Torwards Software Product Lines for (f)MRI analysis

RIPOST WORKSHOP
SIMULA, Oslo,
29/10/2025

Paul TEMPLE
Gauthier LE BARTZ LYAN

MRI analysis challenges

- Various data (organs, people, dimension, resolution,...)
- Noise at acquisition(Movement, residual magnetic fields, ...)
- Computational complexity (interpolation, filtering, ...)
- Analysis noise (floating points error, software configuration, human mistakes / habits, ...)

Élodie Germani. Exploring and mitigating analytical variability in fMRI results using representation learning. Medical Imaging. Université de Rennes, 2024.

MRI analysis challenges

- **Researchers question the reliability of research based on medical imaging.**

- R. Botvinik-Neze *et al*. Variability in the analysis of a single neuroimaging dataset by many teams,” Nature, vol. 582, no. 7810, pp. 84–88, Jun. 2020.
- A. Boucard, A. Marchand, and X. Nogu`es, “Reliability and validity of structural equation modeling applied to neuroimaging data: a simulation study,” Journal of neuroscience methods, vol. 166, 2007
- A. M. Brandmaier, E. Wenger, N. C. Bodammer, S. K¨uhn, N. Raz, and U. Lindenberger, “Assessing reliability in neuroimaging research through intra-class effect decomposition (iced),” Elife, vol. 7, 2018.
- R. L. Billingsley-Marshall, P. G. Simos, and A. C. Papanicolaou, “Reliability and validity of functional neuroimaging techniques for identifying language-critical areas in children and adults,” Developmental neuropsychology, vol. 26, 2004
- J. C. Fournier, H. W. Chase, J. Almeida, and M. L. Phillips, “Model specification and the reliability of fmri results: implications for longitudinal neuroimaging studies in psychiatry,” PLoS One, vol. 9, 2014.

NARPS

- **N.A.R.P.S** stands for **N**euroImaging **A**nalysis **R**eplication and **P**rediction **S**tudy

- **GOAL** : Evaluate consensuality of 70 research teams over a set of 9 **R**esearch **Q**uestions

NARPS : R. Botvinik-Neze *et al*. Variability in the analysis of a single neuroimaging dataset by many teams," Nature, vol. 582, no. 7810, pp. 84–88, Jun. 2020.

NARPS

- **N.A.R.P.S** stands for **N**euroImaging **A**nalysis **R**eplication and **P**rediction **S**tudy

- **GOAL** : Evaluate consensuality of 70 research teams over a set of 9 **R**esearch **Q**uestions

- **RESULTS** : Consensus was never reached.

NARPS : R. Botvinik-Neze *et al*. Variability in the analysis of a single neuroimaging dataset by many teams," Nature, vol. 582, no. 7810, pp. 84–88, Jun. 2020.

Why NARPS is a good playground ?

- **Exogenous variability is partly controlled :**
  - ➢Dataset quality has been assessed before the study
  - ➢All teams results and pipelines have been described and are publicly available

Why NARPS is a good playground ?

- **Exogenous variability is partly controlled :**
  - ➢Dataset quality has been assessed before the study
  - ➢All teams results and pipelines have been described and are publicly available

- **Variability by design :**
  - ➢=/= Technos (Python, Matlab, ...)
  - ➢=/= Data processing software (SPM, FSL, FreeSurfer, fMRIPrep,...)
  - ➢=/= Pipelines (software suites x configuration x teams habits)

Variability layers

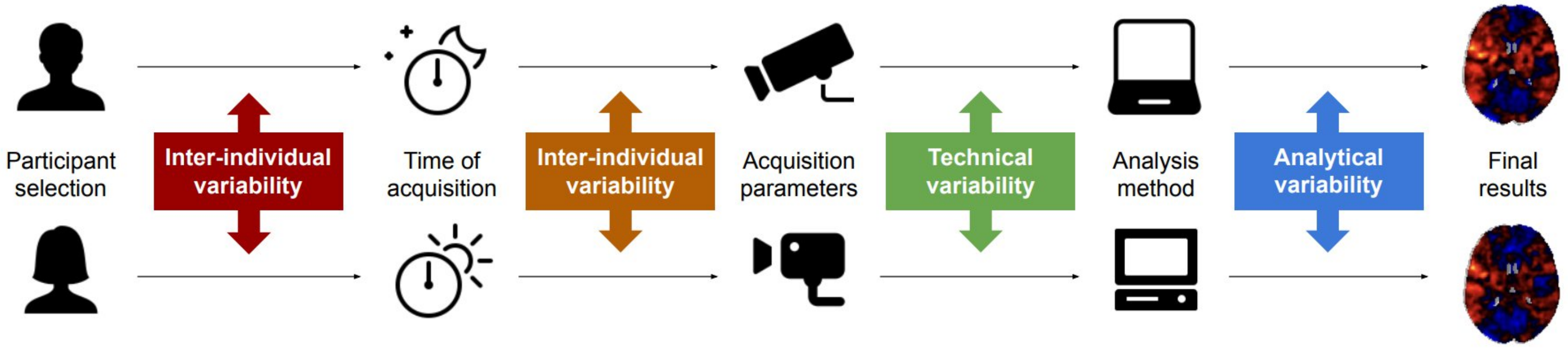


Figure 2.1 – Different sources of variability in neuroimaging studies

Élodie Germani. Exploring and mitigating analytical variability in fMRI results using representation learning. Medical Imaging. Université de Rennes, 2024. English. p.43

