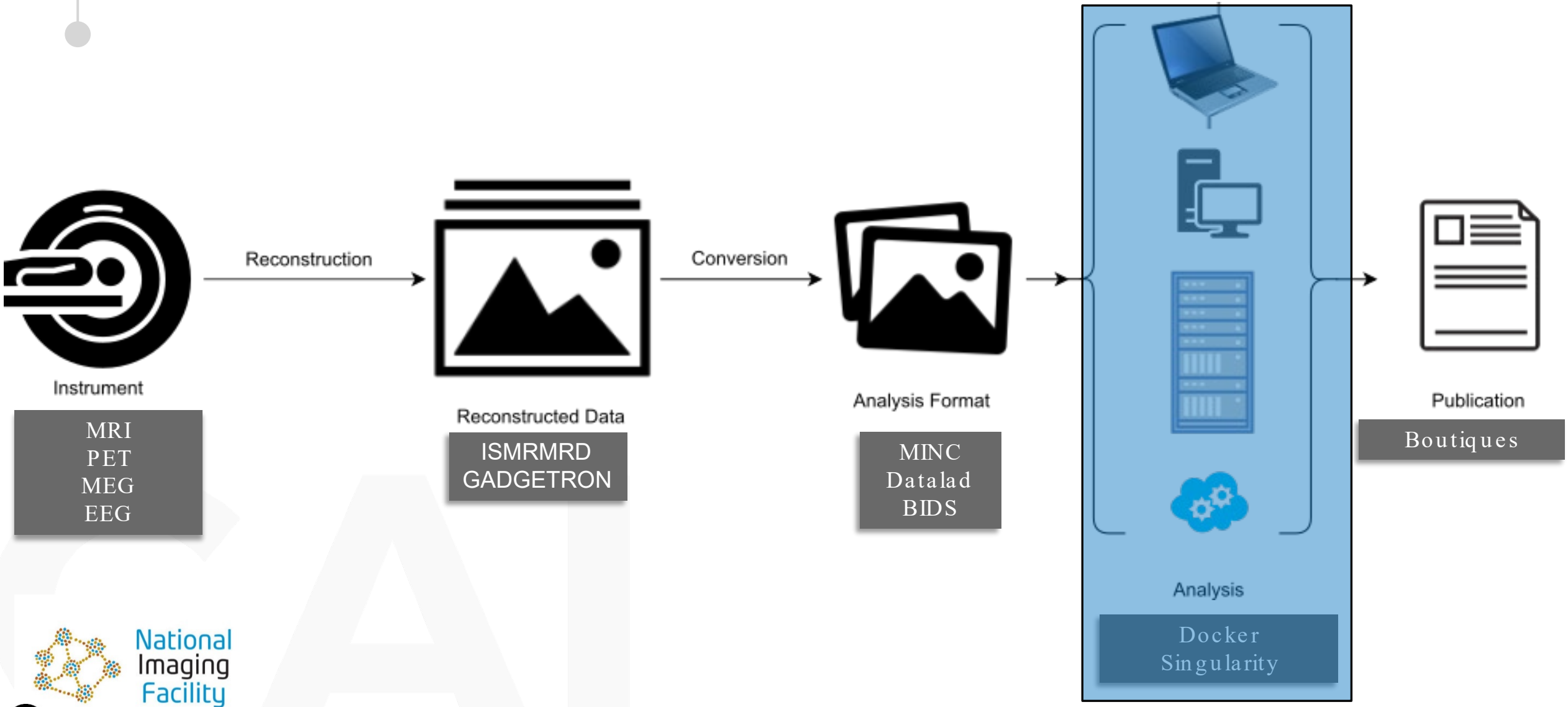
Create change

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

FROM DATA TO PUBLICATION

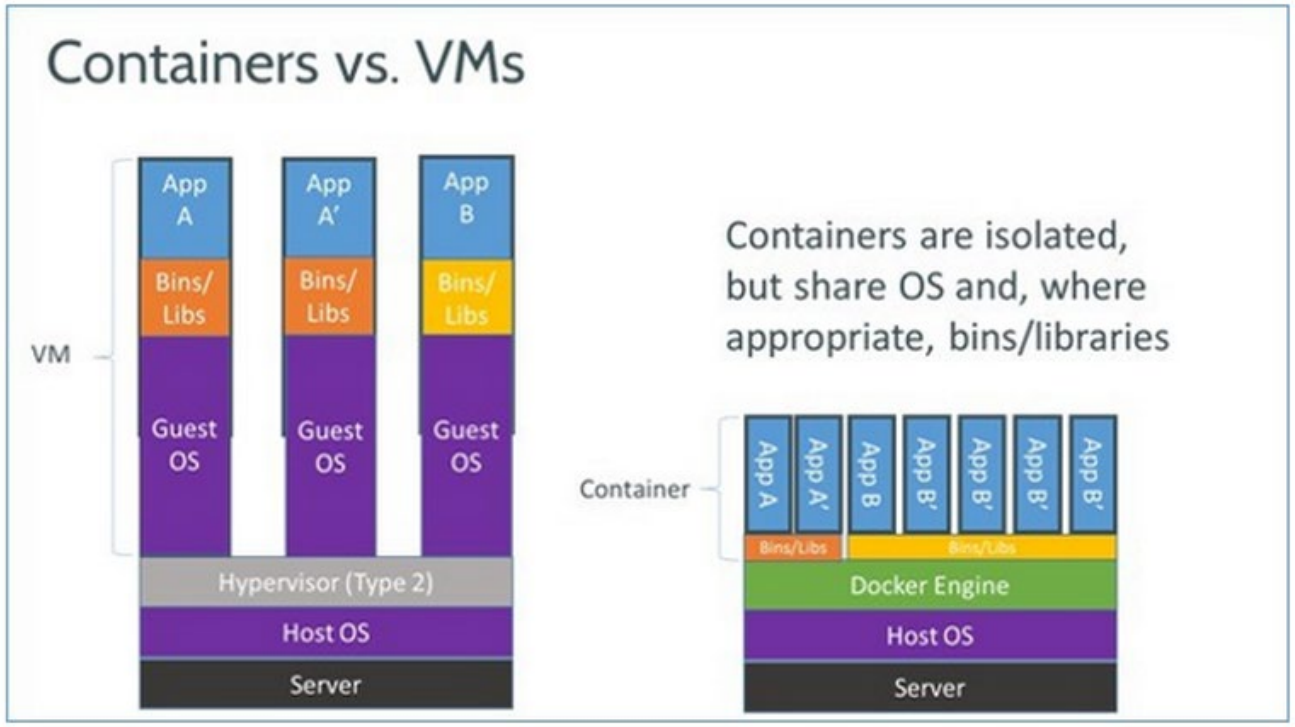


National Imaging Facility

@sbollmann_MRI

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 so-darn -popular/ -is-docker -and -why -is-it-

