

Standardizing neuroimaging: BIDS, aa, NIDM

Tibor Auer, Research Fellow in MRI
Department of Psychology, Royal Holloway University of London



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

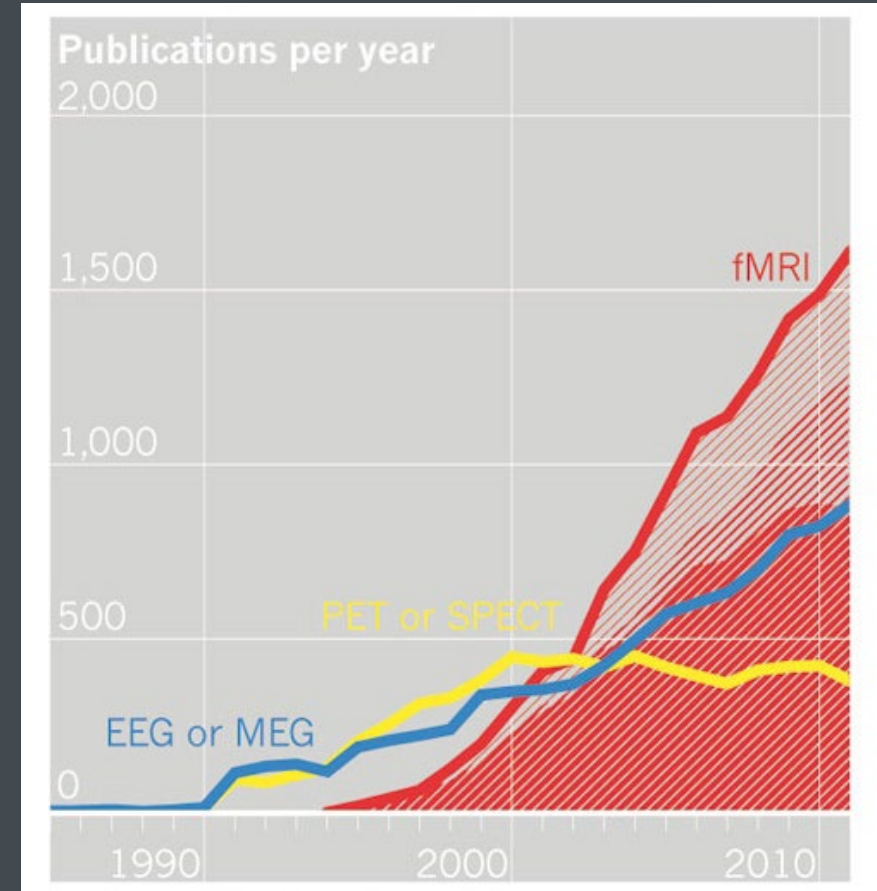
Challenge



- Number of projects is increasing¹
- Cohort sizes are increasing
 - 2008: ~10
 - 2018: hundreds
- Complexity is increasing:
 - Multimodality → integrative view of the brain

↓

 - Requires integration in our practices



<http://www.nature.com/news/brain-imaging-fmri-2-0-1.10365>



- **Integration**

- Experiment: stimulus presentation, response recording, logging
- Data: scripts, multimodal images, behavioural data, log files
- Analysis: different approaches, software, architecture
- Report: transparency, putting into context¹, providing evidence for validity²

- **Issues**

- Demanding
- Difficult to set up and document → Reproducibility and replicability
- Easier to make errors
- Harder to detect errors



- **Global scale:**
 - Harmonizing protocols
 - „Best practices”
 - Long-term planning
- **Local scale**
 - Implementing global solution
 - Building blocks
 - Support

Data



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Data



- **Challenge: How to handle our data?**

- Sharing (even within the same lab) →

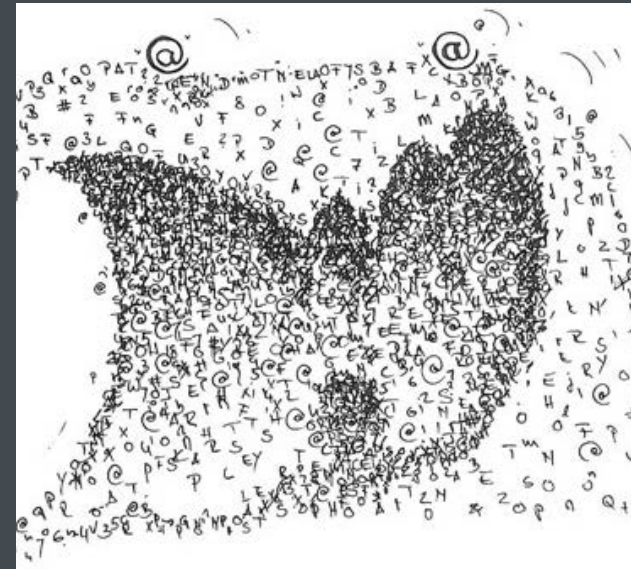
- Rearranging data

- Processing →

- Rearranging data
 - Rewriting scripts
 - Entering tedious manual inputs¹

- Validating →

- Inaccuracy
 - Incompleteness



<http://sonian.com/shut-down-the-email-monster-with-hosted-email-archiving>



- **BIDS: Brain Imaging Data Structure** – <http://bids.neuroimaging.io>
 - Standard for organizing data of a human neuroimaging experiment.
 - Developed by a community (<https://github.com/bids-standard>):
 - Supported and recommended by the INCF



- **Solution for**
 - PIs: More than one person work on the same data over time
 - Users: Software aware of the data structure → less manual entry
 - Databases: Easier to share/exchange data
 - Already accept BIDS: [OpenNEURO](#), COINS, LORIS , SciTran, XNAT
 - Most funders and some journals require data sharing
 - Validator tool



