



ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

Secret Session: Master coding in your research environment

Reproducibility across platforms: why, how, and when to container?

Steffen Bollmann, PhD

Siemens Healthineers

National Imaging Facility Fellow, Centre for Advanced Imaging, The University of Queensland, Australia



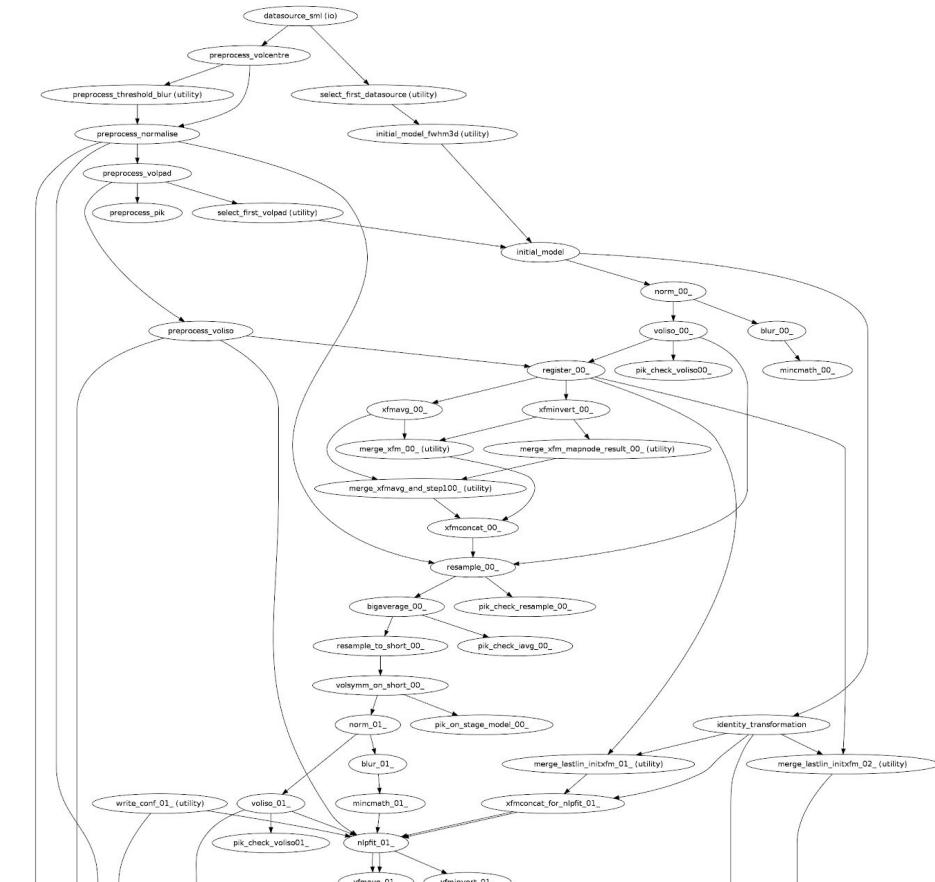
Why Containerization? Sharing an analysis pipeline

one example nipype neuroimaging pipeline:

- convert dicoms to BIDS using heudiconv
- brain extraction using ROBEX
- N4 bias correction
- fslmaths

Problem:

- complex (inter)dependencies make sharing and reproducing this pipeline difficult





ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

Why Containerization? Running analyses on different Platforms

- Notebook: Windows 10 / Mac OS
- Institute cluster: 4.15.0-47-generic Ubuntu 16.04.6 LTS
- University cluster: 3.10.0-693.5.2.el7.x86_64

Problems:

- research software has to be installed/compiled for every system
- compiling software on older HPC operating systems can be very difficult due to outdated dependencies (e.g. Tensorflow ...)
- analysis results can change if operating systems get updated (no control)