Variability layers

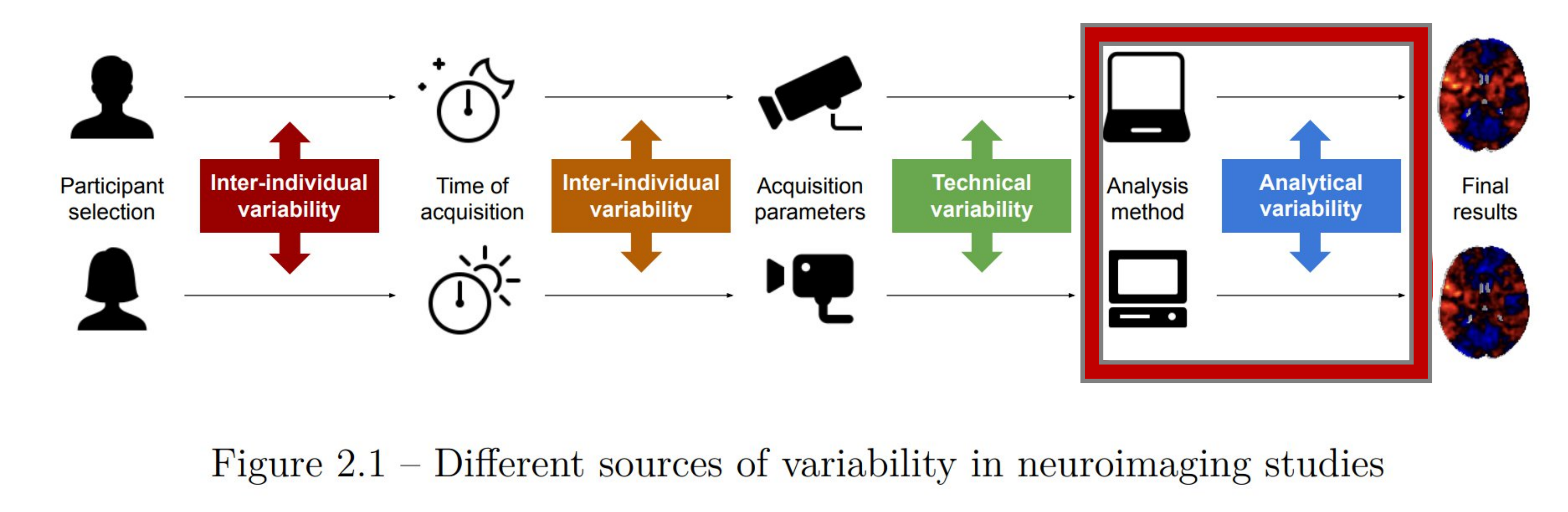
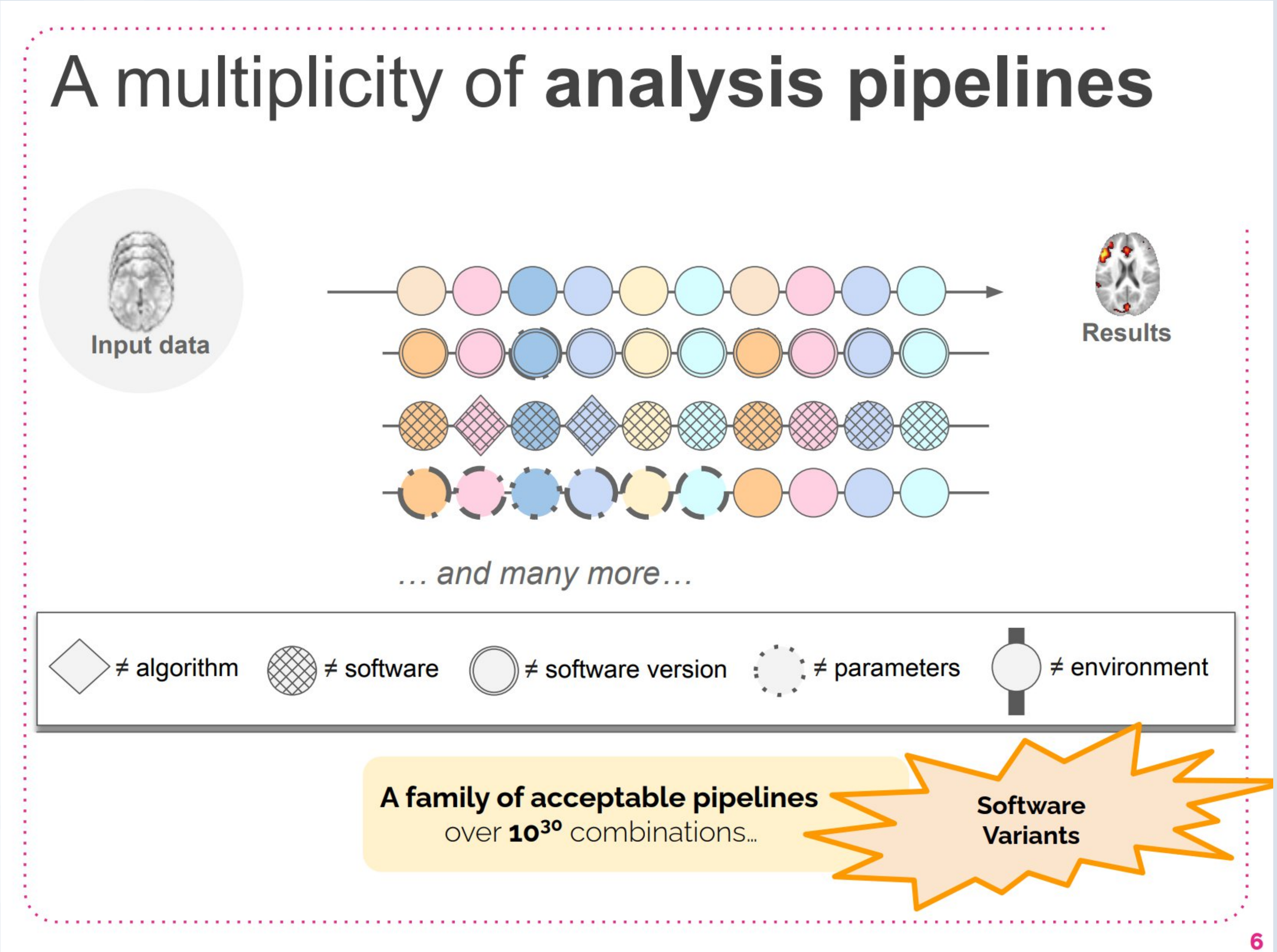


Figure 2.1 – Different sources of variability in neuroimaging studies

Élodie Germani. Exploring and mitigating analytical variability in fMRI results using representation learning. Medical Imaging. Université de Rennes, 2024. English. p.43