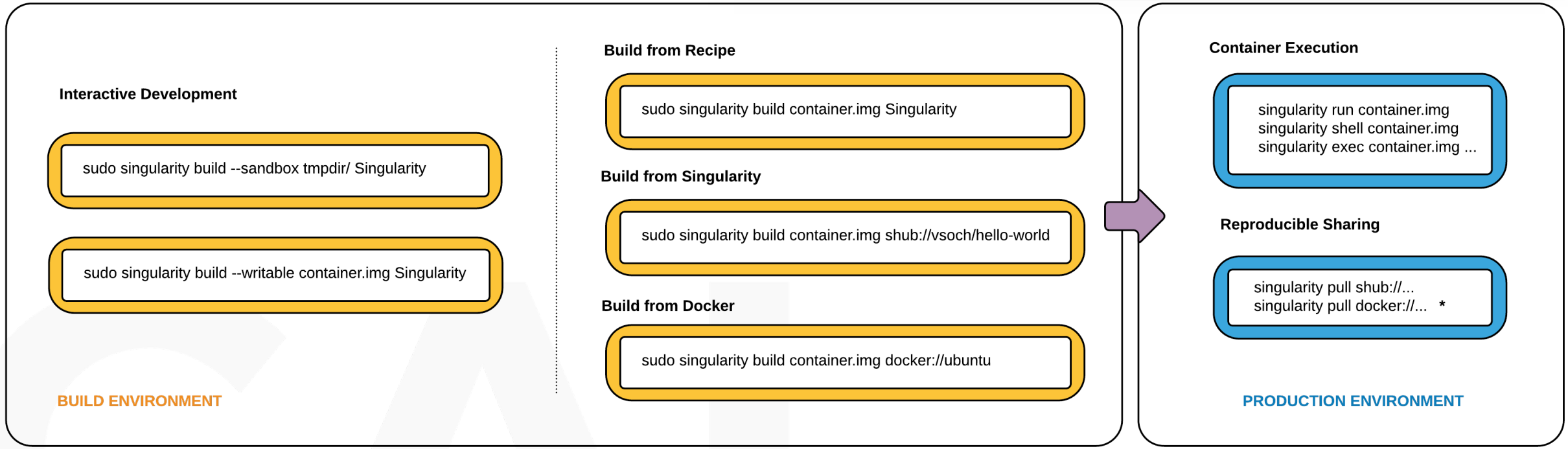
https://github.com/kaczmarj/neurodocker

SINGULARITY

RESEARCH ARTICLE

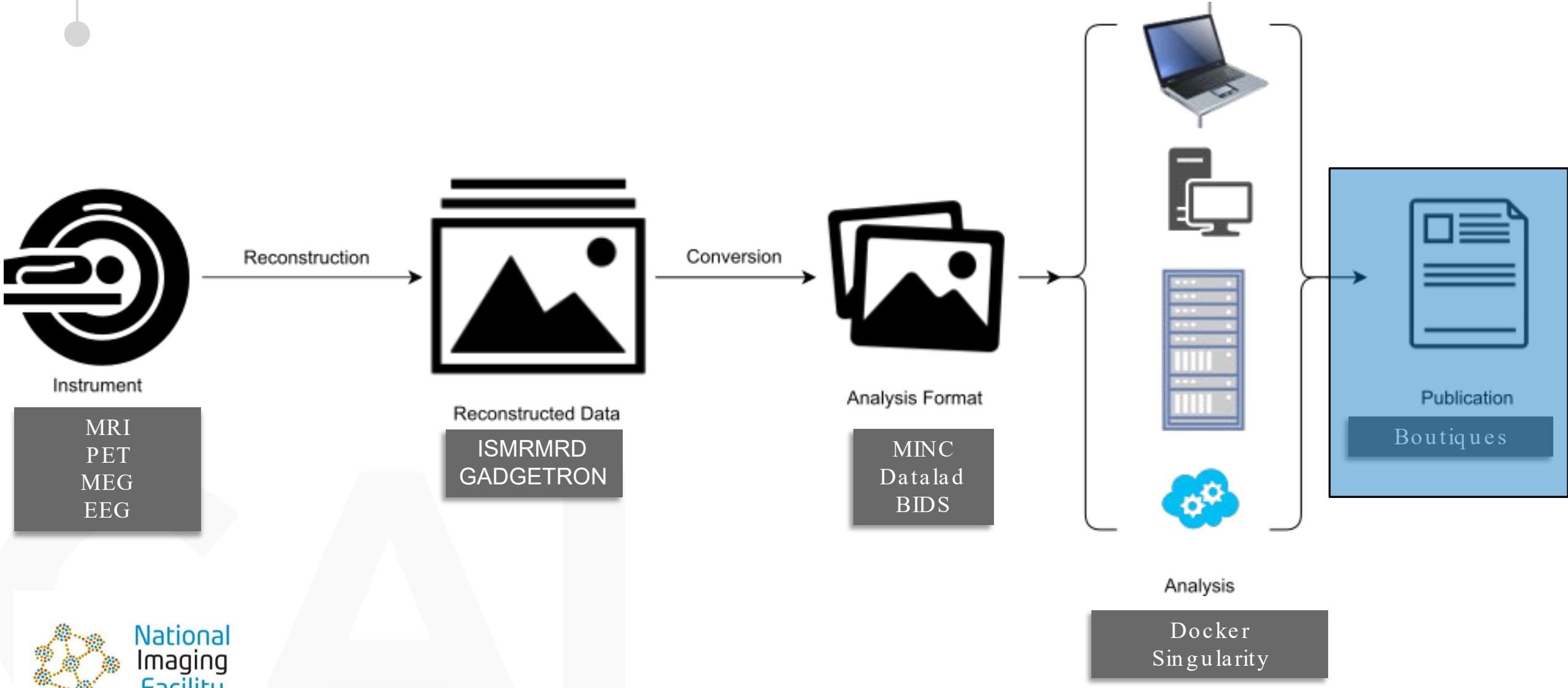
Singularity: Scientific containers for mobility of compute