- **Principles**

- Some metadata is better than no metadata
- Don't rely on external software or complicated file formats
- Aim to intuitively capture most of experiments, but give space to extend the standard

BIDS

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/

- **Some metadata is better than no metadata**



- Folder structure
- Filename



- **Some metadata is better than no metadata**

- JSON files¹ for key–value pairs

```
{  
  
  "RepetitionTime": 3.0,  
  
  "EchoTime": 0.03,  
  
  "FlipAngle": 78,  
  
  "SliceTiming": [0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4,  
                  1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8],  
  
  "InPlanePhaseEncodingDirection": "AP"  
  
  "TaskName": "nback"  
  
}
```

- **Don't rely on external software or complicated file formats**
 - JSON files for key–value pairs
 - Use of compressed NIFTI files for imaging data. 
 - Use of tab separated files for tabular data (demographics, events). 

```
participant_id age sex
sub-001        34  M
Sub-002        12  F
Sub-003        33  F
```

```
onset duration trial_type ResponseTime
1.2    0.6      go          1.435
5.6    0.6      stop         1.739
...
```

- Use of legacy text file formats for b vectors/values



- **Aim to intuitively capture most of experiments...**
 - Supports most types of data, which are common in the field
 - Make certain folder hierarchy levels optional for simplicity.
 - Supports behavioural variables on any level (subjects, sessions and runs).
 - Supports contiguous acquisition covariates (breathing, cardiac etc.)
- **...but give space to extend the standard**
 - Allows for arbitrary files not covered by the spec to be included.
 - BIDS Extension Proposals

Analysis



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

Challenge



- **Increasingly large cohort sizes**
 - 2004: <10
 - 2014: hundreds
 - **Multimodality: fMRI, DWI, anatomy(T1, T2), MTI**
 - Offers a more integrated view of the brain
 - Requires integration of different methods
- ↓
- **Issues:**
 - Difficult integration of methods
 - Difficult documentation → Reproducibility
 - Increased risk of human error
 - Harder to detect errors
- } Transparency



Solution – Automatic analysis (aa)



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- **Description**

- A pipeline system for neuroimaging written in Matlab
 - MRI: structural, fMRI, DTI/DKI, MTI
 - MEG/EEG
- Supports SPM **12**¹ and some functions from FSL and Freesurfer, several toolboxes, etc.
- Proprietary code from contributors and external scientists
- <http://automaticanalysis.org>
- <https://github.com/automaticanalysis/automaticanalysis/blob/master/README.md>

- **Availability**

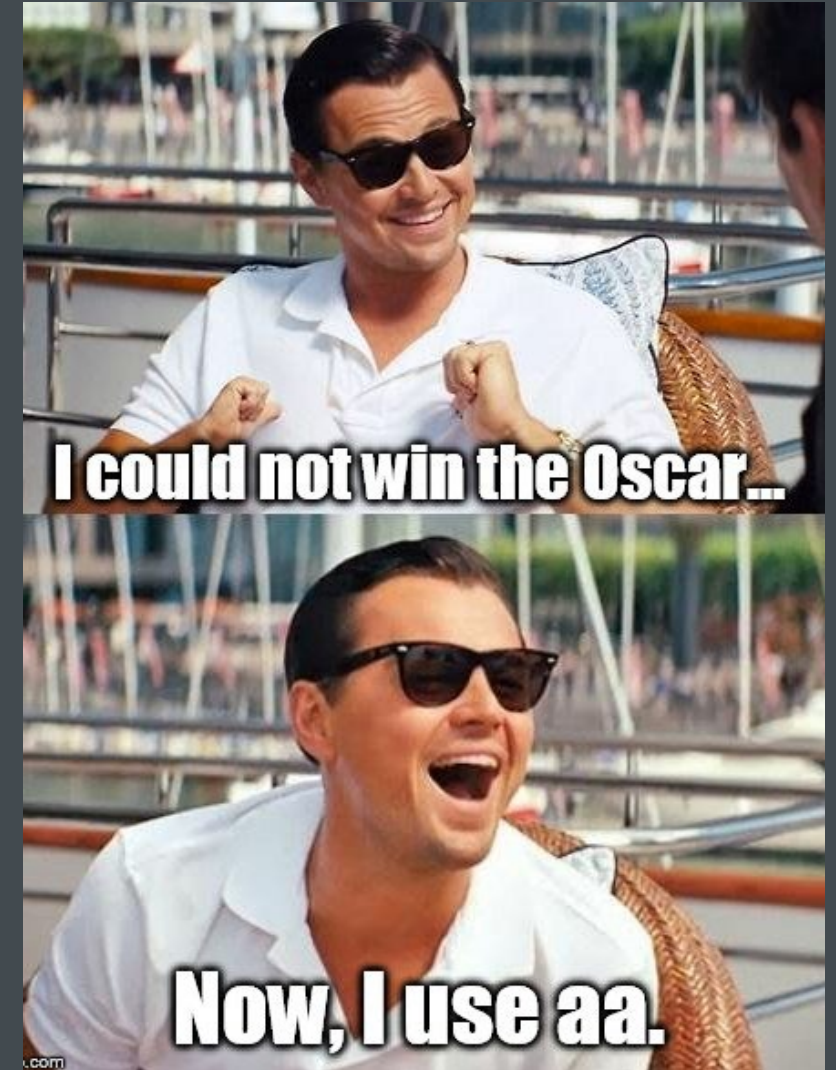
- [GitHub](#)

Solution – Automatic analysis (aa)



