

The Amsterdam Open MRI Collection (AOMIC): A Collection of Publicly Available Population Imaging Datasets.

Lukas Snoek (L.Snoek@uva.nl)

Department of Brain & Cognition, University of Amsterdam, The Netherlands

Maite van der Miesen (maitemvdm@gmail.com)

Brain & Cognitive Sciences Research Master, University of Amsterdam, The Netherlands

Andries van der Leij (A.R.vanderleij@gmail.com)

University of Amsterdam, The Netherlands

BrainsFirst, Amsterdam, The Netherlands

Tinka Beemsterboer (T.Beemsterboer@uva.nl)

LAB Neuroimaging Centre, University of Amsterdam, The Netherlands

H. Steven Scholte (H.S.Scholte@uva.nl)

Department of Brain & Cognition, University of Amsterdam, The Netherlands

Abstract

We present the Amsterdam Open MRI Collection (AOMIC): three datasets with multimodal (3T) MRI data including structural (T1-weighted), diffusion-weighted, resting-state, and task-based functional (BOLD) MRI scans, as well as detailed demographic and individual difference traits (including age, intelligence, and personality scores), from a large set of healthy participants ($N = 933$, $N = 224$, and $N = 238$). All data will be made freely available on the Openneuro data sharing platform. Raw data were anonymized and converted to a standardized format (BIDS) and underwent extensive (automated and manual) quality control. Additionally, the datasets include several derivatives, including quality control reports and metrics, preprocessed (anatomical and functional MRI) data, and preprocessed physiology data (cardiac and respiratory traces). Notably, task-based fMRI was collected during various robust paradigms (targeting cognitive conflict, emotion recognition, working memory, face perception, cognitive control, and response inhibition) for which extensively annotated event-files are available. In addition to the raw data, all code that was used to convert, transform, and (pre)process the data is available online.

Keywords: MRI; data sharing; openneuro.org

Introducing AOMIC

It is becoming increasingly clear that robust effects in neuroimaging studies require very large sample sizes (Button et al., 2013; Yarkoni, 2009), especially when investigating between-subject effects (Dubois & Dolphs, 2016). With this in mind, we have run several large-scale MRI projects over the past decade at the University of Amsterdam. We believe that, at this moment, sharing the data from these projects will benefit the neuroimaging community most. To this end, we present the Amsterdam Open MRI Collection (AOMIC): three large-scale datasets with high-quality, multimodal MRI data

and detailed metadata, which will be made available on the Openneuro data sharing platform¹.

In what follows, we will describe the format and contents of AOMIC, as well as the procedure for preparing the data for release to the public (including quality control and anonymization procedure).

Data curation and format

AOMIC contains three datasets, which we will refer to as "ID1000", "PIOP1", and "PIOP2". All three datasets, scanned on the Philips Achieva 3T scanner at the LAB neuroimaging centre of the University of Amsterdam, contain raw multimodal (i.e., anatomical, diffusion, and BOLD) MRI data and a variety of behavioral, psychometric, and demographic information (see 1). The datasets are formatted according to "Brain Imaging Data Structure" (BIDS; Gorgolewski et al., 2016) and includes detailed metadata about the scanning parameters, experimental tasks, and behavioral/psychometric data. Extensive measures have been taken to ensure participant anonymity, including "defacing" to remove facial characteristics, and data quality, including visual quality control in order to detect and exclude data with reconstruction errors or other artifacts. After exclusion of low-quality data, the datasets contain data of 933 (ID1000), 224 (PIOP1), and 238 (PIOP2) participants. Note that these sample sizes reflect the number of participants for which *some* data is available; the number of participants with *complete* (behavioral and/or MRI) data is lower.

