Tools and standards to make neuroimaging derived data reusable
Camille Maumet

To cite this version:
Camille Maumet. Tools and standards to make neuroimaging derived data reusable. Neuroinformatics 2018, Aug 2018, Montreal, Canada. inserm-01886089

HAL Id: inserm-01886089
https://inria.hal.science/inserm-01886089
Submitted on 2 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Tools and standards to make neuroimaging derived data reusable

Camille Maumet

Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, VisAGeS ERL U-1228
Sample sizes in neuroimaging

2015: 30 subjects / study

Poldrack et. al (2017)
Towards large-scale brain imaging studies
More and more shared data available

**Single study**
- 30 subjects

**Consortium**
- 1000 subjects

**Cohort**
- 1,000 - 100,000 subjects

- + Images
- + Homogeneous
- - Fewer datasets

Logos and links:
- OpenNEURO
- studyforrest.org
- NEUROVAULT
- ABIDE Autism Brain Imaging Data Exchange
- 1000 Functional Connectomes Project
- NITRC
- ADHD200
- CORR
- NKI
- HUMAN Connectome Project
- biobank
New challenges when working with open data
New challenges when working with open data

Raw data

Feature extraction

Derived data

Statistical analysis

Results
New challenges when working with open data

1. Raw data
2. Feature extraction
3. Derived data
4. Statistical analysis
5. Results

Diagram:
- Raw data → Feature extraction → Derived data → Statistical analysis → Results

- Raw data → Feature extraction → Derived data
New challenges when working with open data
New challenges when working with open data

1. Feature extraction
   - Derived data

2. Statistical analysis
   - Results

3. Meta-analyses

4. Raw data
   - Feature extraction
   - Derived data

5. Raw data
   - Feature extraction
   - Derived data

6. Raw data
   - Feature extraction
   - Derived data
2. Analytic variability?
Analytic variability

“Different **acceptable** analysis methods”
Carp et al. (2012)
Analytic variability
Analytic variability

Spatial registration
Segmentation
Cross-modality registration
etc.
Analytic variability
Analytic variability

≠ algorithm

≠ algorithm
Analytic variability

≠ software

≠ algorithm

≠ software
Analytic variability

≠ software version

≠ algorithm

≠ software

≠ software version
Analytic variability

≠ parameters

≠ algorithm
≠ software
≠ software version
≠ parameters
Analytic variability
Analytic variability

≠ software
≠ parameters
≠ environment
≠ software version
≠ algorithm
How does analytic variability impact neuroimaging results?
Background

▷ Choice of Analysis Pipeline (functional MRI)
  Carp (2012)

▷ Choice of Analysis Software Version
  Groenenschild et al (2012)

▷ Choice of operating system

▷ Choice of analysis software package
  Pauli et al (2016)
Background

▷ Choice of Analysis Pipeline (functional MRI)
  *Carp* (2012)

▷ Choice of Analysis Software Version
  *Groenenschild et al* (2012)

▷ Choice of operating system

▷ Choice of analysis software package
  *Pauli et al* (2016)

Joint work with Alexander Bowring and Thomas Nichols

*Slide by Alexander Bowring, adapted*
Three published studies with available data

Pipelines aimed to replicate the original analyses

Impact of Analysis Software on Task fMRI Results

Reproducing the main figure

Impact of Analysis Software on Task fMRI Results

Reproducing the main figure

<table>
<thead>
<tr>
<th>ds000001</th>
<th>ds000109</th>
<th>ds000120</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFNI</td>
<td>AFNI</td>
<td>AFNI</td>
</tr>
<tr>
<td>FSL</td>
<td>FSL</td>
<td></td>
</tr>
<tr>
<td>SPM</td>
<td>SPM</td>
<td></td>
</tr>
</tbody>
</table>

Dice coefficients

Impact of Analysis Software on Task fMRI Results

Reproducing the main figure

<table>
<thead>
<tr>
<th></th>
<th>ds000001</th>
<th>ds000109</th>
<th>ds000120</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFNI</td>
<td>![AFNI Image]</td>
<td>![AFNI Image]</td>
<td>![AFNI Image]</td>
</tr>
<tr>
<td>x=4</td>
<td>![AFNI Image]</td>
<td>![AFNI Image]</td>
<td>![AFNI Image]</td>
</tr>
<tr>
<td>FSL</td>
<td>![FSL Image]</td>
<td>![FSL Image]</td>
<td>![FSL Image]</td>
</tr>
<tr>
<td>x=4</td>
<td>![FSL Image]</td>
<td>![FSL Image]</td>
<td>![FSL Image]</td>
</tr>
<tr>
<td>SPM</td>
<td>![SPM Image]</td>
<td>![SPM Image]</td>
<td>![SPM Image]</td>
</tr>
<tr>
<td>x=4</td>
<td>![SPM Image]</td>
<td>![SPM Image]</td>
<td>![SPM Image]</td>
</tr>
</tbody>
</table>

Dice coefficients

![Dice Coefficients Chart]

Unthresholded statistics

![Unthresholded Statistics]

Disclosing analytical details

Details about the analysis pipeline are in the code.

In manuscripts, it is best practice to:

- Cite custom code
- Cite libraries and tools
How to cite software? The software citation principles

Software citation principles:

- **Importance**
- **Unique Identification**
- **Credit and Attribution**
- **Persistence**
- **Specificity**
- **Accessibility**

Software is a critical part of modern research...