A multiplicity of **analysis pipelines**

Input data

Results

… and many more…

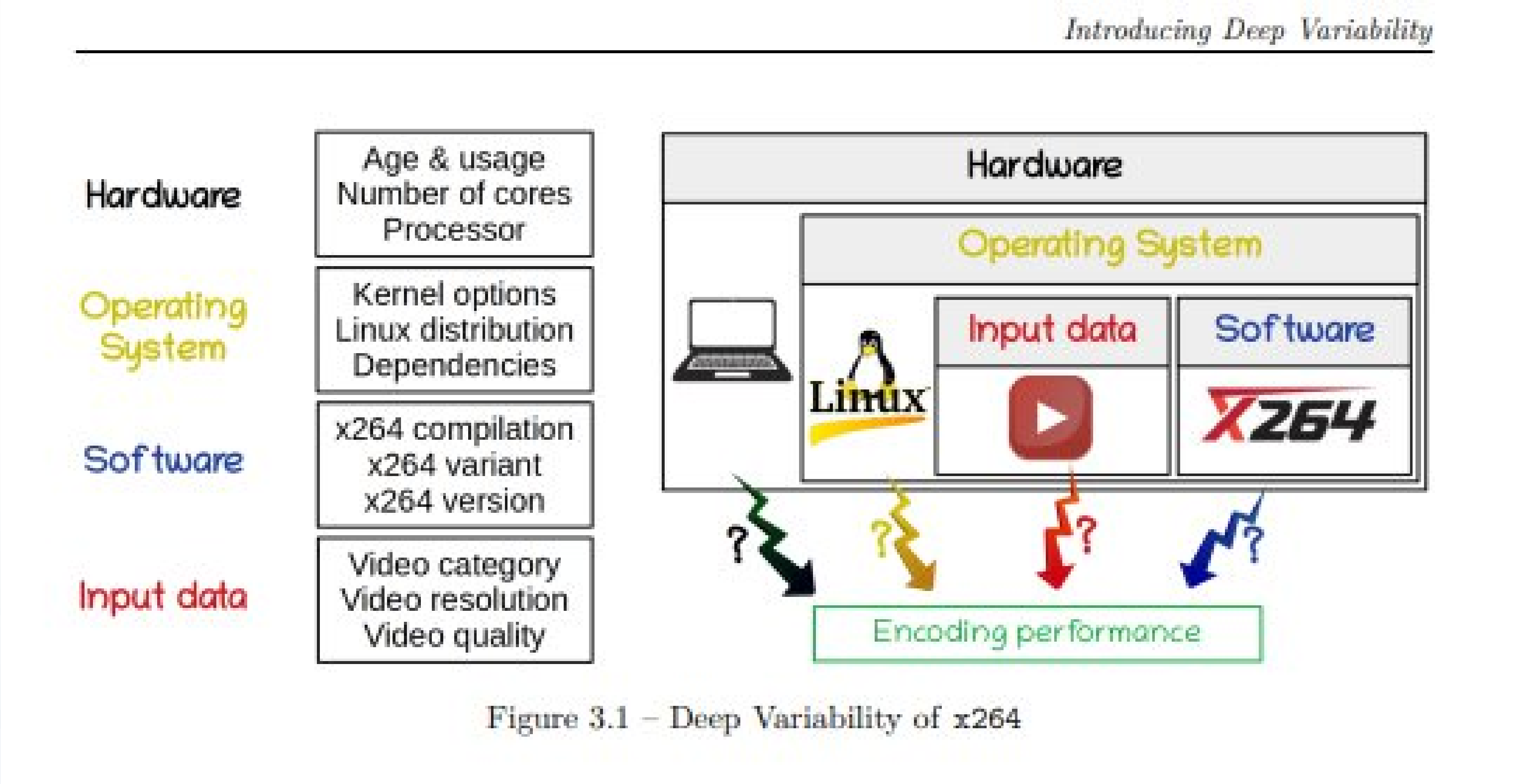
≠ algorithm ≠ software ≠ software version ≠ parameters ≠ environment

A family of acceptable pipelines
over **10^{30}** combinations...

Software Variants

Camille Maumet. Variability in brain imaging studies across different analysis pipelines. VAMOS 2025 - 19th International Working Conference on Variability Modelling of Software-Intensive Systems, Feb 2025, Rennes, France. 2025. ffhal-04926453v2f

Deep Software Variability



Introducing Deep Variability

Hardware	Age & usage Number of cores Processor
Operating System	Kernel options Linux distribution Dependencies
Software	x264 compilation x264 variant x264 version
Input data	Video category Video resolution Video quality

Figure 3.1 – Deep Variability of x264

Luc Lesoil. Deep software variability for resilient performance models of configurable systems. Other [cs.OH]. Université de Rennes, 2023. English. p 43

What we already know

- Use of different Hardware can induce variation in results

  G. Vila, E. Medernach, I. Gonzalez Pepe, A. Bonnet, Y. Chatelain, M. Sdika, T. Glatard, and S. Camarasu Pop, “The impact of hardware variability on applications packaged with docker and guix: a case study in neuroimaging,” in Proceedings of the 2nd ACM Conference on Reproducibility and Replicability, ser. ACM REP ’24.

- Software configuration can lead to discrepancies in fMRI prepped data

  Y. Chatelain, L. Tetrel, C. J. Markiewicz, M. Goncalves, G. Kiar, O. Esteban, P. Bellec, and T. Glatard, “A numerical variability approach to results stability tests and its application to neuroimaging,” IEEE Trans. Comput., vol. 74, no. 1, p. 200–209, Jan. 2025.

- Software versions can lead to divergence between series of runs in brain **structural** analysis workflows

  A. Sokołowski, N. Bhagwat, D. Kirbizakis, Y. Chatelain, M. Dugr´e, J.-B. Poline, M. Sharp, and T. Glatard, “The impact of freesurfer versions on structural neuroimaging analyses of parkinson’s disease,” bioRxiv, pp. 2024–11, 2024.

Experimentation over NARPS

- What kind of Exps. are we talking about ?
  - Fix version of SPM (SPM12) with different revisions/releases
  - Matlab (comes with SPM)
  - 100 subjects
  - All 9 RQs

=> Were able to re-run (only) 2 pipelines from 2 different teams (U26C and 2T6S)

NARPS open pipelines : https://github.com/Inria-Empenn/narps_open_pipelines

Experimentation over NARPS

- Ad-hoc docker containers with customizable Docker files
  - Aim : reproducibility AMAP.

Experimentation over NARPS

- Ad-hoc docker containers with customizable Docker files
  - Aim : reproducibility AMAP.

- Tens of runs over different versions of matlab, SPM12

- Tens of runs over different hardware : x86, ARM

- Hundreds of GB of data generated

Experimentation over NARPS

- Ad-hoc docker containers with customizable Docker files
  - Aim : reproducibility AMAP.
- Tens of runs over different versions of matlab, SPM12
- Tens of runs over different hardware : x86, ARM
- Hundreds of GB of data generated
- Ad-hoc code for statistical (matrices) analysis and dataviz

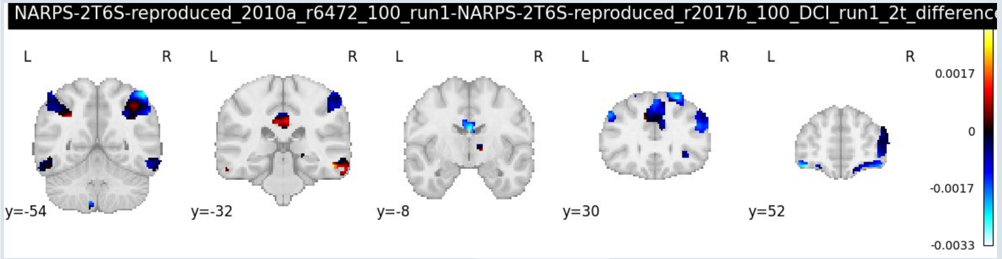
Differences in Results

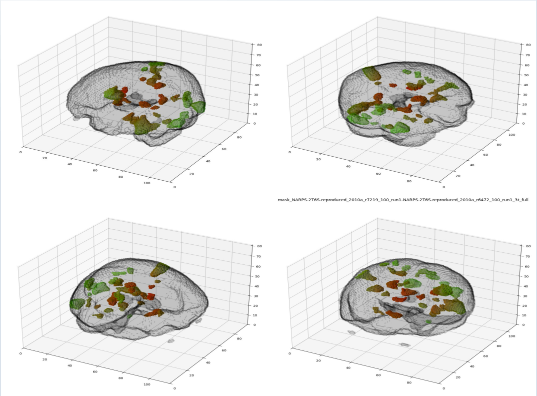
- **Minor statistical discrepancies on preproc. AND final data:**
  - *~[6E-5 ; 1E-3]* between comparable runs
  - Meaningful data usually ranges in [-8 ; 8]