In addition to the raw data, the datasets contain several "derivatives", including outputs from a state-of-the-art quality control pipeline (MRIQC, v0.15.0; Esteban et al., 2017) and preprocessing pipeline (fMRIPrep, v1.3.2; Esteban et al., 2019), as well as RETROICOR regressors derived from the physiology data (using the PhysIO toolbox; Kasper et al., 2017). Derivatives from the preprocessing pipeline, specifi-

¹<https://openneuro.org>

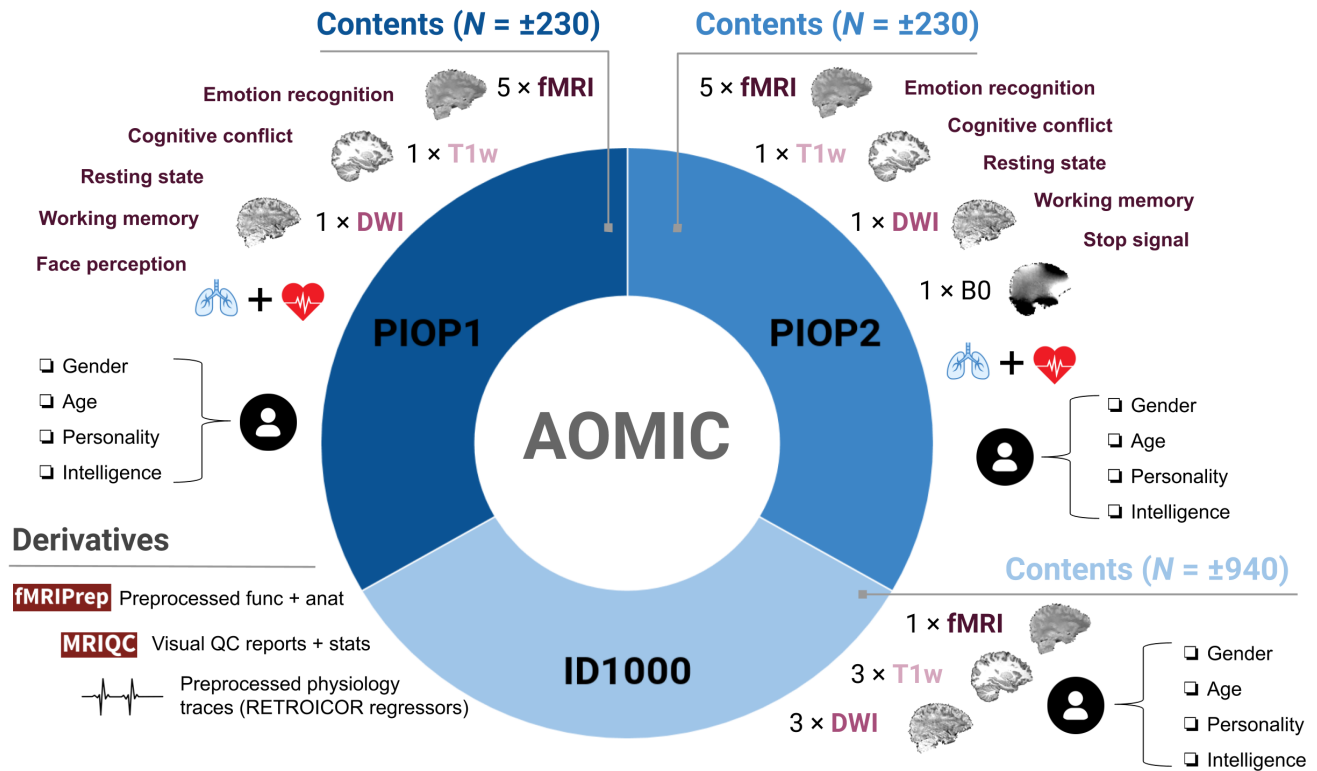


Figure 1: Overview of AOMIC contents. Note that ID1000 contains mostly structural and diffusion MRI data, while PIOP1 and PIOP2 contain mostly functional MRI data, acquired during different experimental tasks.

cally, include Freesurfer surface reconstructions (Fischl, 2012) and preprocessed fMRI and T1-weighted images.

In the next sections, we will describe the contents of the three datasets in more detail.