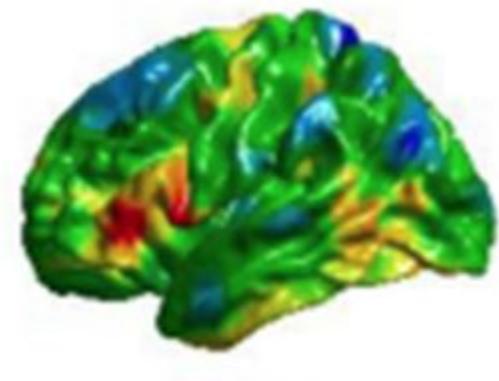
Why Containerization? GLIBC 2.5 vs 2.18

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



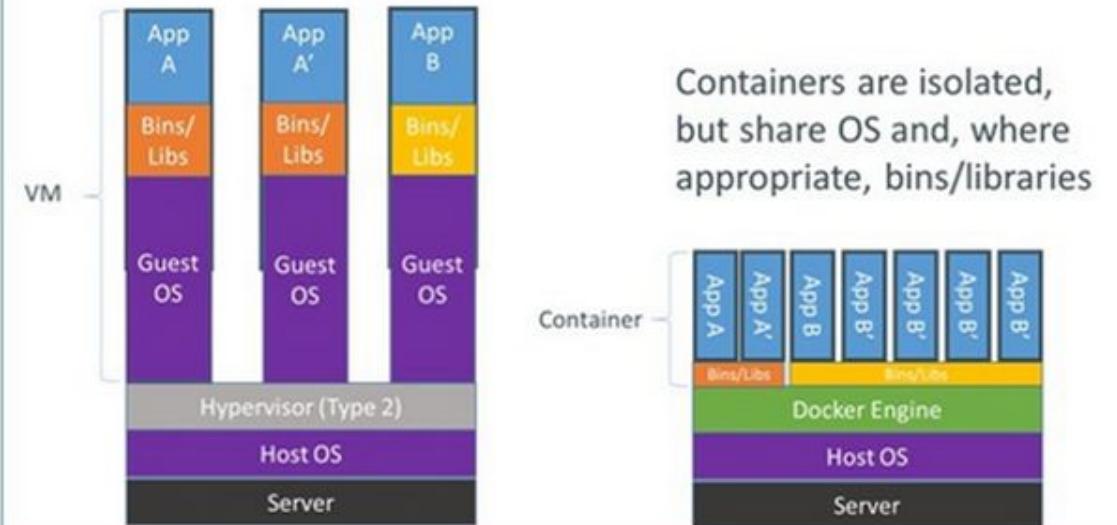
ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

What is Docker?

- Containers are lighter than VMs 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

Containers vs. VMs





ISMRM 27TH ANNUAL MEETING & EXHIBITION

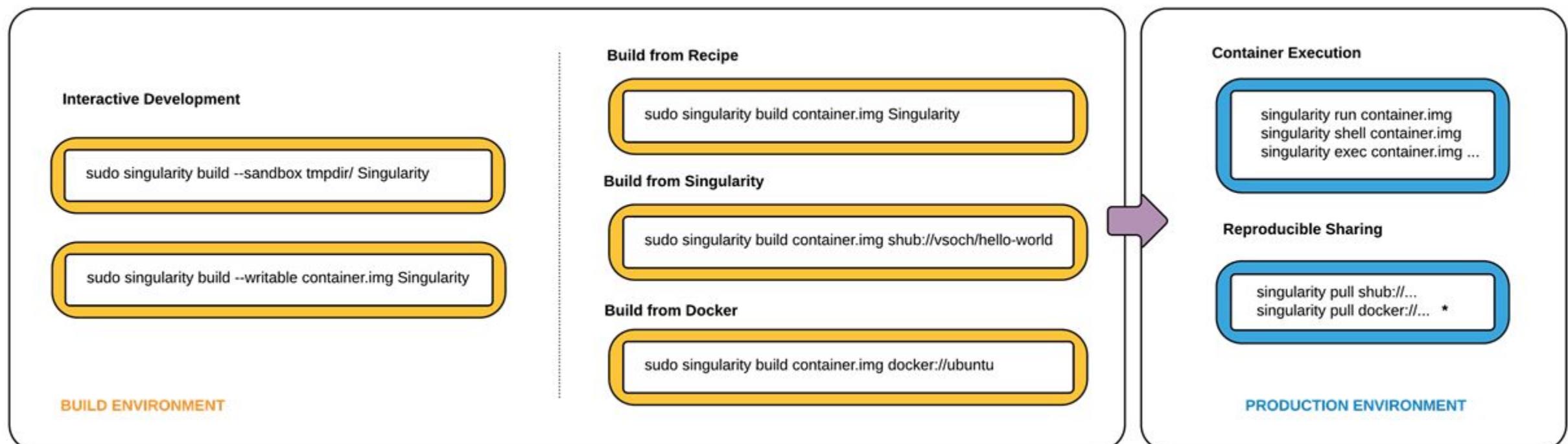
Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

What is Singularity?

