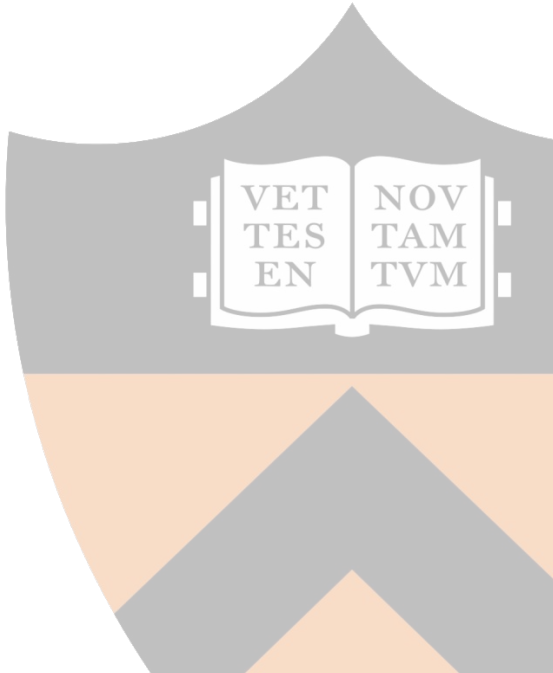
Developing community resources for reproducible neuroimaging

Research Inside-Out
February 25, 2021

Samuel A. Nastase
Princeton Neuroscience Institute
snastase@princeton.edu



@samnastase

Developing community resources for reproducible neuroimaging

Research Inside-Out
February 25, 2021

Samuel A. Nastase
Princeton Neuroscience Institute
snastase@princeton.edu

Uri Hasson

Ken Norman

@samnastase

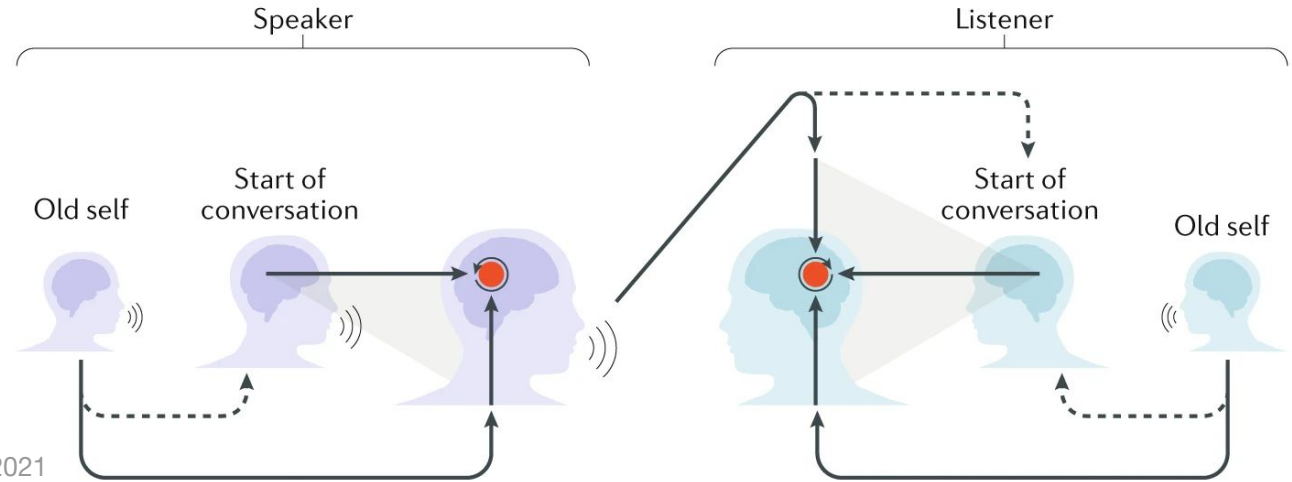
Ethel Franklin Betts (1908)
The Orphant Annie Book, by James Whitcomb Riley

The neurobiological basis of natural language and communication

The problem of language

—language is uniquely human, limiting the use of nonhuman animal models

—the dynamic and contextual nature of language is unamenable to experimental design



Yeshurun et al, *Nat Rev Neurosci*, 2021

The neurobiological basis of natural language and communication

The dominant experimental paradigm

—controlled experimental tasks and manipulations targeting particular linguistic phenomena

—models derived from small, manufactured datasets are difficult to synthesize

OPINION

The cortical organization of speech processing

Gregory Hickok and David Poeppel

Review

A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading

Cathy J. Price

Cerebral Cortex December 2009;19:2767-2796
doi:10.1093/cercor/bhp055
Advance Access publication March 27, 2009

FEATURE ARTICLE
Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies

Jeffrey R. Binder, Rutvik H. Desai, William W. Graves and Lisa L. Conant

Language Imaging Laboratory, Department of Neurology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

Physiol Rev 91: 1357–1392, 2011
doi:10.1152/physrev.00006.2011

THE BRAIN BASIS OF LANGUAGE PROCESSING: FROM STRUCTURE TO FUNCTION

Angela D. Friederici

The neurobiological basis of natural language and communication

The revolution in natural language processing

—neural network models have dramatically transformed natural language processing

—these models are very complex and require very large-volumes of real-world data

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

XLNet: Generalized Autoregressive Pretraining for Language Understanding

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V. Le

RoBERTa: A Robustly Optimized BERT Pretraining Approach

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov

Language Models are Few-Shot Learners

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei

The neurobiological basis of natural language and communication

Our goal:

Create a large-scale open neuroimaging dataset that will allow researchers to evaluate and compare neurobiological models of natural language comprehension across diverse spoken narratives.