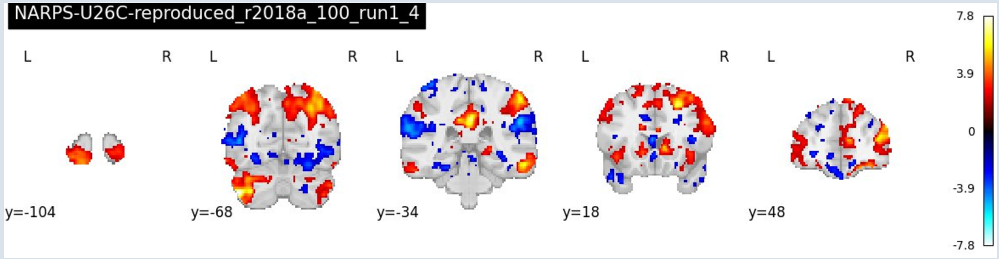
Differences in Results

- **Minor statistical discrepancies on preproc. AND final data:**
  - *~[6E-5 ; 1E-3]* between comparable runs
  - Meaningful data usually ranges in [-8 ; 8]

- **Minor visual differences in outcome data, most of it bound to statistical differences :**
  - We had to emphasize statistical differences a lot to be able to filter them out


NARPS-2T6S-reproduced_2010a_r6472_100_run1-NARPS-2T6S-reproduced_r2017b_100_DCI_run1_2t_differenc



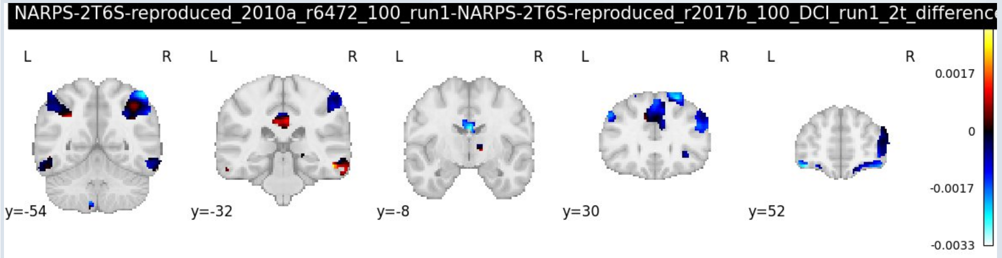

NARPS-U26C-reproduced_r2018a_100_run1_4

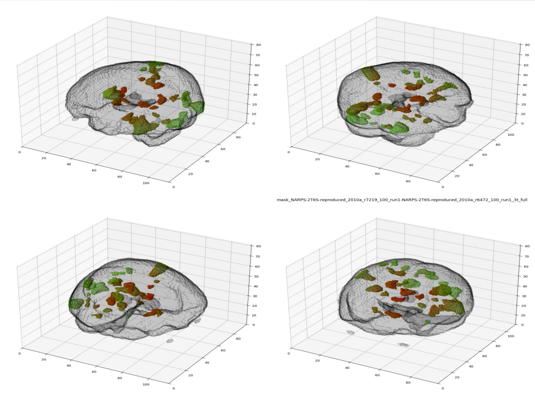
Differences in Results

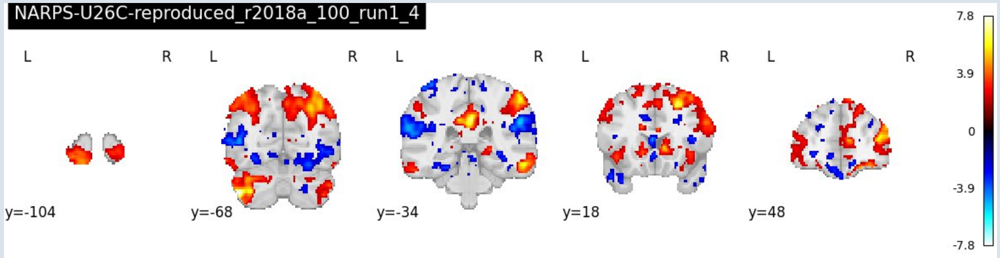
- **Minor statistical discrepancies on preproc. AND final data:**
  - *~[6E-5 ; 1E-3]* between comparable runs
  - Meaningful data usually ranges in [-8 ; 8]

- **Minor visual differences in outcome data, most of it bound to statistical differences :**
  - We had to emphasize statistical differences a lot to be able to filter them out

- **Nothing Major came out of it BUT the complexity of the software environment, pipeline usage, configuration, and fMRI silos.**


NARPS-2T6S-reproduced_2010a_r6472_100_run1-NARPS-2T6S-reproduced_r2017b_100_DCI_run1_2t_differenc
L
R
y=-54
y=-32
y=-8
y=30
y=52
0.0017
0
-0.0017
-0.0033


mask_NARPS-2T6S-reproduced_2010a_r7219_100_run1-NARPS-2T6S-reproduced_2010a_r6472_100_run1_3t_full



Experimentation challenges

- Heavy software environment (many dependencies, discrete pieces of software, configuration and versions)