... yet there is little support for its acknowledgement and citation.

https://blog.datacite.org/software-citation-principles/ by Laura Rueda, adapted.
3. Results

[...] All analysis scripts, results reports, and notebooks for each of study are available through Zenodo (Nielsen and Smith, 2014) at https://zenodo.org/record/1203654 (Alexander Bowring et al., 2018).
Let others cite your code

Attribution

Unique identifier

Specificity

Let others cite your code
Citing libraries and tools

2.2 Data Analyses

All data analyses were conducted using AFNI 17.0.18 (RRID:SCR_005927, (Cox, 1996)), FSL 5.0.10 (RRID:SCR_002823, (Jenkinson et al., 2012)), and SPM12 v6906 (RRID:SCR_007037, (Penny et al., 2011)).
2.2 Data Analyses

All data analyses were conducted using **AFNI** 17.0.18 (RRID:SCR_005927, (Cox, 1996)), **FSL** 5.0.10 (RRID:SCR_002823, (Jenkinson et al., 2012)), and **SPM12** v6906 (RRID:SCR_007037, (Penny et al., 2011)).
Citing libraries and tools

2.2 Data Analyses

All data analyses were conducted using **AFNI 17.0.18 (RRID:SCR_005927, (Cox, 1996))**, **FSL 5.0.10 (RRID:SCR_002823, (Jenkinson et al., 2012))**, and **SPM12 v6906 (RRID:SCR_007037, (Penny et al., 2011))**.
Citing libraries and tools

2.2 Data Analyses

All data analyses were conducted using AFNI 17.0.18 (RRID:SCR_005927, (Cox, 1996)), FSL 5.0.10 (RRID:SCR_002823, (Jenkinson et al., 2012)), and SPM12 v6906 (RRID:SCR_007037, (Penny et al., 2011)).
Citing libraries and tools

2.2 Data Analyses

All data analyses were conducted using AFNI 17.0.18 (RRID:SCR_005927, (Cox, 1996)), FSL 5.0.10 (RRID:SCR_002823, (Jenkinson et al., 2012)), and SPM12 v6906 (RRID:SCR_007037, (Penny et al., 2011)).

https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
Citing libraries and tools

2.2 Data Analyses

All data analyses were conducted using **AFNI** 17.0.18 (RRID:SCR_005927, (Cox, 1996)), **FSL** 5.0.10 (RRID:SCR_002823, (Jenkinson et al., 2012)), and **SPM12 v6906** (RRID:SCR_007037, (Penny et al., 2011)).

Also applies to other (non-neuroimaging) tools, repositories and libraries...

2.5 Scripting of analyses and figures

A Python **Jupyter Notebook** (Kluyver et al., 2016) was created for each of the three studies. Each notebook harvests our results data from **Neurovault (RRID:SCR_003806)** and applies the variety of methods discussed in the previous section using **NiBabel 2.2.0** (Brett et al., 2017), **NumPy 1.13.3** (Walt et al., 2011) and **Pandas 0.20.3** (McKinney and Others, 2010) packages. Figures were created using **Matplotlib 2.1.0** (Hunter, 2007) and **Nilearn 0.4.0** (Abraham et al., 2014).
Provenance is information about entities, activities, and people involved in producing a piece of data or thing” — W3C PROV
Machine-readable provenance
Machine-readable provenance

NIDM: neuroimaging data model
International collaboration with INCF SIG on Reproducibility and Best Practices in Human Brain Imaging
Machine-readable provenance

NIDM Experiment

NIDM Workflow

NIDM Results

Raw data

Pipelines

Results

NIDM: neuroimaging data model
International collaboration with INCF SIG on Reproducibility and Best Practices in Human Brain Imaging.

NIDM-Results → fMRI, VBM results
Harmonised model and provenance.

Joint work with INCF NIDASH working group - https://github.com/incf-nidash/
NIDM-Results in brief

Maumet et al. (2016)
NIDM-Results in brief

Design of the experiment, pre-processed data...

Statistical estimation

Effects, standard error and statistical maps...

Inference

Thresholded maps, list of peaks and clusters...

Maumet et. al, (2016)
Derived data reuse with NIDM-Results: meta-analysis


Maumet et. al, (2016)
Derived data inspection with NIDM-Results: SPM- and FSL-like views

https://github.com/incf-nidash/nidmresults-spmhtml
https://github.com/incf-nidash/nidmresults-fslhtm
Derived data inspection with NIDM-Results: SPM- and FSL-like views
How to use NIDM-Results for my study?

1. Export

`$ nidmfs fsl_ds107_group 49 -g Control`

https://github.com/incf-nidash/nidmresults-spm
https://github.com/incf-nidash/nidmresults-fsl

2. Publication on NeuroVault
One step further: linking it all together!

“Gather, view and monitor all research outcomes”

https://beta.connect.openaire.eu/

Principe (2018)
One step further: linking it all together!

“Gather, view and monitor all research outcomes”

https://beta.connect.openaire.eu/

Principe (2018)
Many ways to contribute

You can...

- Make your research outputs available and citable
- Cite existing software
- Use derived data standards and tools
- Contribute to the development of new specifications and their ecosystems
- Make datasets in neuroimaging data repository discoverable

Join the NIDASH community! [https://github.com/incf-nidash/nidm-specs](https://github.com/incf-nidash/nidm-specs)

References


Photo and template credits

Brains on slide 4: Neil Conway - link
Transparency on slide 29: Public domain (CC0) - link
Presentation template by SlidesCarnival, adapted.
Acknowledgements

**NIDM working group**

**Neuroimaging software teams**
Gang Chen, Richard Reynolds, Robert Cox (AFNI), Mark Jenkinson, Matthew Webster, Paul McCarthy, Eugene Duff, Steve Smith (FSL), Guillaume Flandin (SPM).

**Meta-analysis datasets** Tracey group at FMRIB.

Thank you! To all INCF NIDASH task force members.
Thank you!

@cmaumet
camille.maumet@inria.fr

Join the NIDASH community:
https://github.com/incf-nidash/nidm-specs