Narratives

fMRI data for evaluating models of naturalistic language comprehension

uploaded by Sam Nastase on 2019-12-10 - about 1 year ago

last modified on 2021-01-13 - about 1 month ago

authored by Samuel A. Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J. Honey, Yaara Yeshurun, Mor Regev, Mai Nguyen, Claire H. C. Chang, Christopher Baldassano, Olga Lositsky, Erez Simony, Michael A. Chow, Yuan Chang Leong, Paula P. Brooks, Emily Micciche, Gina Choe, Ariel Goldstein, Tamara Vanderwal, Yaroslav O. Halchenko, Kenneth A. Norman, Uri Hasson

115 18036

OpenNEURO **ds002345**

https://openneuro.org/datasets/ds002345

http://datasets.datalad.org/?dir=/labs/hasson/narratives

OPEN DATA

The *Narratives* data collection

Sharing the "long tail" of neuroscience data

—public, well-curated datasets accelerate research

—re-use of publicly available fMRI datasets saves billions USD in public funding

The "richness" of natural language data

—naturalistic paradigms such as story-listening can accommodate many hypotheses, increasing the re-use value

Ferguson et al., *Nat Neurosci*, 2014
Milham et al., *Nat Commun*, 2018
DuPre et al., *NeuroImage*, 2020
Willems, Nastase et al., *Trends Neurosci*, 2020

The *Narratives* data collection

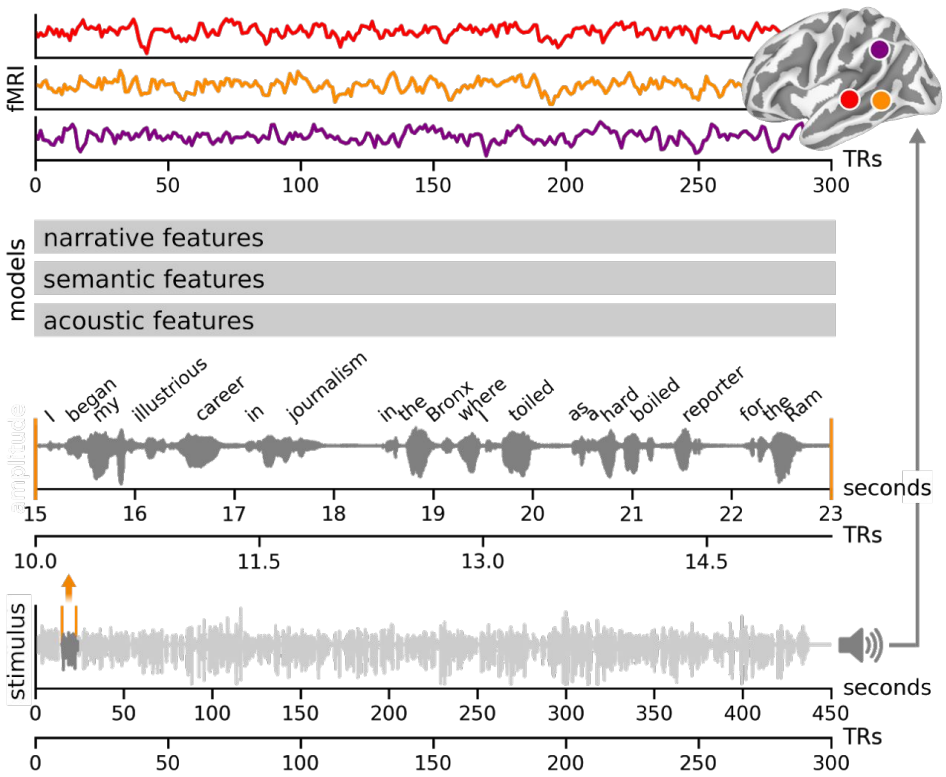
Sharing the “long tail” of neuroscience data

—public, well-curated datasets accelerate research

—re-use of publicly available fMRI datasets saves billions USD in public funding

The “richness” of natural language data

—naturalistic paradigms such as story-listening can accommodate many hypotheses, increasing the re-use value



The *Narratives* data collection

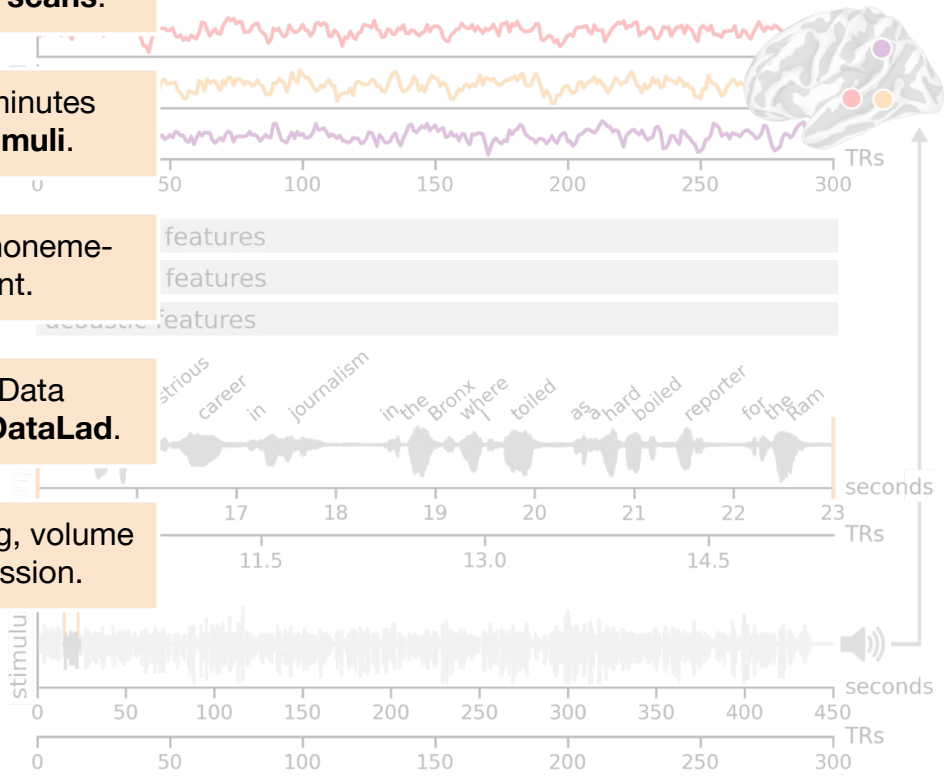
345 unique subjects (ages 18–53 years, mean age 22.2 ± 5.1 years, 204 reported female) participating in **891 functional scans**.