- Huge dataset and storage needs

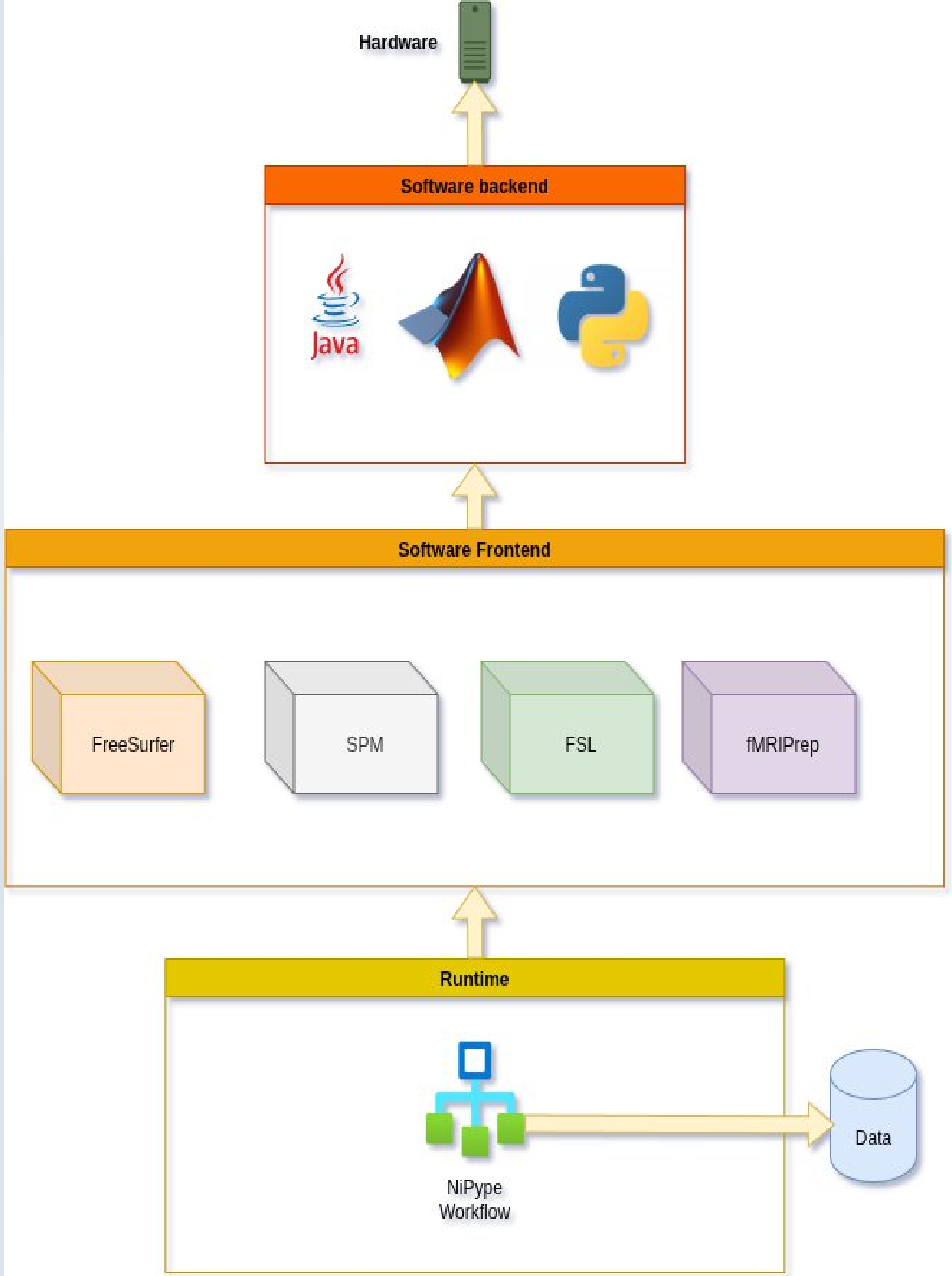
Experimentation challenges

- Heavy software environment (many dependencies, discrete pieces of software, configuration and versions)
- Huge dataset and storage needs
- Lack of domain expert knowledge
- Interpretability of outputs

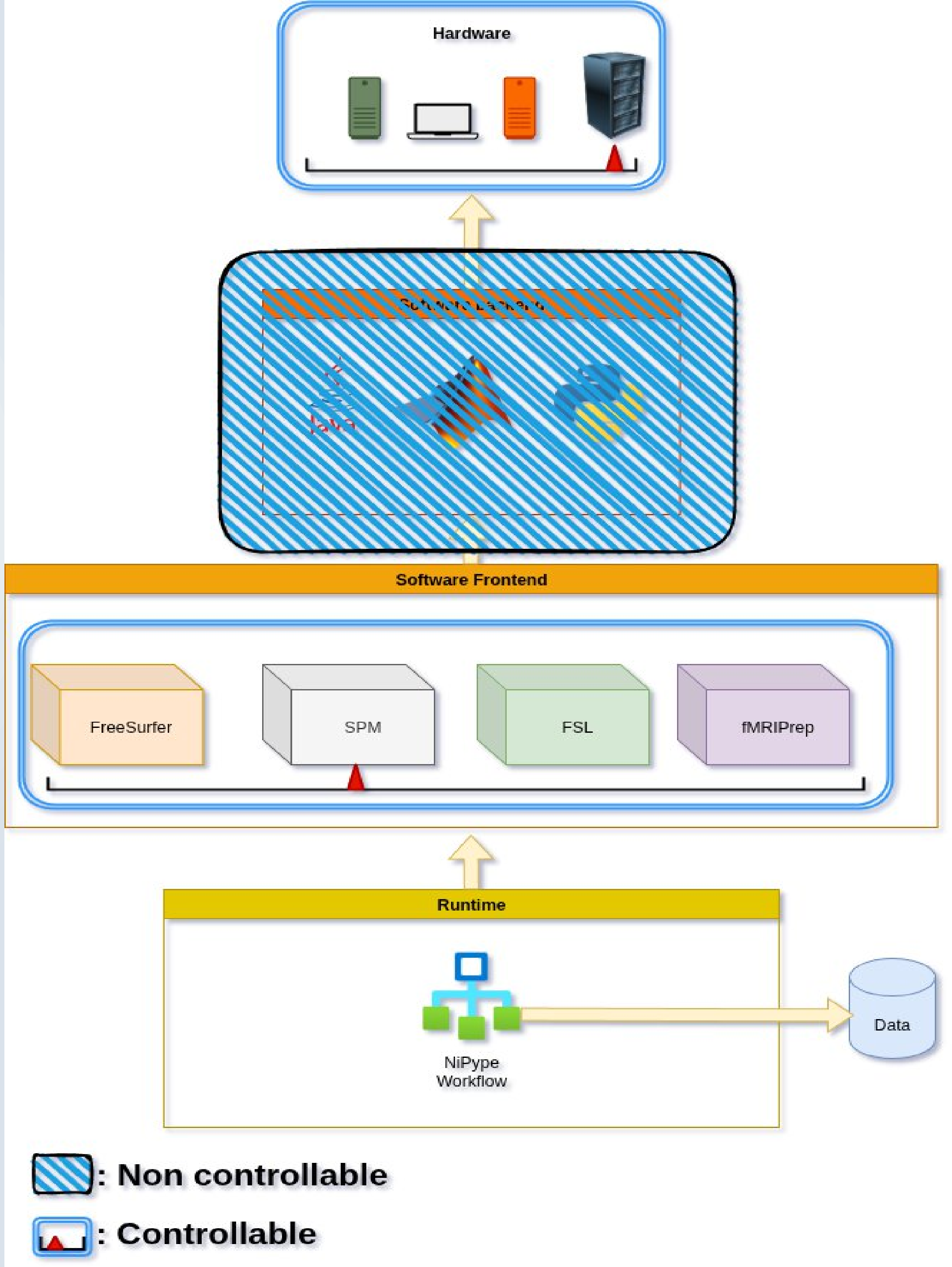
Experimentation challenges

- Heavy software environment (many dependencies, discrete pieces of software, configuration and versions)
- Huge dataset and storage needs
- Lack of domain expert knowledge
- Interpretability of outputs
- **Complexity of (re)implementation of pipelines**