Gregory M. Kurtzer[1], Vanessa Sochat[2]*, Michael W. Bauer[1,3,4]



https://sylabs.io/singularity/

FROM DATA TO PUBLICATION

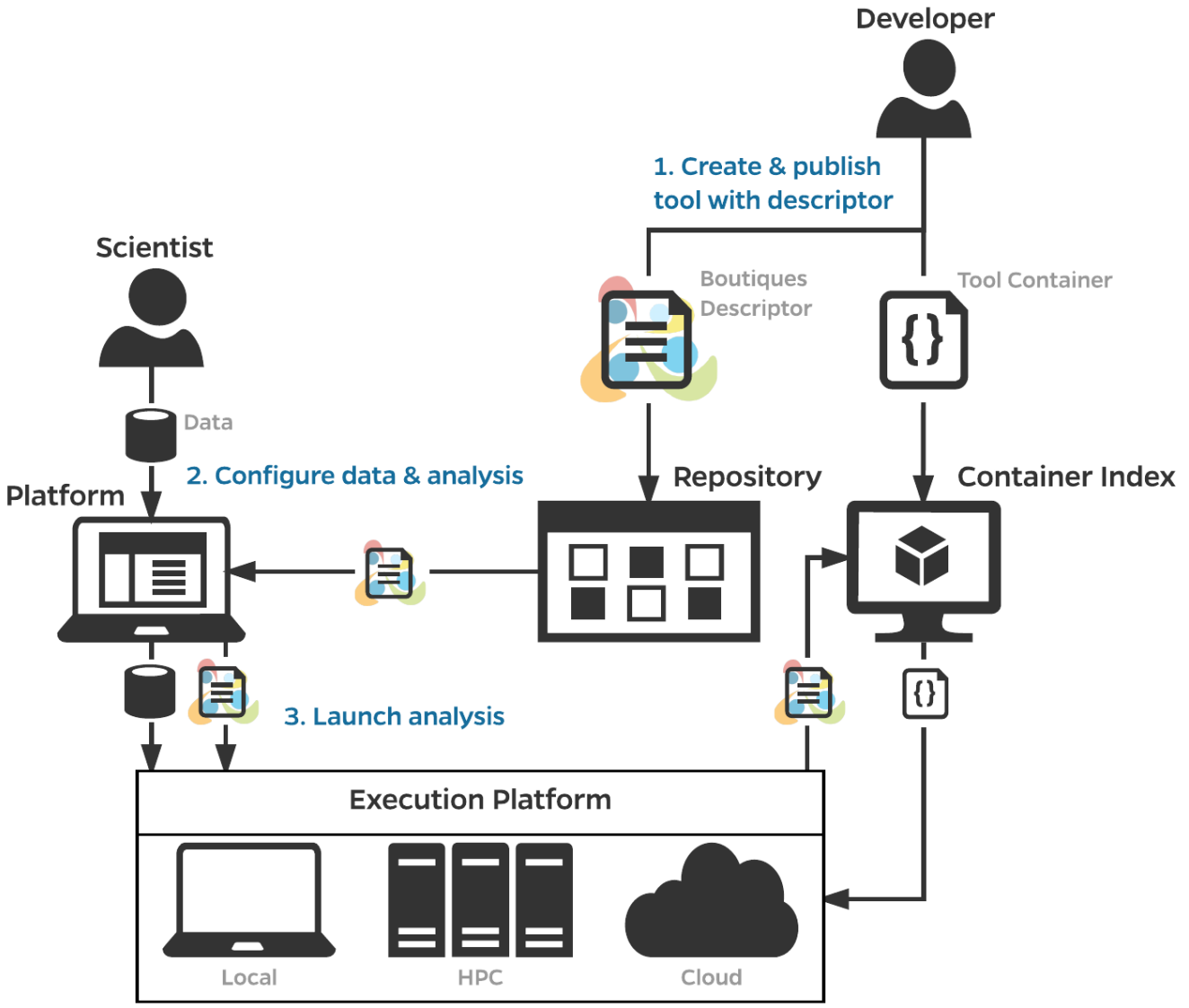


National Imaging Facility