27 spoken story stimuli ranging from ~3 minutes to ~56 minutes (mean ≈ 12 minutes) for a total of **~4.6 hours of unique stimuli**.

Story stimuli are shared with accompanying word- and phoneme-level **time-stamped transcripts** via Gentle forced alignment.

MRI data are standardized according to the Brain Imaging Data Structure (**BIDS**) and the full provenance is tracked using **DataLad**.

Ready-to-use derivatives include fMRIPrep preprocessing, volume and surface normalization, smoothing, and confound regression.



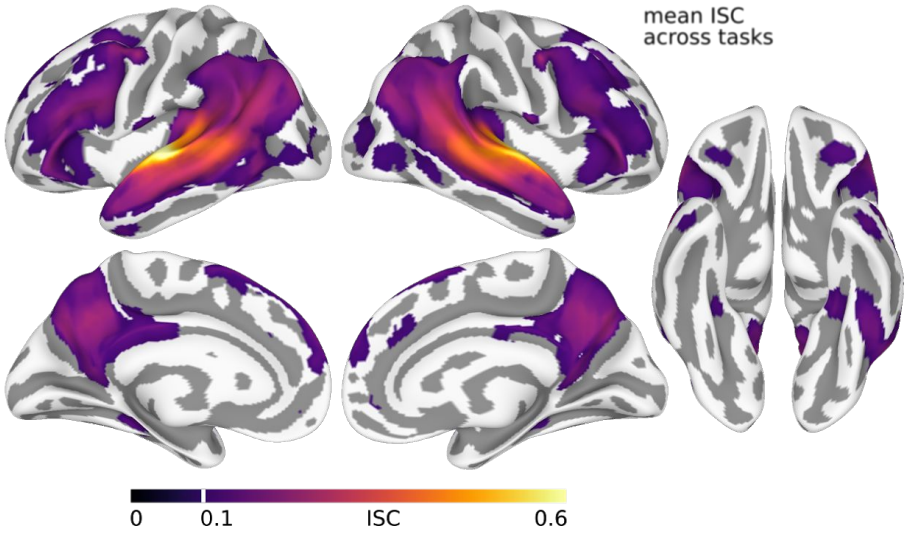
Evaluating the *Narratives* collection

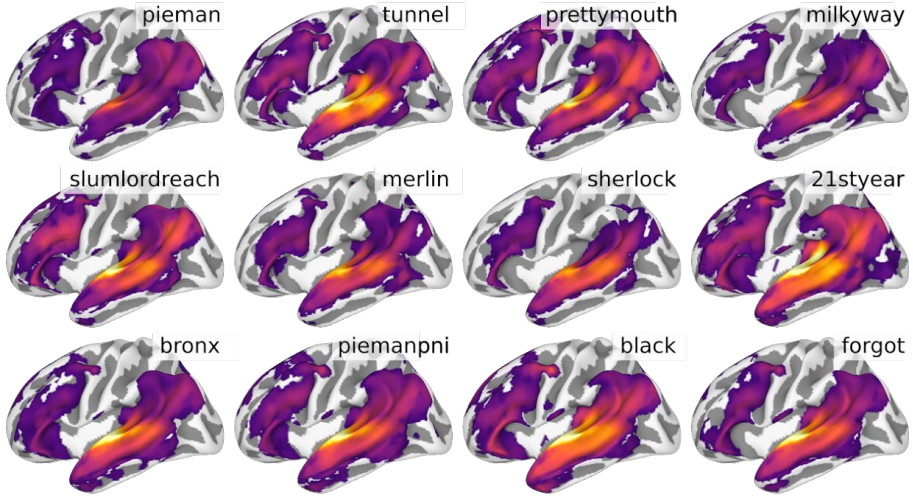
Quality control

—automated quality control using MRIQC

—head motion, spatial smoothness, and tSNR

—intersubject correlation (ISC) analysis





The *Narratives* data collection

Story	Duration	TRs	Words	Subjects
“Pie Man”	07:02	282	957	82
“Tunnel Under the World”	25:34	1,023	3,435	23
“Lucy”	09:02	362	1,607	16
“Pretty Mouth and Green My Eyes”	11:16	451	1,970	40
“Milky Way”	06:44	270	1,058	53
“Slumlord”	15:03	602	2,715	18
“Reach for the Stars One Small Step at a Time”	13:45	550	2,629	18
“It’s Not the Fall That Gets You”	09:07	365	1,601	56
“Merlin”	14:46	591	2,245	36
“Sherlock”	17:32	702	2,681	36
“Schema”	23:12	928	3,788	31
“Shapes”	06:45	270	910	59
“The 21st Year”	55:38	2,226	8,267	25
“Pie Man (PNI)”	06:40	267	992	40
“Running from the Bronx (PNI)”	08:56	358	1,379	40
“I Knew You Were Black”	13:20	534	1,544	40
“The Man Who Forgot Ray Bradbury”	13:57	558	2,135	40
Total:	4.6 hours	11,149 TRs	42,989 words	