The ID1000 dataset

ID1000, named as such because the project aimed to investigate Individual Differences across 1000 participants, is a dataset containing structural, diffusion, and functional (BOLD) MRI, as well as concurrently acquired physiological data (respiratory/cardiac traces). Unlike the name suggests, the dataset contains data from 933 subjects after discarding data with artifacts or otherwise corrupted data, who were scanned between 2009 and 2011. Participants were selected to reflect a random sample of the (healthy) Dutch population, encompassing a wide range of ages (18-40) and educational level. Each participant took part in a single session of four hours, which contained three hours of behavioral tests and filling in questionnaires and one hour of MRI acquisition.

The MRI acquisition consisted of three T1-weighted anatomical scans with identical scan parameters (MPRAGE, 1 mm³ spatial resolution), three diffusion-weighted scans (2 mm³ spatial resolution, consisting of a single spin-echo B=0 scan plus B=1000 diffusion-weighted volumes in 32 directions), and one functional MRI (fMRI) scan (GRE-EPI,

3.3 × 3 × 3 mm spatial resolution, TR=2200 ms, TE=28 ms) during which participants watched the same movie clip (*Koyaanisqatsi: Life Out of Balance*, 1983) for 11 minutes. Physiological measurements (cardiac and respiratory traces) were acquired during fMRI acquisition.

In the behavioral part of the session, participants filled in questionnaires aimed to measure personality, intelligence, and various demographics (age, gender, BMI).

The PIOP1 and PIOP2 datasets

Between 2015 and 2017, we set up two “Population Imaging of Psychology” (PIOP) projects, aimed at generating two high-powered datasets for neuroimaging research on individual differences. Similar to the ID1000 project, the neuroimaging centre organized and collected MRI and behavioral/psychometric data from a large set of participants. As opposed to the ID1000 dataset, which imaging data consists mostly of structural and diffusion MRI data, the PIOP datasets contain mostly functional MRI data (acquired during different experimental paradigms). The particular set of MRI scans (and associated paradigms for the functional scans) were chosen such that it would accommodate a wide variety of individual difference studies. Each participant took part in a single session of four hours, which contained three hours of behavioral tests and filling in questionnaires and one hour of MRI acquisition.

The MRI acquisition consisted of one T1-weighted anatomical scan (MPRAGE, 1 mm³ spatial resolution), one diffusion-weighted scan (2 mm³ spatial resolution, consisting of a single spin-echo B=0 scan plus B=1000 diffusion-weighted volumes in 32 directions), and six (PIOP1) or five (PIOP2) functional MRI (fMRI) scans. All functional scans were acquired using the same parameters (3 × 3 × 3.3 spatial resolution, TR=2000ms, TE=28 ms), except for two multiband-accelerated scans in PIOP1 (the resting-state and face-perception scans; multiband factor 3, 3 × 3 × 3.3 mm spatial resolution, TR=750 ms, TE=28 ms). Below, we outline the different paradigms used during fMRI acquisition in the PIOP1 and PIOP2 projects.

Working memory (PIOP1+2) The “working memory” task was adapted from the task used in Pessoa, Gutierrez, Bandettini, and Ungerleider (2002). Participants were shown an visual display with eight orientated bars, which after a delay period of 4 or 6 seconds was shown again with either an identical visual display or one with one bar changed in orientation. Participants had to report whether the display changed or not. In addition to these “working memory” trials, the task included 8 control trials (in which there was no visual display, but participants had to respond with a random button press) and 18 null trials. Trial order was identical for all participants.

Emotion recognition (PIOP1+2) The “emotion recognition” task was adapted from the task used in Hariri, Bookheimer, and Mazziotta (2000). Participants were briefly presented with a target stimulus (top) and two probe stimuli (bottom) and were instructed to decide which of the probe stimuli (left or right) matched the target stimulus. Stimuli were either faces, in which participants had to match the emotional expression (anger or fear) of the target face (*emotion* condition), or color-matched ovals, in which participants had to match the orientation (horizontal or vertical) of the target oval.