@sbollmann_MRI

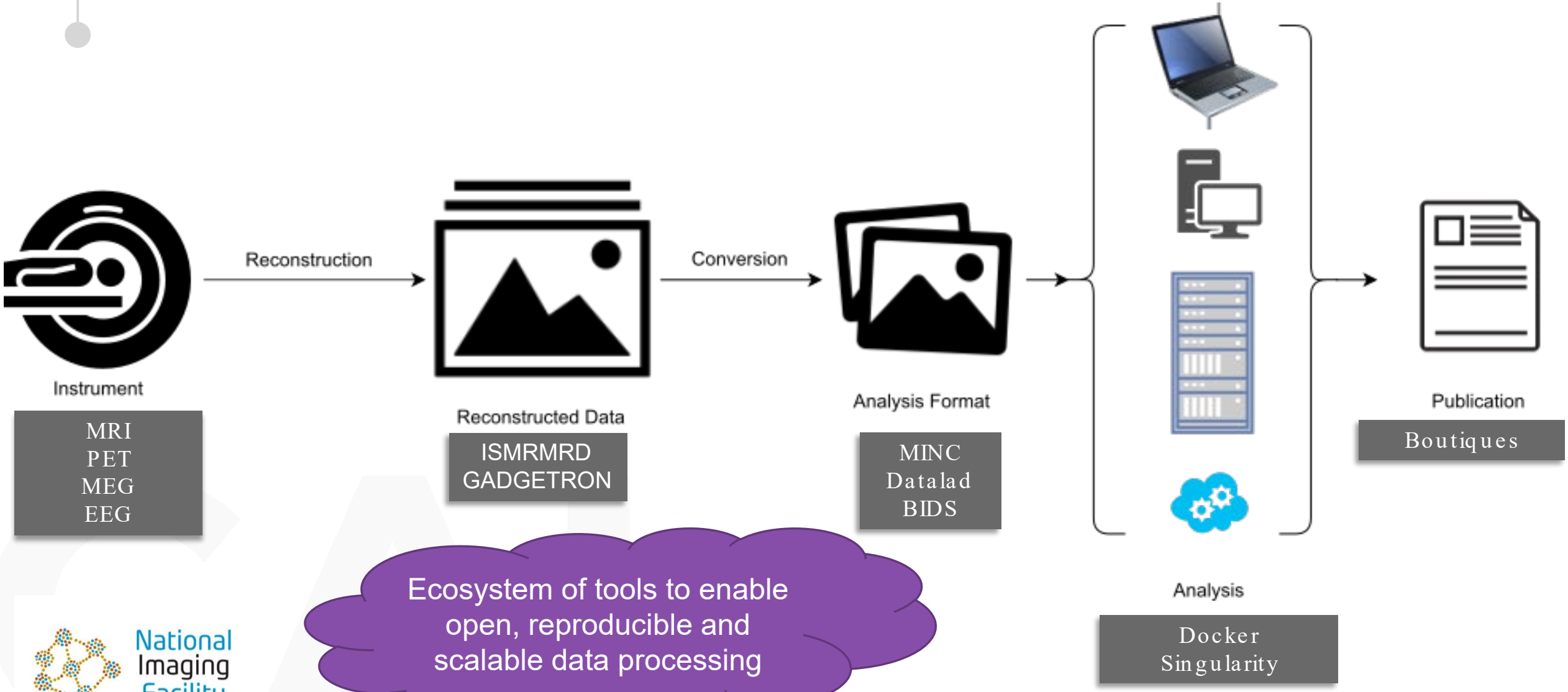
BOUTIQUES

- automatically publish, integrate, and execute applications
- applications summarized in a JSON description