The *Narratives* data collection

Story	Duration	TRs	Words	Subjects
“Pie Man”	07:02	282	957	82
“Tunnel Under the World”	25:34	1,023	3,435	23
“Lucy”	09:02	362	1,607	16
“Pretty Mouth and Green My Eyes”	11:16	451	1,970	40
“Milky Way”	06:44	270	1,058	53
“Slumlord”	15:03	602	2,715	18
“Reach for the Stars One Small Step at a Time”	13:45	550	2,629	18
“It’s Not the Fall That Gets You”	09:07	365	1,601	56
“Merlin”	14:46	591	2,245	36
“Sherlock”	17:32	702	2,681	36
“Schema”	23:12	928	3,788	31
“Shapes”	06:45	270	910	59
“The 21st Year”	55:38	2,226	8,267	25
“Pie Man (PNI)”	06:40	267	992	40
“Running from the Bronx (PNI)”	08:56	358	1,379	40
“I Knew You Were Black”	13:20	534	1,544	40
“The Man Who Forgot Ray Bradbury”	13:57	558	2,135	40
Total:	4.6 hours	11,149 TRs	42,989 words	
Total across subjects:	6.4 days	369,496 TRs	1,399,655 words	

The *Narratives* data collection

"Pie Man" by Jim O'Grady

I began my illustrious career in journalism in the Bronx where I toiled as a hard-boiled reporter for the Ram, the student newspaper at Fordham University. And one day I'm walking toward the campus center and out comes the elusive Dean McGowan, architect of a policy to replace Fordham's traditionally working- to middle-class students with wealthier, more prestigious ones. So I whip out my notebook and I go up to him and I say, "Dean McGowan, is it true that Fordham University plans to raise tuition substantially above the inflation rate, and if so, wouldn't that be a betrayal of its mission?" And he stops and looks at me and he says, "Listen up, punk." And right then, there's a blur in the corner of my eye which becomes this figure holding a cream pie which becomes the guy standing next to me mashing a cream pie into Dean McGowan's face. And then runs away. And the Dean is covered with cream. So I give him a moment, and then I say, "Dean McGowan, would you care to comment on this latest attack?" And the Dean says, "Yes, I would care to comment. Fuck you." So I race back to the newsroom with my scoop and I find the editor, Jim Dwyer,

T R AH D IH SH AH N AH L IY

Ochshorn & Hawkins, 2019

CHANGES
README
code
dataset_description.json
derivatives
participants.json
participants.tsv
stimuli
sub-001
sub-002
sub-003
sub-004
sub-005
sub-006
sub-007
sub-008
sub-009
sub-010
sub-011
sub-012
sub-013
sub-014
sub-015
sub-016
sub-017
sub-018
sub-019
sub-020
sub-021
sub-022
sub-023
sub-024
sub-025
sub-026
sub-027
sub-028
sub-029
sub-030
sub-031
sub-032
sub-033
sub-034
sub-035
sub-036
sub-037
sub-038
sub-039
sub-040
sub-041
sub-042
sub-043
sub-044
sub-045
sub-046
sub-047
sub-048
sub-049
sub-050
sub-051
sub-052
sub-053
sub-054
sub-055
sub-056
sub-057
sub-058
sub-059
sub-060
sub-061
sub-062
sub-063
sub-064
sub-065
sub-066
sub-067
sub-068
sub-069
sub-070
sub-071
sub-072
sub-073
sub-074
sub-075
sub-076
sub-077
sub-078
sub-079
sub-080
sub-081
sub-082
sub-083
sub-084
sub-085
sub-086
sub-087
sub-088
sub-089
sub-090
sub-091
sub-092
sub-093
sub-094
sub-095
sub-096
sub-097
sub-098
sub-099
sub-100
sub-101
sub-102
sub-103
sub-104
sub-105
sub-106
sub-107
sub-108
sub-109
sub-110
sub-111
sub-112
sub-113
sub-114
sub-115
sub-116
sub-117
sub-118
sub-119
sub-120
sub-121
sub-122
sub-123
sub-124
sub-125
sub-126
sub-127
sub-128
sub-129
sub-130
sub-131
sub-132
sub-133
sub-134
sub-135
sub-136
sub-137
sub-138
sub-139
sub-140
sub-141
sub-142
sub-143
sub-144
sub-145
sub-146
sub-147
sub-148
sub-149
sub-150
sub-151
sub-152
sub-153
sub-154
sub-155
sub-156
sub-157
sub-158
sub-159
sub-160
sub-161
sub-162
sub-163
sub-164
sub-165
sub-166
sub-167
sub-168
sub-169
sub-170
sub-171
sub-172
sub-173
sub-174
sub-175
sub-176
sub-177
sub-178
sub-179
sub-180
sub-181
sub-182
sub-183
sub-184
sub-185
sub-186
sub-187
sub-188
sub-189
sub-190
sub-191
sub-192
sub-193
sub-194
sub-195
sub-196
sub-197
sub-198
sub-199
sub-200
sub-201
sub-202
sub-203
sub-204
sub-205
sub-206
sub-207
sub-208
sub-209
sub-210
sub-211
sub-212
sub-213
sub-214
sub-215
sub-216
sub-217
sub-218
sub-219
sub-220
sub-221
sub-222
sub-223
sub-224
sub-225
sub-226
sub-227
sub-228
sub-229
sub-230
sub-231
sub-232
sub-233
sub-234
sub-235
sub-236
sub-237
sub-238
sub-239
sub-240
sub-241
sub-242
sub-243
sub-244
sub-245
sub-246
sub-247
sub-248
sub-249
sub-250
sub-251
sub-252
sub-253
sub-254
sub-255
sub-256
sub-257
sub-258
sub-259
sub-260
sub-261
sub-262
sub-263
sub-264
sub-265
sub-266
sub-267
sub-268
sub-269
sub-270
sub-271
sub-272
sub-273
sub-274
sub-275
sub-276
sub-277
sub-278
sub-279
sub-280
sub-281
sub-282
sub-283
sub-284
sub-285
sub-286
sub-287
sub-288
sub-289
sub-290
sub-291
sub-292
sub-293
sub-294
sub-295
sub-296
sub-297
sub-298
sub-299
sub-300
sub-301
sub-302
sub-303
sub-304
sub-305
sub-306
sub-307
sub-308
sub-309
sub-310
sub-311
sub-312
sub-313
sub-314
sub-315
sub-316
sub-317
sub-318
sub-319
sub-320
sub-321
sub-322
sub-323
sub-324
sub-325
sub-326
sub-327
sub-328
sub-329
sub-330
sub-331
sub-332
sub-333
sub-334
sub-335
sub-336
sub-337
sub-338
sub-339
sub-340
sub-341
sub-342
sub-343
sub-344
sub-345