- **Properties**

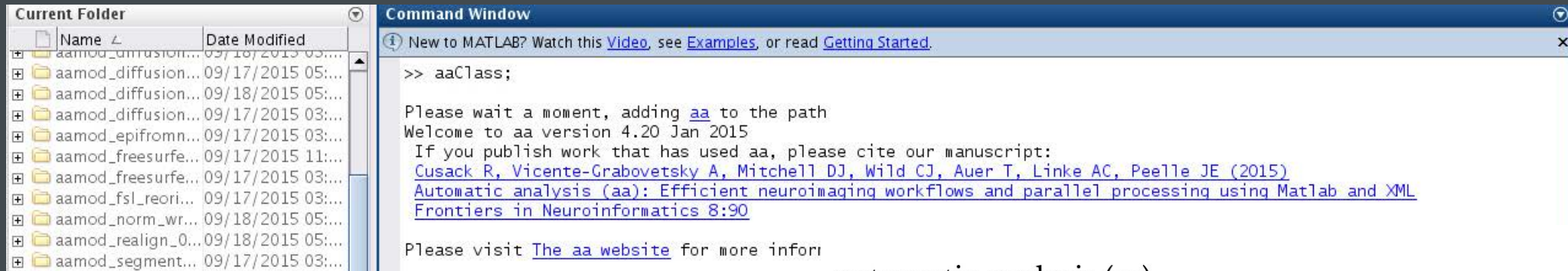
- High-level based on standardised recipes
 - Automatic, transparent, replicable
 - Capture provenance
 - Code recycling/sharing/publishing
- Tracks processes → Restartable
- Notifies via e-mail
- Record keeping → Diagnostics¹
- NiFTI-4D support → “Economic”





- Website: <http://automaticanalysis.org>
- GitHub: <https://github.com/automaticanalysis/automaticanalysis/blob/master/README.md>
- MRC-CBSU Wiki: <http://imaging.mrc-cbu.cam.ac.uk/imaging/AA>

- Running – aa intro



TECHNOLOGY REPORT ARTICLE

Front. Neuroinform., 15 January 2015 | <http://dx.doi.org/10.3389/fninf.2014.00090>

Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML

Rhodri Cusack¹, Alejandro Vicente-Grabovetsky², Daniel J. Mitchell³, Conor J. Wild¹, Tibor Auer³, Annika C. Linke¹ and Jonathan E. Peelle⁴

¹Brain and Mind Institute, Western University, London, ON, Canada

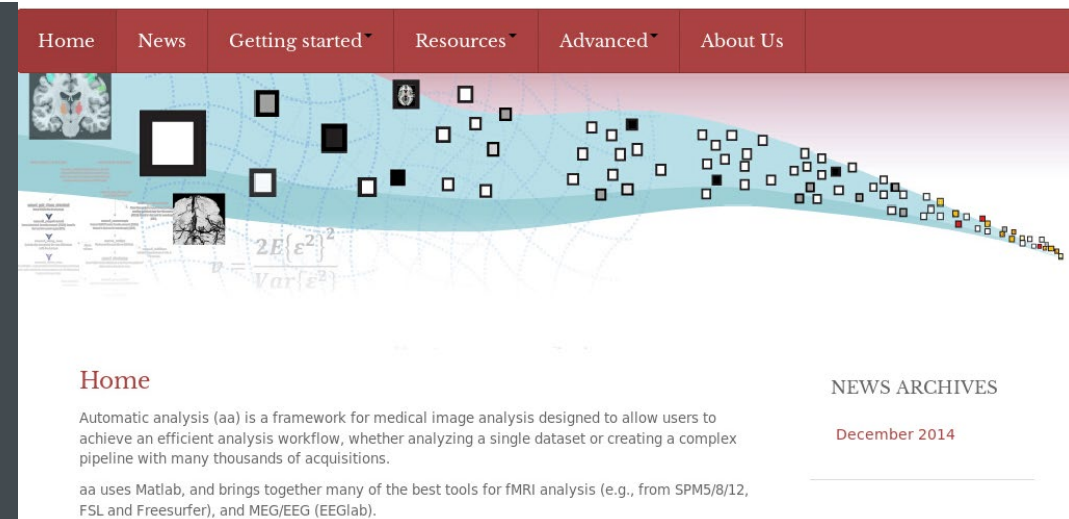
²Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