https://boutiques.github.io/

FROM DATA TO PUBLICATION



Ecosystem of tools to enable open, reproducible and scalable data processing

National Imaging Facility

@sbollmann_MRI

Thank you

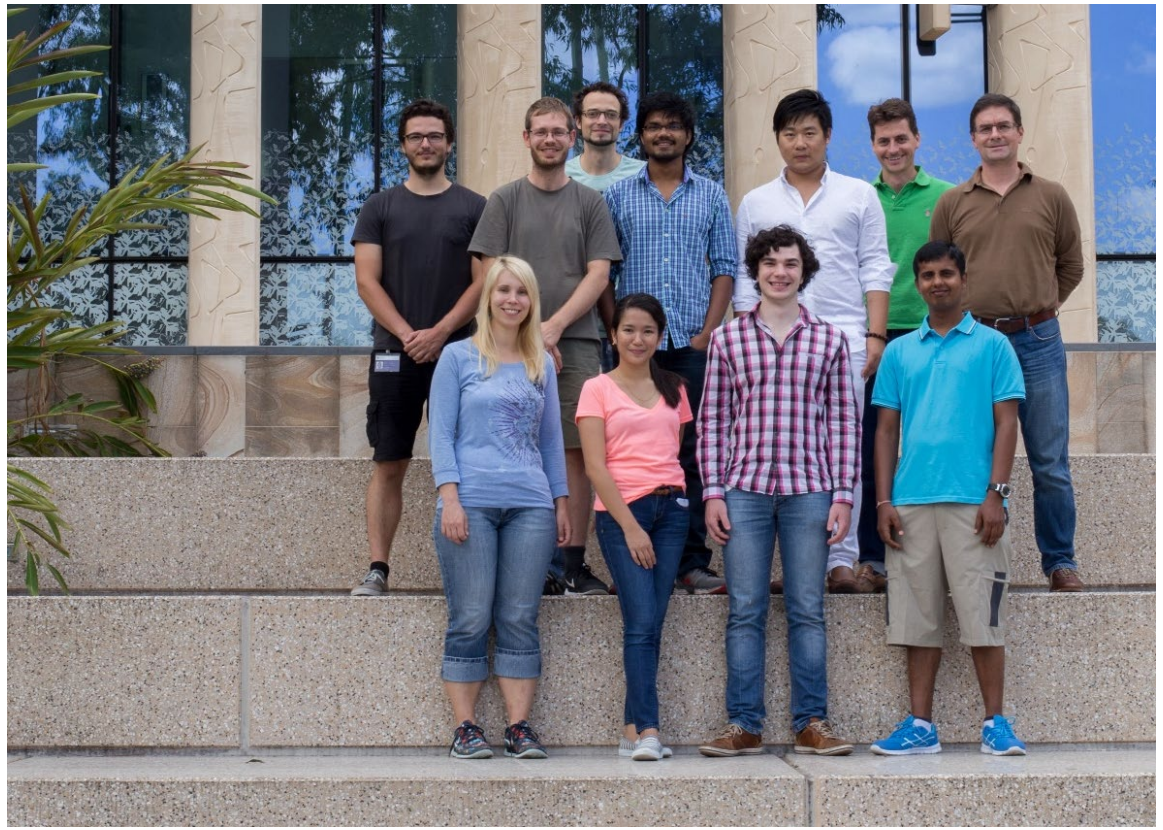
The University Of Queensland
AUSTRALIA

Centre for Advanced Imaging

www.cai.uq.edu.au/bollmann

@sbollmann_MRI

steffen.bollmann@cai.uq.edu.au



SIEMENS Healthineers

National Imaging Facility

Australian Government
Australian Research Council

AUSTRALIAN CANCER RESEARCH FOUNDATION

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

Queensland Government