CHANGES
README
code
dataset_description.json
derivatives
participants.json
participants.tsv
stimuli
sub-001
sub-002
sub-003
sub-004
sub-005
sub-006
sub-007
sub-008
sub-009
sub-010
sub-011
sub-012
sub-013
sub-014
sub-015
sub-016
sub-017
sub-018
sub-019
sub-020
sub-021
sub-022
sub-023
sub-024
sub-025
sub-026
sub-027
sub-028
sub-029
sub-030
sub-031
sub-032
sub-033
sub-034
sub-035
sub-036
sub-037
sub-038
sub-039
sub-040
sub-041
sub-042
sub-043
sub-044
sub-045
sub-046
sub-047
sub-048
sub-049
sub-050
sub-051
sub-052
sub-053
sub-054
sub-055
sub-056
sub-057
sub-058
sub-059
sub-060
sub-061
sub-062
sub-063
sub-064
sub-065
sub-066
sub-067
sub-068
sub-069
sub-070
sub-071
sub-072
sub-073
sub-074
sub-075
sub-076
sub-077
sub-078
sub-079
sub-080
sub-081
sub-082
sub-083
sub-084
sub-085
sub-086
sub-087
sub-088
sub-089
sub-090
sub-091
sub-092
sub-093
sub-094
sub-095
sub-096
sub-097
sub-098
sub-099
sub-100
sub-101
sub-102
sub-103
sub-104
sub-105
sub-106
sub-107
sub-108
sub-109
sub-110
sub-111
sub-112
sub-113
sub-114
sub-115
sub-116
sub-117
sub-118
sub-119
sub-120
sub-121
sub-122
sub-123
sub-124
sub-125
sub-126
sub-127
sub-128
sub-129
sub-130
sub-131
sub-132
sub-133
sub-134
sub-135
sub-136
sub-137
sub-138
sub-139
sub-140
sub-141
sub-142
sub-143
sub-144
sub-145
sub-146
sub-147
sub-148
sub-149
sub-150
sub-151
sub-152
sub-153
sub-154
sub-155
sub-156
sub-157
sub-158
sub-159
sub-160
sub-161
sub-162
sub-163
sub-164
sub-165
sub-166
sub-167
sub-168
sub-169
sub-170
sub-171
sub-172
sub-173
sub-174
sub-175
sub-176
sub-177
sub-178
sub-179
sub-180
sub-181
sub-182
sub-183
sub-184
sub-185
sub-186
sub-187
sub-188
sub-189
sub-190
sub-191
sub-192
sub-193
sub-194
sub-195
sub-196
sub-197
sub-198
sub-199
sub-200
sub-201
sub-202
sub-203
sub-204
sub-205
sub-206
sub-207
sub-208
sub-209
sub-210
sub-211
sub-212
sub-213
sub-214
sub-215
sub-216
sub-217
sub-218
sub-219
sub-220
sub-221
sub-222
sub-223
sub-224
sub-225
sub-226
sub-227
sub-228
sub-229
sub-230
sub-231
sub-232
sub-233
sub-234
sub-235
sub-236
sub-237
sub-238
sub-239
sub-240
sub-241
sub-242
sub-243
sub-244
sub-245
sub-246
sub-247
sub-248
sub-249
sub-250
sub-251
sub-252
sub-253
sub-254
sub-255
sub-256
sub-257
sub-263
sub-264
sub-265
sub-266
sub-267
sub-268
sub-269
sub-270
sub-271
sub-272
sub-273
sub-274
sub-275
sub-276
sub-277
sub-278
sub-279
sub-280
sub-281
sub-282
sub-283
sub-284
sub-285
sub-286
sub-287
sub-288
sub-289
sub-290
sub-291
sub-292
sub-293
sub-294
sub-295
sub-296
sub-297
sub-298
sub-299
sub-301
sub-302
sub-308
sub-309
sub-310
sub-311
sub-312
sub-313
sub-314
sub-315
sub-316
sub-317
sub-318
sub-319
sub-320
sub-321
sub-322
sub-323
sub-324
sub-325
sub-326
sub-327
sub-328
sub-329
sub-330
sub-331
sub-332
sub-333
sub-334
sub-335
sub-336
sub-337
sub-338
sub-339
sub-340
sub-341
sub-342
sub-343
sub-344
sub-345

Gorgolewski et al., *Sci Data*, 2016

BIDS

BRAIN IMAGING DATA STRUCTURE

Brain Imaging Data Structure (BIDS)

What is BIDS?