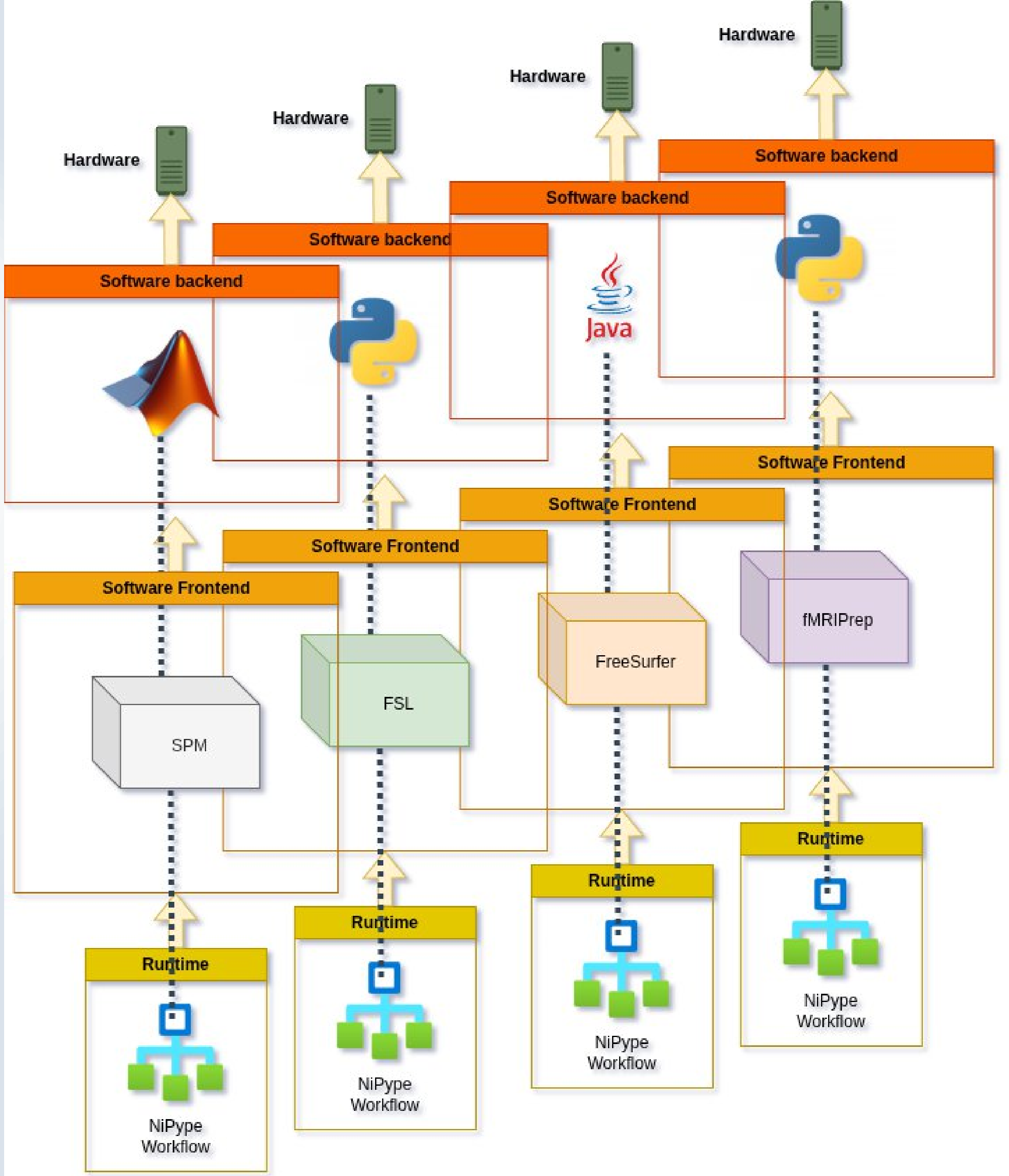
Pipelines structure



Pipelines control



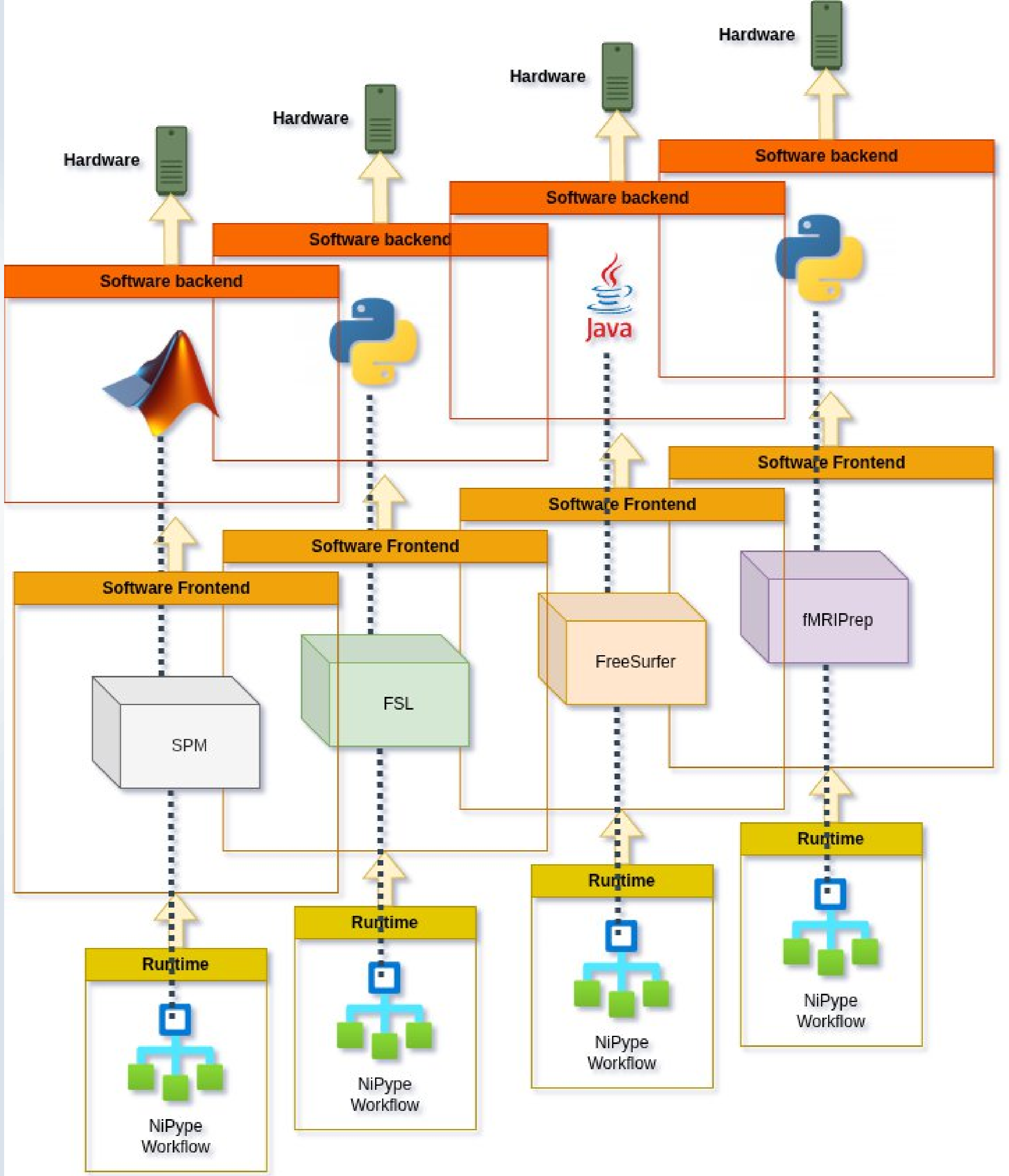
Pipelines issues



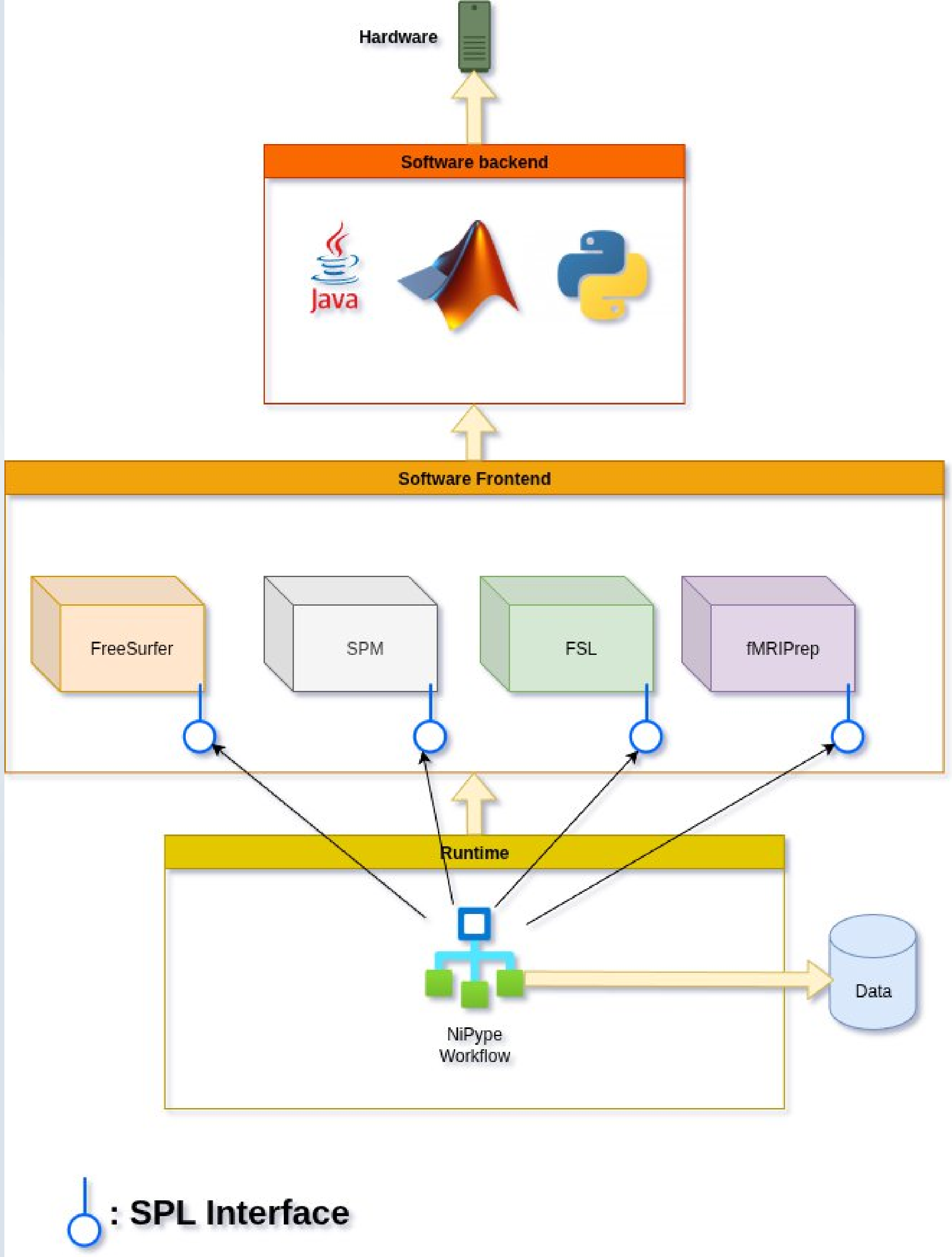
Pipelines issues

Problems :

1. No inter-operability
2. Code duplication
3. Errors spreading
4. No OOB comparison between tools
5. No standard tooling
6. Maintenance
7. Interfacing for middleware alike NiPype
8. NO TESTING