³MRC Cognition and Brain Sciences Unit, Cambridge, UK

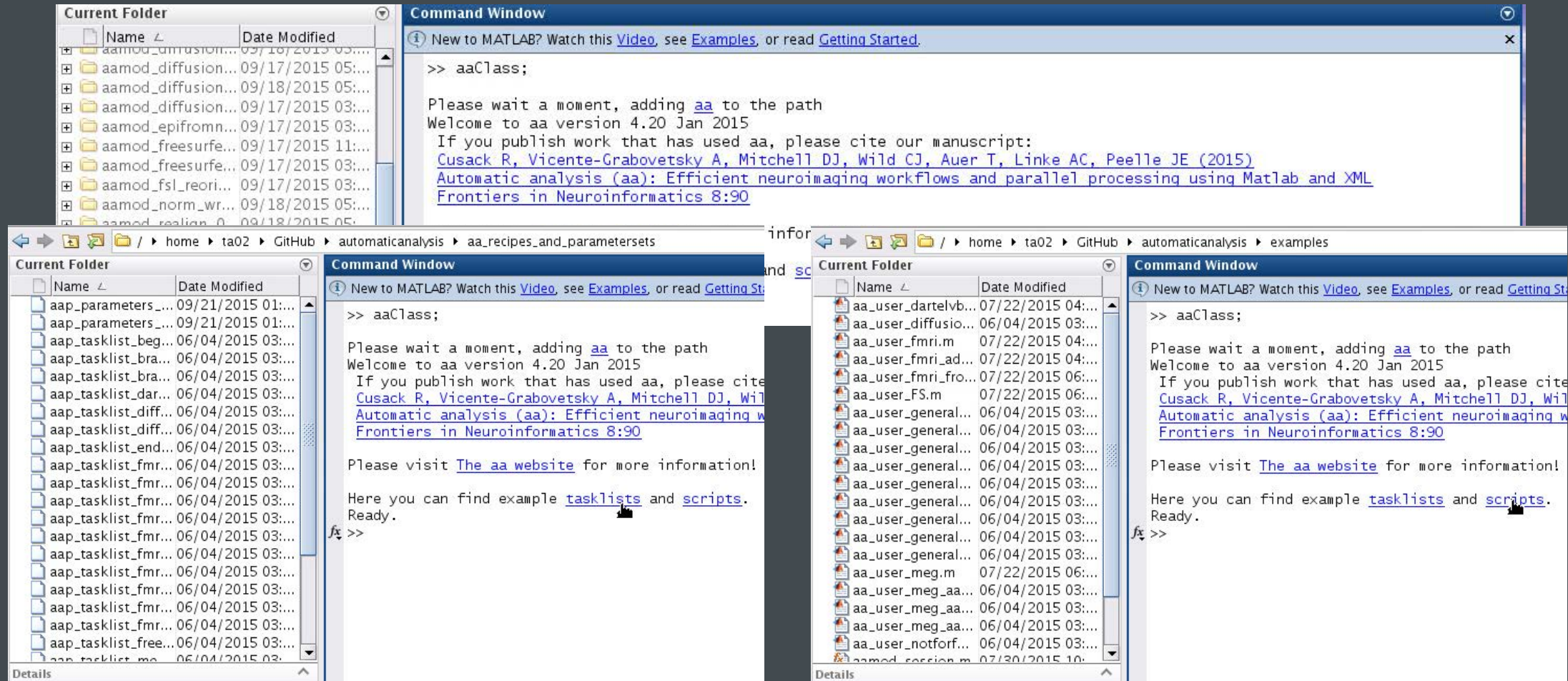
⁴Department of Otolaryngology, Washington University in St. Louis, St. Louis, MO, USA

and [Sci](#) automatic analysis (aa)

Efficient workflows for medical imaging



- Running – aa intro



The image displays three screenshots of the MATLAB Command Window, illustrating the initial setup and welcome message for the 'aa' (Automatic Analysis) package. Each window shows the 'Current Folder' pane on the left and the 'Command Window' on the right.

Top Screenshot: The Command Window shows the command `>> aaClass;` being executed. The output includes a welcome message for 'aa version 4.20 Jan 2015' and a citation request: 'If you publish work that has used aa, please cite our manuscript: Cusack R, Vicente-Grabovetsky A, Mitchell DJ, Wild CJ, Auer T, Linke AC, Peelle JE (2015) Automatic analysis (aa): Efficient neuroimaging workflows and parallel processing using Matlab and XML Frontiers in Neuroinformatics 8:90'.

Bottom Left Screenshot: The Command Window shows the same command `>> aaClass;` being executed. The output includes the same welcome message and citation request, followed by a prompt to visit the 'aa website' for more information and a link to example 'tasklists' and 'scripts'.

Bottom Right Screenshot: The Command Window shows the same command `>> aaClass;` being executed. The output includes the same welcome message and citation request, followed by a prompt to visit the 'aa website' for more information and a link to example 'tasklists' and 'scripts'.