—a standard for organizing neuroimaging data that facilitates **re-use** and **automated processing**

—the standard specifies machine-readable directory structure, filenames, file formats, and metadata

—consensus-based community-driven development capitalizing on existing conventions

—emphasis on simplicity, readability, and accessibility, i.e. ease of use and adoption

—developed out of the **OpenfMRI** (now **OpenNeuro**) open neuroimaging data repository

Gorgolewski et al, *Sci Data*, 2016

Brain Imaging Data Structure (BIDS)

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Data publication and archiving
» Research data

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski[1], Tibor Auer[2], Vince D. Calhoun[3,4], R. Cameron Craddock[5,6], Samir Das[7], Eugene P. Duff[8], Guillaume Flandin[9], Satrajit S. Ghosh[10,11], Tristan Glatard[7,12], Yaroslav O. Halchenko[13], Daniel A. Handwerker[14], Michael Hanke[15,16], David Keator[17], Xiangrui Li[18], Zachary Michael[19], Camille Maumet[20], B. Nolan Nichols[21,22], Thomas E. Nichols[20,23], John Pellman[6], Jean-Baptiste Poline[24], Ariel Rokem[25], Gunnar Schaefer[1,26], Vanessa Sochat[27], William Triplett[1], Jessica A. Turner[3,28], Gaël Varoquaux[29] & Russell A. Poldrack[1]

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

Brain Imaging Data Structure (BIDS)

SCIENTIFIC DATA

OPEN SCIENTIFIC DATA

SUBJECT CATEGORIES
» Data publication and archiving
» Research data

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

OPEN

Comment: MEG-BIDS, the brain imaging data structure extended to magnetoencephalography

Guiomar Niso[1,2], Krzysztof J. Gorgolewski[3], Elizabeth Bock[1], Teon L. Brooks[3], Guillaume Flandin[4], Alexandre Gramfort[5,6], Richard N. Henson[7], Mainak Jas[5], Vladimir Litvak[4], Jeremy T. Moreau[1], Robert Oostenveld[8,9], Jan-Mathijs Schoffelen[8], Francois Tadel[1,10,11], Joseph Wexler[3] & Sylvain Baillet[1]

Received: 14 November 2017
Accepted: 3 May 2018
Published: 19 June 2018

Brain Imaging Data Structure (BIDS)

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Data publication and archiving
» Research data

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

SCIENTIFIC DATA

OPEN

Received: 14 November 2017
Accepted: 3 May 2018
Published: 19 June 2018

SCIENTIFIC DATA

OPEN

COMMENT

EEG-BIDS, an extension to the brain imaging data structure for electroencephalography

Cyril R. Pernet[1], Stefan Appelhoff[2], Krzysztof J. Gorgolewski[3], Guillaume Flandin[4], Christophe Phillips[5], Arnaud Delorme[6,7] & Robert Oostenveld[8,9]

Received: 16 January 2019
Accepted: 7 May 2019
Published online: 25 June 2019

Brain Imaging Data Structure (BIDS)

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Data publication and archiving
» Research data

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

SCIENTIFIC DATA

OPEN

Received: 14 November 2017
Accepted: 3 May 2018
Published: 19 June 2018

SCIENTIFIC DATA

OPEN

COMMENT

Received: 16 January 2019
Accepted: 7 May 2019
Published online: 25 June 2019

SCIENTIFIC DATA

OPEN

COMMENT

Received: 29 January 2019
Accepted: 24 May 2019
Published online: 25 June 2019

iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology

Christopher Holdgraf[1,16], Stefan Appelhoff[2], Stephan Bickel[3], Kristofer Bouchard[4], Sasha D'Ambrosio[5], Olivier David[6], Orrin Devinsky[7], Benjamin Dichter[8], Adeen Flinker[7], Brett L. Foster[9], Krzysztof J. Gorgolewski[8], Iris Groen[10], David Groppe[11], Aysegul Gunduz[12], Liberty Hamilton[13], Christopher J. Honey[14], Mainak Jas[15], Robert Knight[16], Jean-Philippe Lachaux[17], Jonathan C. Lau[18], Christopher Lee-Messer[8], Brian N. Lundstrom[19], Kai J. Miller[20], Jeffrey G. Ojemann[21], Robert Oostenveld[22], Natalia Petridou[23], Gio Piantoni[24], Andrea Pigorini[5], Nader Pouratian[25], Nick F. Ramsey[24], Arjen Stolk[16], Nicole C. Swann[26], François Tadel[6,27], Bradley Voytek[28], Brian A. Wandell[8], Jonathan Winawer[10], Kirstie Whitaker[29,32], Lyuba Zehl[30] & Dora Hermes[8,24,31]

Converting data to BIDS

The hard way...

There are many tools that can facilitate BIDS conversion:
HeuDiConv, **dcm2niix**, **PyBIDS**, **bidsify**, **bidskit**, **pyBIDSconv**, **dcm2BIDS**, etc

This is the most unpleasant part…
a single-use script manually tailored to the idiosyncrasies of each data set (☉_◎)

Converting data to BIDS

The hard way...

There are many tools that can facilitate BIDS conversion: **HeuDiConv**, **dcm2niix**, **PyBIDS**, **bidsify**, **bidskit**, **pyBIDSconv**, **dcm2BIDS**, etc.

The easy way!

Using a prespecified naming convention when creating program cards on the scanner console can allow for automated BIDS conversion—e.g. **ReproIn** (for Siemens).

This is the most unpleasant part…
a single-use script manually tailored to the idiosyncrasies of each data set (☉_◎)

Converting data to BIDS

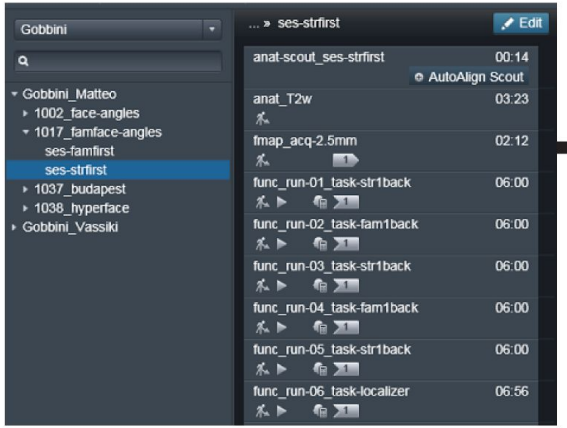
The hard way...