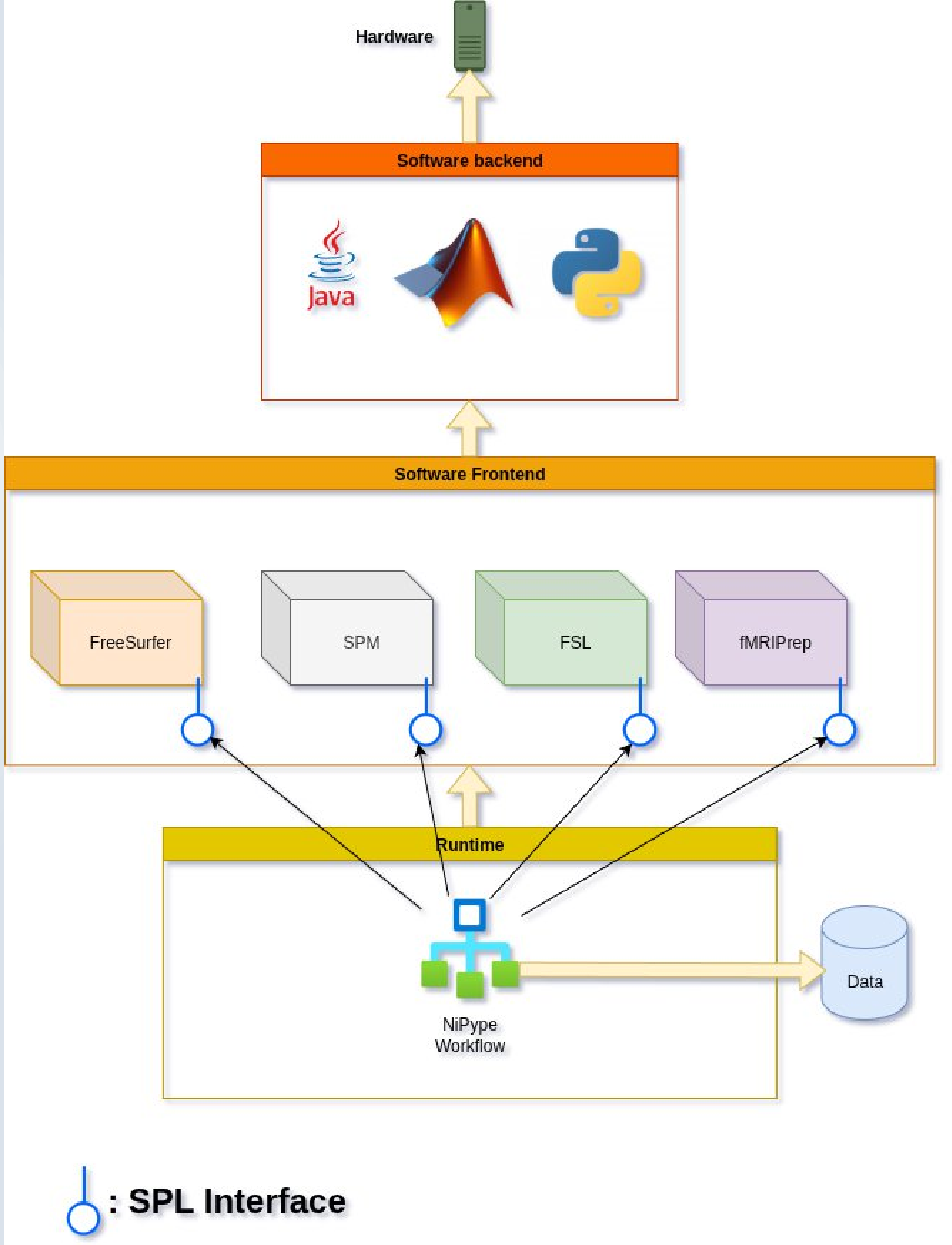
Pipelines SPL



Pipelines SPL

Advantages:

1. Inter-operability
2. Less code duplication
3. Less errors spreading
4. Possible OOB comparison between tools
5. Standardized tooling
6. Possibly less maintenance
7. <u>TESTING FINALLY AVAILABLE</u>



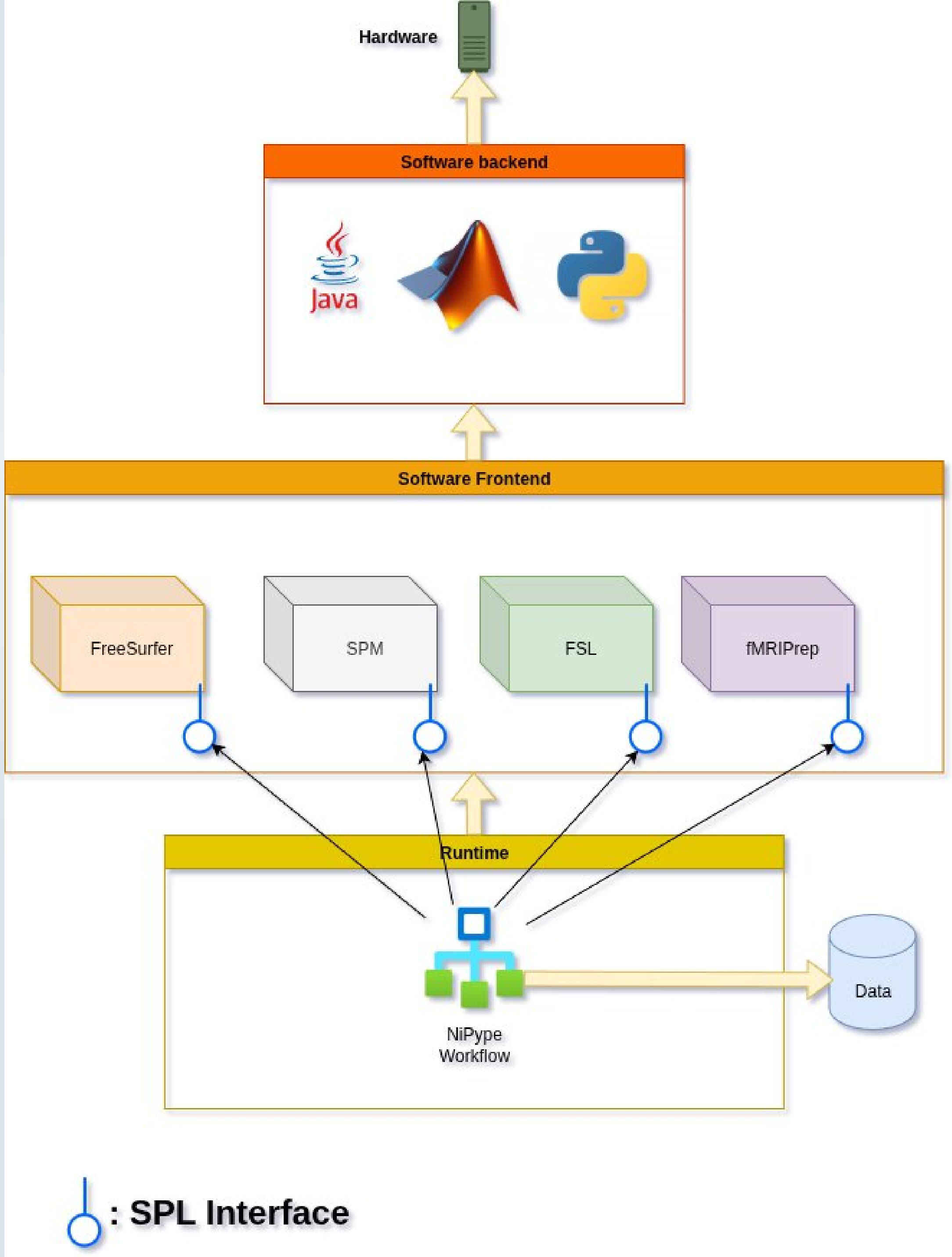
Pipelines SPL

Advantages:

1. Inter-operability
2. Less code duplication
3. Less errors spreading
4. Possible OOB comparison between tools
5. Standardized tooling
6. Possibly less maintenance
7. <u>TESTING FINALLY AVAILABLE</u>

Caveats:

1. Huge step over for stakeholders
2. Stakeholders must reunite and agree on a standard interface
3. Possibly time and money demanding
4. Many stakeholders are OS community based



Wrap-up

- Medical Imaging analysis is hard to perform
- Lot of variability (material, software, pipeline, etc.)
- Why not propose an SPL approach ?
  - An interface => select the pipeline to run
  - Ease for reproducing
  - Ease for comparison
  - More resilient
  - ...
- We are not there yet...
  - No inter-operability
  - Heavy maintenance
  - No standard tooling
  - ...

Should it lead to a paper ?