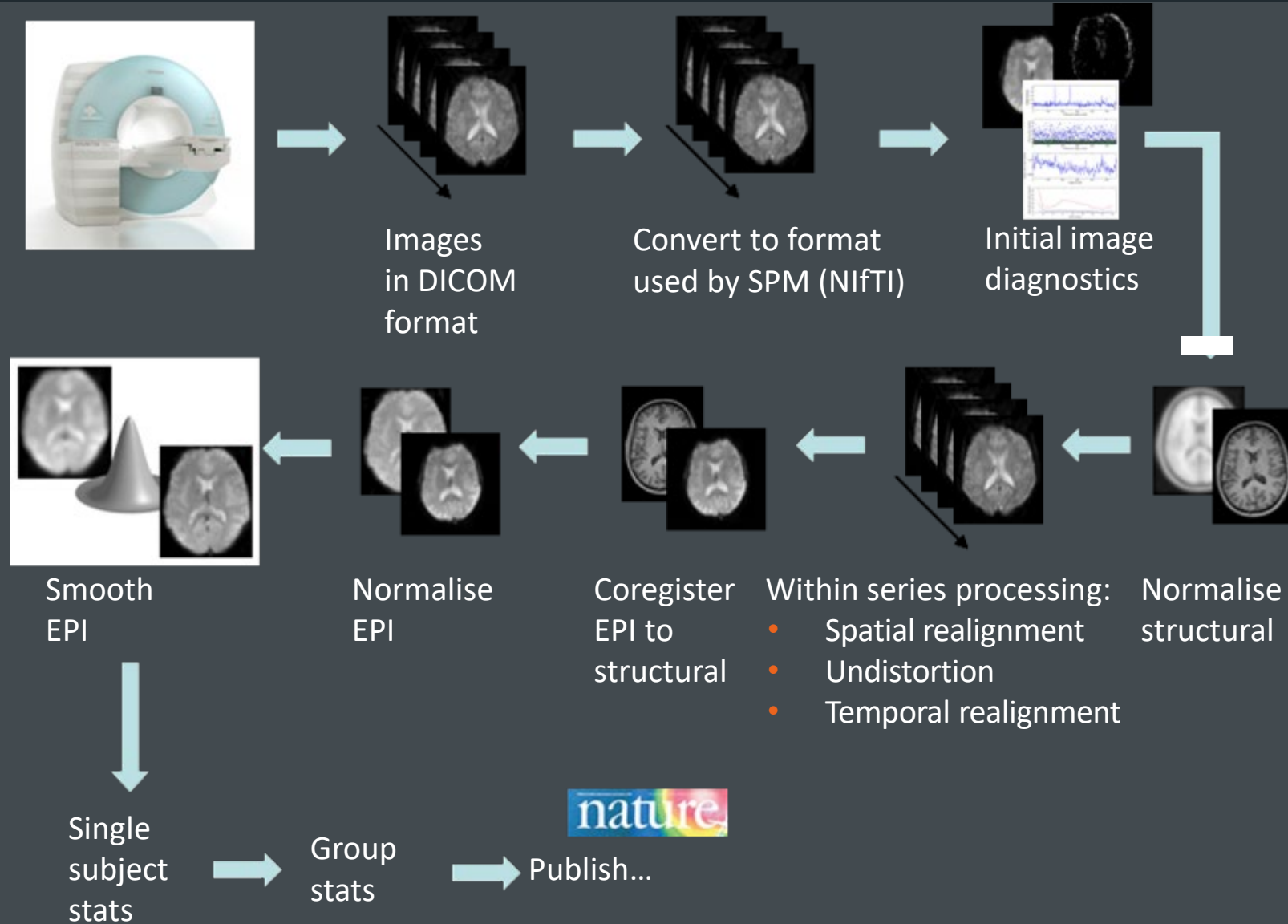


- **High-level scripting** (examples also bundled)
 - **Tasklist**: pipelines describing a series of modules to be executed
 - Easy reading
 - Easy reordering
 - Branching
 - **User Master Script**: specifies the analysis⁰
 - Loads in: default parameters and the tasklist
 - Customises: parameters and tasks
 - Specifies¹: data and model
 - Generates³: report
 - Cleans up⁴: garbage (from data encapsulation)

Automatic analysis (aa)

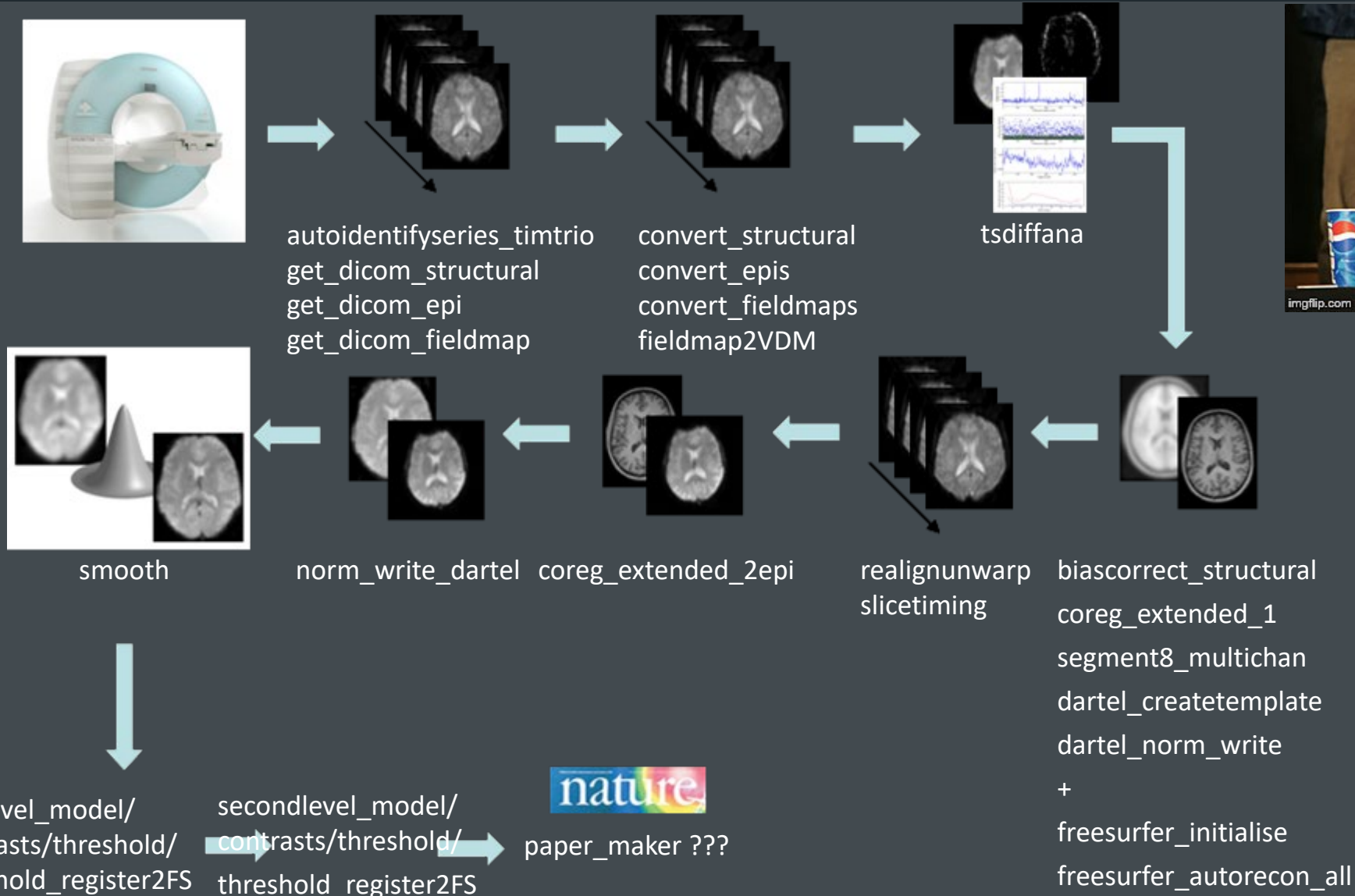


aa-natomy



Automatic analysis (aa)

aa-natomy



Automatic analysis (aa)

aa-natomy



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- Modules

- [Header](#)¹ – Data encapsulation
 - Defines inputs and outputs (streams)
 - Defines domain (i.e. once per study/subject/session/scan)
 - Set parameter defaults²
 - Ensures independence → Parallel processing on cluster or cloud³
- [Body](#)⁴ – Code encapsulation
 - Low overhead, Expandable