Cognitive conflict (PIOP1+2) The “cognitive conflict” paradigm is a variant of the “gender Stroop” task, which we adapted from (Egner, Ely, & Grinband, 2010). Participants observed faces, which were either male or female, with superimposed words, which were either (the Dutch words for) “woman” or “man”, and were instructed to report the actual sex of the face. Trials were either congruent (sex of face is the same as the superimposed word; $n = 48$) or incongruent (sex of face is different from the superimposed word; $n = 48$).

Resting state (PIOP1+2) During the resting state scans, participants were instructed to keep their gaze fixated on a fixation cross in the middle of the screen with a gray background. The resting state scans lasted 6 minutes (PIOP1) and 8 minutes (PIOP2).

Face perception (PIOP1) In the “face perception” task, participants passively viewed dynamic facial expressions taken from the Amsterdam Facial Expression Set (ADFES; Hawk, Van der Schalk, & Fischer, 2008), which displayed either anger, contempt, happiness, or pride, or no expression (“neu-

tral”).

Stop signal (PIOP2) In the “stop signal” task, participants observed images of male or female faces, and were instructed to press left for female and right for male faces as fast as possible, unless they heard a short tone (i.e, the stop signal) right after the onset of the image, in which case they had to withhold a response.

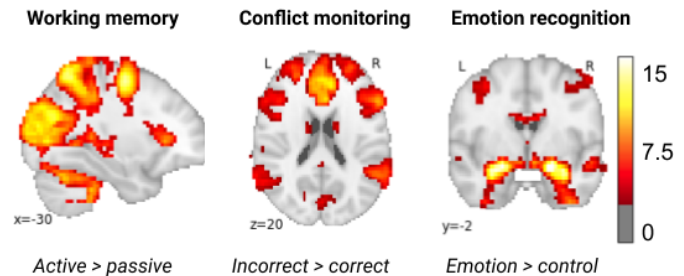


Figure 2: Group-average statistic maps (t-values) for a selection of fMRI tasks. Brain maps were FDR-corrected at the 0.005 level.

How do I get started with AOMIC?

At the moment of writing, we are finalizing the preparation for publication of the data. The raw data and associated derivatives will be published on the Openneuro data sharing platform. To be informed about the release of AOMIC, you can follow the NILAB-UvA Github organization (<https://github.com/NILAB-UvA>), which contains code repositories for the three datasets. Releases of the ID1000, PIOP1, and PIOP2, as well as future datasets acquired by the centre, will be documented here.

Acknowledgments

We thank all research assistants who helped collect the data of the three projects and Jasper Wijnen for advice and guidance with respect to privacy and GDPR-related concerns.

References

- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365.
- Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fmri. *Trends in cognitive sciences*, 20(6), 425–443.
- Egner, T., Ely, S., & Grinband, J. (2010). Going, going, gone: characterizing the time-course of congruency sequence effects. *Frontiers in psychology*, 1, 154.
- Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). Mriqc: Advancing the automatic prediction of image quality in mri from unseen sites. *PloS one*, 12(9), e0184661.

- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... others (2019). Fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1), 111.
- Fischl, B. (2012). Freesurfer. *Neuroimage*, 62(2), 774–781.
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., ... others (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, 160044.
- Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport*, 11(1), 43–48.
- Hawk, S., Van der Schalk, J., & Fischer, A. (2008). Moving faces, looking places: The amsterdam dynamic facial expressions set (adfs). In *12th european conference on facial expressions, geneva, switzerland*.
- Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinze, J., Iglesias, S., ... others (2017). The physio toolbox for modeling physiological noise in fmri data. *Journal of neuroscience methods*, 276, 56–72.
- Koyaanisqatsi: Life out of balance*. (1983). Institute for Regional Education, American Zoetrope.
- Pessoa, L., Gutierrez, E., Bandettini, P. A., & Ungerleider, L. G. (2002). Neural correlates of visual working memory: fmri amplitude predicts task performance. *Neuron*, 35(5), 975–987.
- Yarkoni, T. (2009). Big correlations in little studies: Inflated fmri correlations reflect low statistical powercommentary on vul et al.(2009). *Perspectives on Psychological Science*, 4(3), 294–298.