There are many tools that can facilitate BIDS conversion: **HeuDiConv**, **dcm2niix**, **PyBIDS**, **bidsify**, **bidskit**, **pyBIDSconv**, **dcm2BIDS**, etc.

The easy way!

Using a prespecified naming convention when creating program cards on the scanner console can allow for automated BIDS conversion—e.g. **ReproIn** (for Siemens).

Scanner



DICOM

```
001-anat-scout_ses-strfirst
005-anat_T2w
  000001.dcm
  000002.dcm
  ...
006-fmap_acq-2.5mm
007-fmap_acq-2.5mm
008-func_run-01_task-str1back
011-func_run-01_task-str1back
018-func_run-02_task-fam1back
025-func_run-03_task-str1back
032-func_run-04_task-fam1back
039-func_run-05_task-str1back
046-func_run-06_task-localizer
```

$ heudiconv

Datalad

BIDS

```
anat
  sub-sid000005_ses-strfirst_T2w.json
  sub-sid000005_ses-strfirst_T2w.nii.gz
fmap
  sub-sid000005_ses-strfirst_acq-25mm_magnitude1.json
  sub-sid000005_ses-strfirst_acq-25mm_magnitude1.nii.gz
  ...
func
  sub-sid000005_ses-strfirst_task-fam1back_run-02_bold.json
  sub-sid000005_ses-strfirst_task-fam1back_run-02_bold.nii.gz
  sub-sid000005_ses-strfirst_task-fam1back_run-02_events.tsv
  ...
sub-sid000005_ses-strfirst_scans.tsv
```

```
$ git grep TODO

CHANGES:TODOs:
README:TODO: Provide description for the dataset ...
dataset_description.json: "Acknowledgements": "TODO...",
dataset_description.json:   "TODO:",
dataset_description.json: "DatasetDOI": "TODO: ...
task-fam1back_bold.json:  "CogAtlasID": "TODO",
task-fam1back_bold.json:  "TaskName": "TODO: full task name",
task-localizer_bold.json:  "CogAtlasID": "TODO",
task-localizer_bold.json:  "TaskName": "TODO: full task name",
task-str1back_bold.json:  "CogAtlasID": "TODO",
task-str1back_bold.json:  "TaskName": "TODO: full task name",
...
```

Visconti di Oleggio Castello et al, *Zenodo*, 2020

Converting data to BIDS

The hard way...

There are many tools that can facilitate BIDS conversion: **HeuDiConv**, **dcm2niix**, **PyBIDS**, **bidsify**, **bidskit**, **pyBIDSconv**, **dcm2BIDS**, etc.

The easy way!

Using a prespecified naming convention when creating program cards on the scanner console can allow for automated BIDS conversion—e.g. **ReproIn** (for Siemens).

This is the most unpleasant part…
a single-use script manually tailored to the idiosyncrasies of each data set (☉_◎)

Moral of the story:
standardize for sharing from the start and not as an afterthought (๑>ᴗ<)ง

Halchenko & Hanke, *GigaScience*, 2015

Brain Imaging Data Structure (BIDS)

What is BIDS?

—a standard for organizing neuroimaging data that facilitates **re-use** and **automated processing**

—the standard specifies machine-readable directory structure, filenames, file formats, and metadata

—consensus-based community-driven development capitalizing on existing conventions

—emphasis on simplicity, readability, and accessibility, i.e. ease of use and adoption

—developed out of the **OpenfMRI** (now **OpenNeuro**) open neuroimaging data repository

The importance of automated processing

Neuroimaging analysis is complex and flexible—relying on multi-stage processing workflows with many possible analysis choices at each stage (i.e. “researcher degrees of freedom”).

The machine-readable BIDS organization with rich metadata allows for adaptive, automated processing and analysis (i.e. BIDS Apps) that:

—minimize error-prone manual intervention and “procedural overfitting”

—maximize reproducible execution via containerization and content-tracking

Carp, *Front Neurosci*, 2012
Ghosh et al, *F1000Research*, 2017
Botvinik-Nezer et al, *Nature*, 2020

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

Krzysztof J. Gorgolewski[1]*, Fidel Alfaro-Almagro[2], Tibor Auer[3], Pierre Bellec[4,5], Mihai Capotă[6], M. Mallar Chakravarty[7,8], Nathan W. Churchill[9], Alexander Li Cohen[10], R. Cameron Craddock[11,12], Gabriel A. Devenyi[7,8], Anders Eklund[13,14,15], Oscar Esteban[1], Guillaume Flandin[16], Satrajit S. Ghosh[17,18], J. Swaroop Guntupalli[19], Mark Jenkinson[2], Anisha Keshavan[20], Gregory Kiar[21,22], Franziskus Liem[23], Pradeep Reddy Raamana[24,25], David Raffelt[26], Christopher J. Steele[7,8], Pierre-Olivier Quirion[15], Robert E. Smith[26], Stephen C. Strother[24,25], Gaël Varoquaux[27], Yida Wang[6], Tal Yarkoni[28], Russell A. Poldrack[1]

Automated processing and containerization

What is a BIDS App?