Automatic analysis (aa)

aa-natomy



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- Data streams¹

- Modules explicitly define their inputs and outputs and default parameters.

e.g. (fragment from [aamod_realignunwarp.xml](#))

```
<inputstreams>
```

```
<stream>epi</stream>
```

```
<stream>fieldmap</stream>
```

```
</inputstreams>
```

```
<outputstreams>
```

```
<stream>realignment_parameter</stream>
```

```
<stream>epi</stream>
```

```
<stream>meanepi</stream>
```

```
</outputstreams>
```

→ Takes:

→ a set of EPI volumes

→ and a fieldmap;

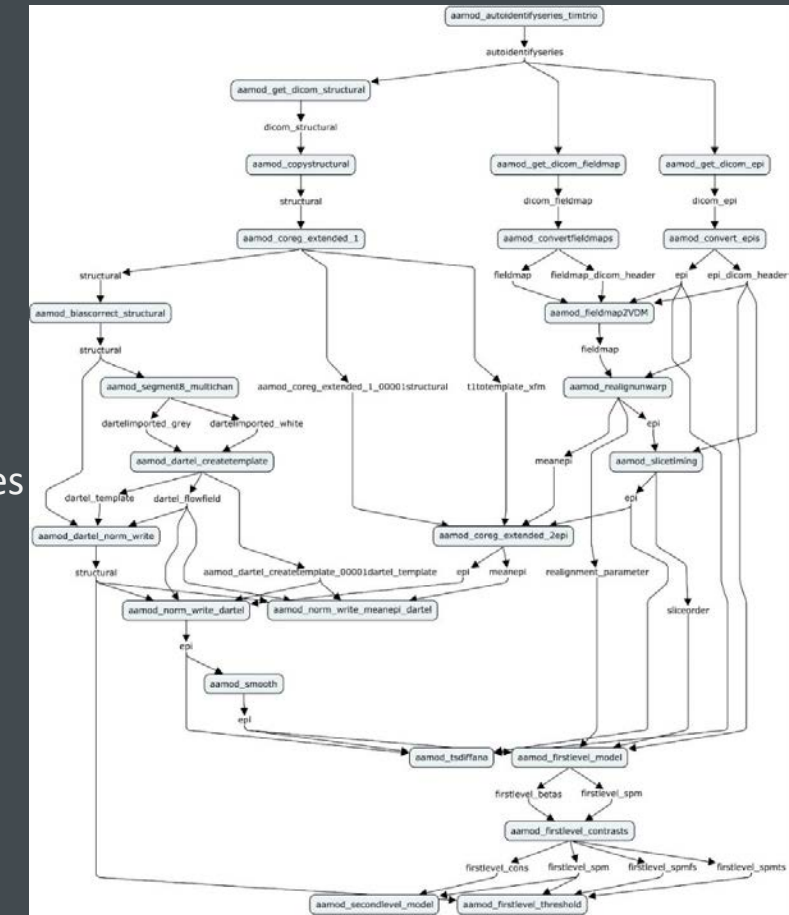
→ Produces:

→ realignment parameters,

→ another set of EPI volumes

→ and a mean EPI volume

- Easy reordering of the modules without worrying for prefixes
- Provenance (flow of data) → parallel computing, report generating





- **Diagnostics**

- Between-subject summaries with descriptive stats to identify outliers
 - Motion correction
 - Registration (Normalisation)
 - First-level activations
- Within-subject summary to localize the erroneous stage



report_moco.mht



report_reg.mht



report_C01.mht



report_S01.mht

Automatic analysis (aa)

Extra features



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- “Large Study” features:
 - NiFTI-4D format also for SPM-based modules (data maintenance)
 - Site-/user-/study-specific configuration defaults
 - Multiple raw data (DICOM) sources
 - Pipeline connection: direct aa streams from a common pipeline
 - Can select only a subset of subjects and sessions
 - Keeps dependency: pipeline aware of source changes
 - E.g.:
 - Multimodal study: separate pipelines for each modality
 - Complex study: common preprocessing pipeline + multiple models



- Brain Imaging Data Structure (BIDS) as input:

% Add data

```
aap.directory_conventions.rawdatadir = '<path to BIDS dataset>';
```

```
aap = aas_processBIDS(aap);
```



- For: structural (t1w, t2w), functional, fieldmap, diffusion
 - Adds: sessions (within-subject or separate), subjects, events (all, not BIDS-Model)
-
- Cave:
 - Tasklist must be set accordingly

Automatic analysis (aa)

BIDS-aware



- **BIDS-App:** <https://github.com/BIDS-Apps/aa> and <https://hub.docker.com/r/bids/aa>
- Includes: MATLAB run-time, FSL, FreeSurfer
- **Usage:**
 - `run <bids_dir> <output_dir> {participant|group}`
`[--participant_label <participant_label>]`

`[--freesurfer_license <license_file>]`
`[--connection <pipeline to connect to>]`
`[<tasklist> <user_customisation>]`
- **Arguments**
 - BIDS-specific
 - aa-specific
 - Connection to a previously processed pipeline
 - [Tasklist](#)
 - [User customisation](#)

Automatic analysis (aa)