RESEARCH ARTICLE

Singularity: Scientific containers for mobility of compute

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





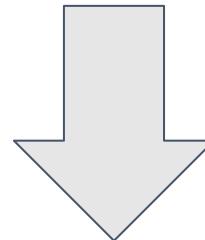
ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

What is Transparent Singularity? <https://github.com/CAIsr/transparent-singularity>

- wrapper to use tools inside a singularity container as if they where installed locally

```
singularity exec --pwd $PWD /software/minc_1p9p16_visual_20180731.simg mincmath -add 1.mnc 2.mnc out.mnc
```



```
mincmath -add 1.mnc 2.mnc out.mnc
```

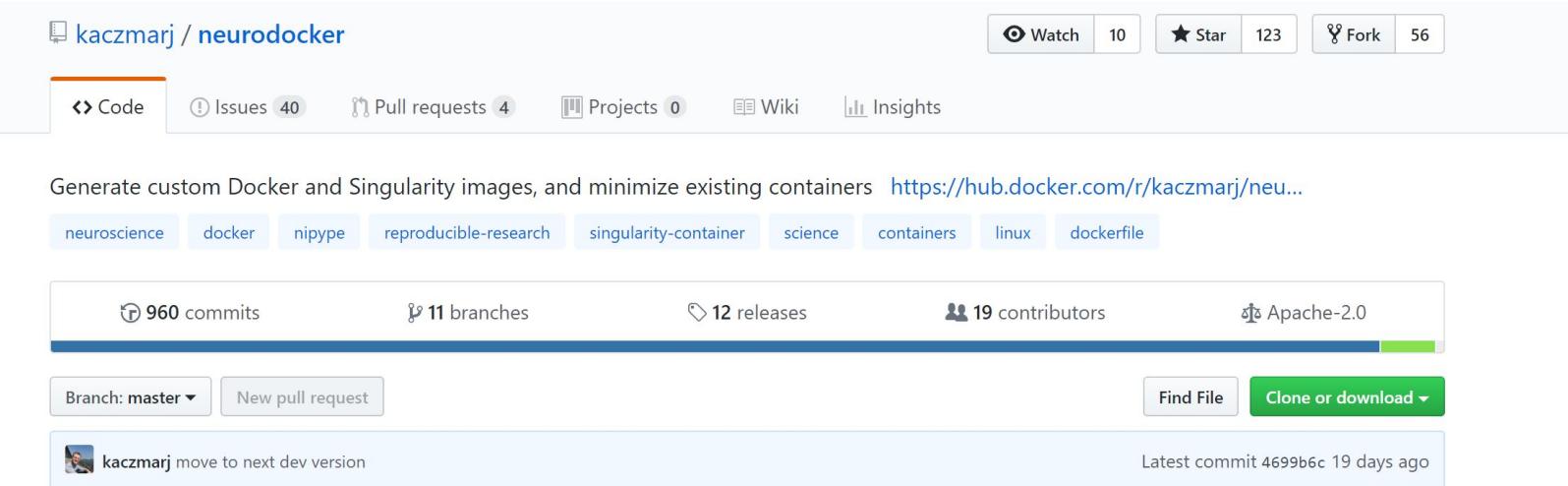


ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

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

- builds recipe files for docker / singularity to create containers with most neuroimaging software: AFNI, ANTs, Convert3D, dcm2niix, FreeSurfer, FSL, Matlab Compiler Runtime, MINC, Miniconda, MRtrix3, PETPVC, SPM12 + everything in Neurodebian!



The screenshot shows the GitHub repository page for `kaczmarj/neurodocker`. The page includes the repository name, a navigation bar with links to Code, Issues (40), Pull requests (4), Projects (0), Wiki, and Insights. It also displays statistics such as 960 commits, 11 branches, 12 releases, 19 contributors, and an Apache-2.0 license. A note at the top states: "Generate custom Docker and Singularity images, and minimize existing containers <https://hub.docker.com/r/kaczmarj/neu...>". The repository has 10 watchers, 123 stars, and 56 forks. At the bottom, there are buttons for Branch: master, New pull request, Find File, Clone or download, and a message from `@kaczmarj` about moving to next dev version.



ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

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



ISMRM 27TH ANNUAL MEETING & EXHIBITION

Palais des congrès de Montréal  Montréal, QC, Canada  11–16 May 2019

Food for Thought

- scientists are not software developers (and shouldn't be): But we use and write tools to implement our ideas - what are software development workflows that we can efficiently use?
- how can we make more of our tools easier to access?
- how can we run software across multiple platforms?
- how can we make it more desirable to contribute to existing projects instead of re-inventing?
- should we release everything we write open source? What is our responsibility after releasing (e.g. AFNI bug)? Refactoring before release? Testing? Community? Documentation? Who pays for longterm maintenance?