BIDS-aware



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON

- **BIDS-App:** <https://github.com/BIDS-Apps/aa> and <https://hub.docker.com/r/bids/aa>
- Includes: MATLAB run-time, FSL, FreeSurfer
- E.g.:
 - `docker pull bids/aa`
 - `docker run -ti --rm --read-only -v /tmp:/tmp -v /var/tmp:/var/tmp \\
-v ${HOME}/Download:/download -v {HOME}/Download/ds114_test1:/bids_dataset -v {HOME}/Download/aa:/outputs \\
bids/aa \\
/bids_dataset /outputs participant --participant_label 01 \\
--freesurfer_license /download/license.txt \\
/download/BIDS114_tasklist.xml /download/BIDS114_test1_aa.xml`

Provenance



ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON



- **Challenge**

- Data/**Methods**/Results are hard to find, access and use
- Metadata are often missing
- Provenance is vaguely described

Lack of interoperability
Lack of reproducibility



- **NIDM: Neuroimaging Data Model** – <http://nidm.nidash.org>

- ⁰Collection of specification documents describing human neuroimaging
- INCF Neuroimaging Data Sharing Task Force (NIDASH-TF)



- **Contributors (part)**

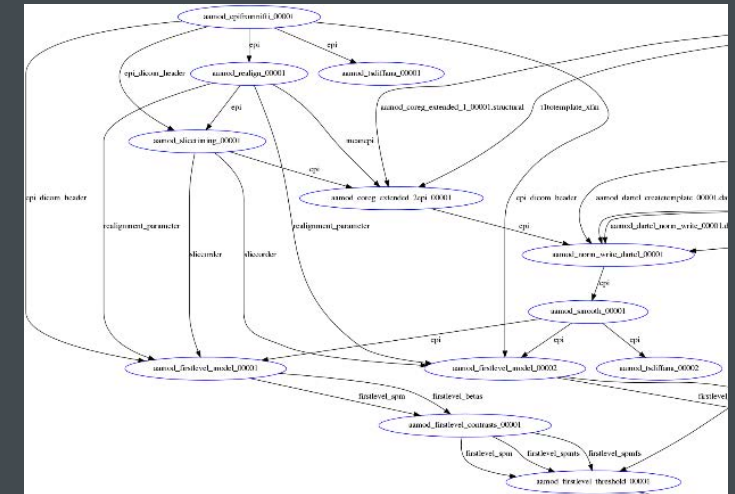
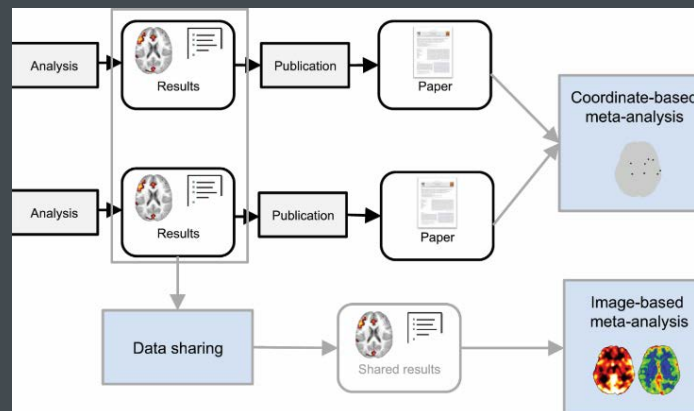
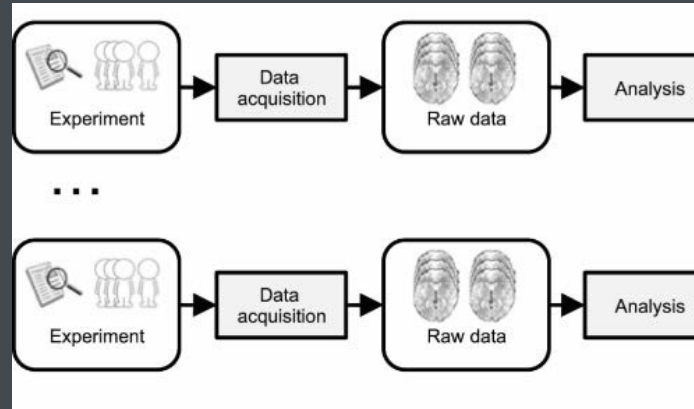
- **Tibor Auer**, Gully Burns, Samir Das, Fariba Fana, Guillaume Flandin, Satrajit Ghosh, Tristan Glatard, Chris Gorgolewski, Karl Helmer, David Keator, Camille Maumet, Nolan Nichols, Thomas Nichols, Jean-Baptiste Poline, Vanessa Sochat, Jason Steffener, Jessica Turner
- Cameron Craddock, Stephan Gerhard, Yaroslav Halchenko, Michael Hanke, Christian Haselgrove, David Kennedy, Arno Klein, Daniel Marcus, Franck Michel, Simon Milton, Russell Poldrack, Rich Stoner

Provenance



NIDM

- NIDM-Experiment
 - **Validate** data
 - Analyse data with **minimum input**
- NIDM-Workflow
 - **Apply workflow** with minimum input
- NIDM-Results
 - **Image-based** meta-analysis



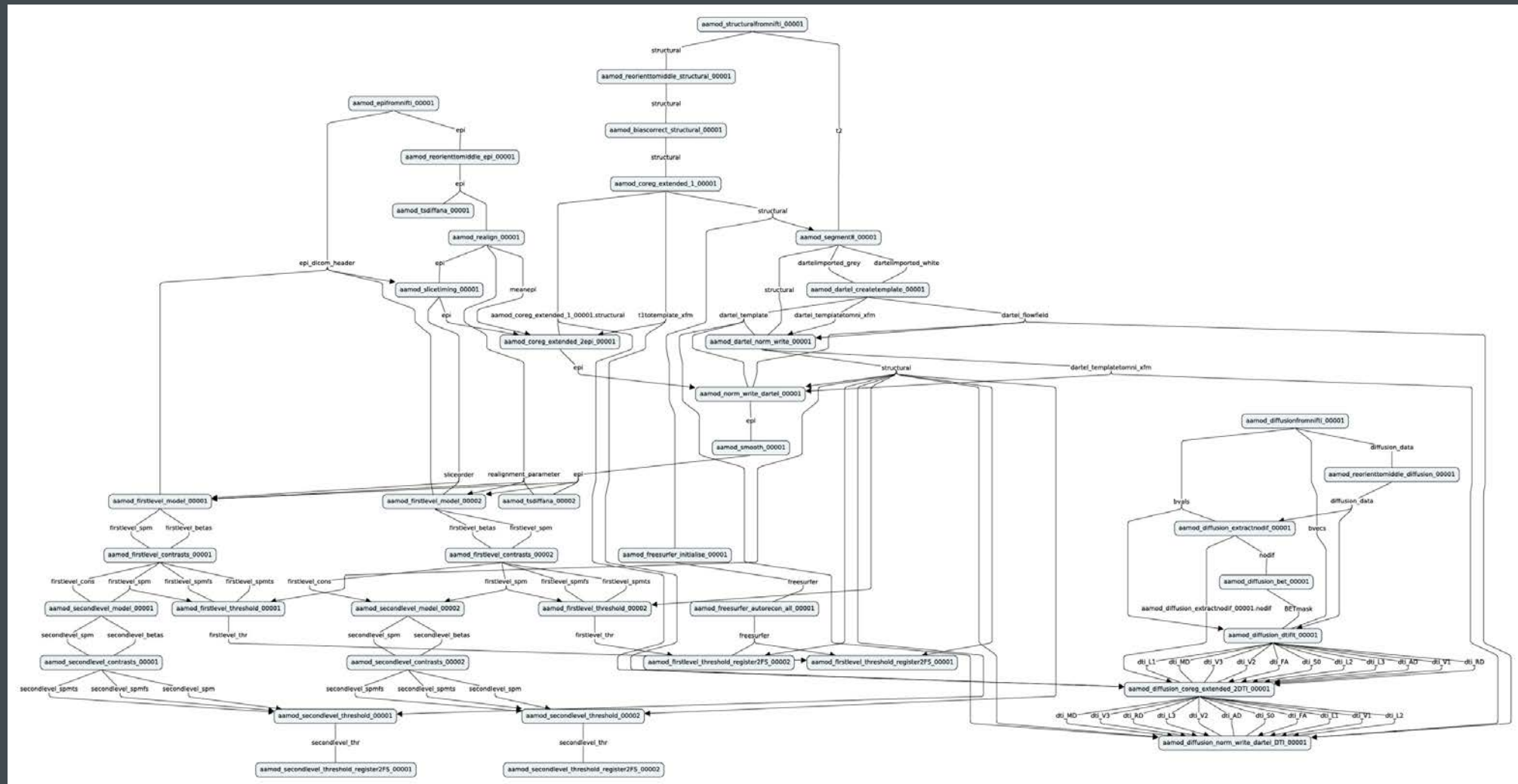
Provenance



**ROYAL
HOLLOWAY
UNIVERSITY
OF LONDON**

Automatic analysis (aa)

- **Workflow – CmapTools:**

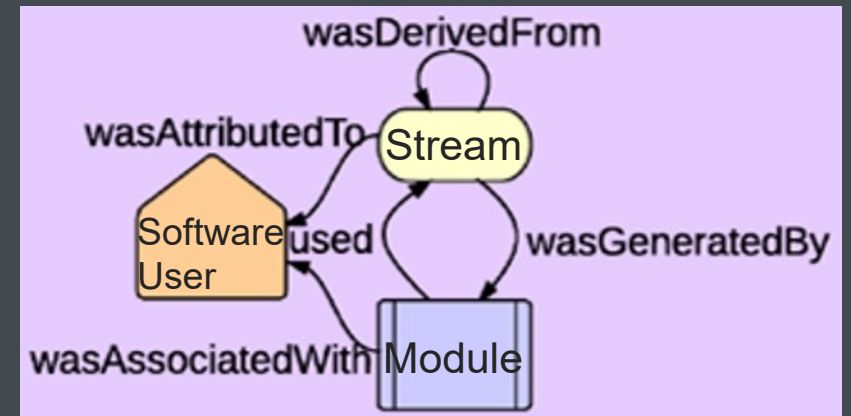


Automatic analysis (aa)

- **Workflow – NIDM-W – ProvONE:**

- Using: [spm provenance](#) by Guillaume Flandin
- aa-concept ~ PROV concept

| aa | NIDM-W/ProveONE | Prov |
|---------------------------------|----------------------|----------|
| Software User | id. | Agent |
| Module | ProcessExec | Activity |
| Stream | Data/Collection | Entity |
| Tasklist | ProcessSpecification | Plan |
| Toolboxes Hardware resources | Requirements | |





- **Cloud-based platform**
 - Data and analysis: <https://openneuro.org>
 - Results: <https://neurovault.org>
- **Local platform**
 - <http://www.reproducibleimaging.org/index.html> (in progress)
- **Solution**
 - Data: <https://www.datalad.org> – git-based data management
 - Provenance: <https://www.reprozip.org>

- Increasing awareness globally and locally
- Growing number of initiatives, platforms
 - [UK Network of Open Science Working Groups \(UK-OSWG\)](#)
 - [OHBM Open Science Special Interest Group](#)
 - [INCF SIG on Reproducible Research and Open Neuroscience](#)
- Growing number of solutions → Fragmentation vs. Harmonisation
- Need to identify and tackle barriers
 - Education, Support
 - Career building

Open Science Survey