—the machine-readable BIDS format enables automated processing via **BIDS Apps**

—BIDS Apps use **containerization** to facilitate portability and reproducibility

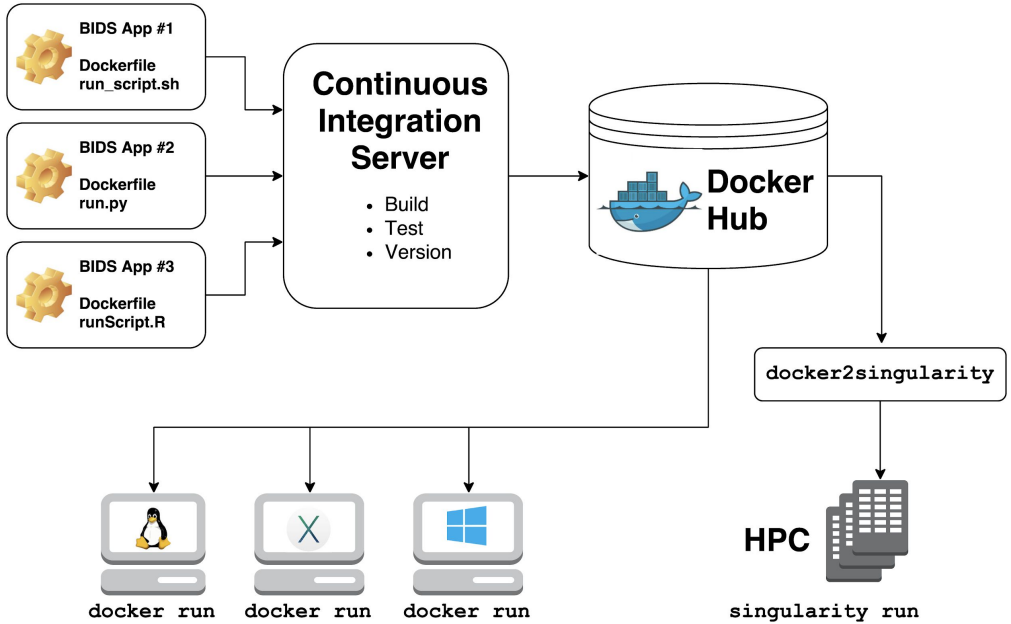
Software containers

Containerization solutions package software with all dependencies in single, encapsulated image.

Docker (for local use)

Singularity (for servers)

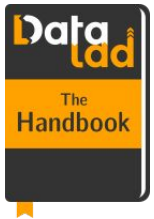
Kurtzer et al, *PLOS One*, 2017



Data provenance

Content tracking with DataLad

All stages of data processing are version-controlled and distributed using DataLad.

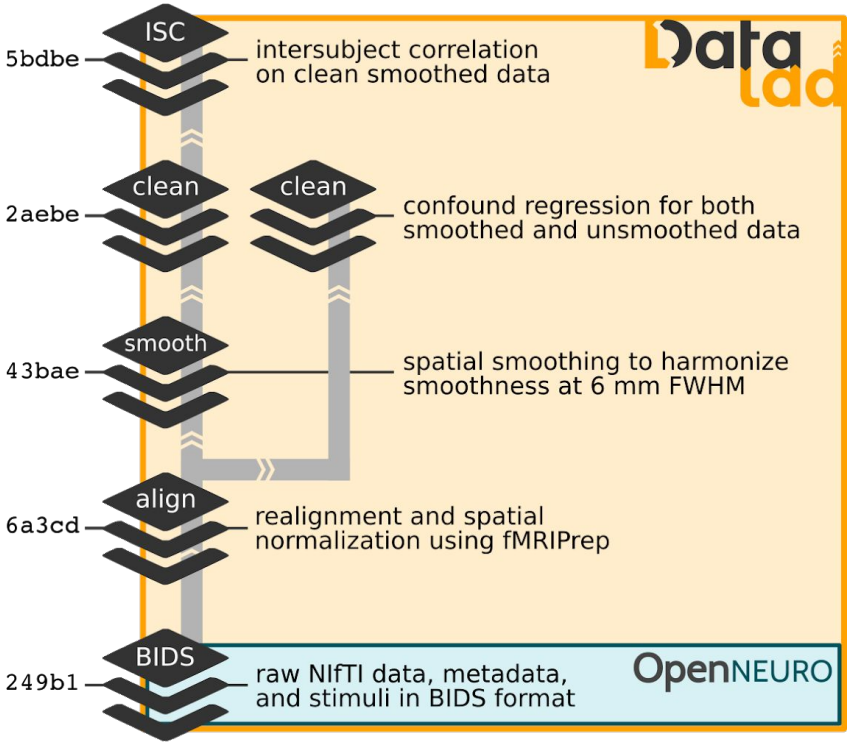


https://github.com/snastase/narratives
http://datalad.org — discover
http://github.com/datalad/datalad — contribute
http://handbook.datalad.org — learn

Review article

Michael Hanke*, Franco Pestilli, Adina S. Wagner, Christopher J. Markiewicz, Jean-Baptiste Poline and Yaroslav O. Halchenko

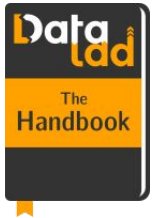
In defense of decentralized research data management



Data provenance

Content tracking with DataLad

All stages of data processing are version-controlled and distributed using DataLad.



https://github.com/snastase/narratives
http://datalad.org — discover
http://github.com/datalad/datalad — contribute
http://handbook.datalad.org — learn

Review article

Michael Hanke*, Franco Pestilli, Adina S. Wagner, Christopher J. Markiewicz, Jean-Baptiste Poline and Yaroslav O. Halchenko

In defense of decentralized research data management

```
commit ea46e14fd0dc949acd654947e136fe2c354ef780 (HEAD -> master)
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Thu Oct 17 00:40:28 2019 -0400

    Updated some missing condition labels (prettymouth)

commit a46de0940b96057d7b0297a48ab2f7a8d08a1c1e
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Tue Oct 15 18:06:08 2019 -0400

    Added audio files to stimuli subdataset

commit a1019e628f4e6a40730941d3ec467e15fbdc8560
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Tue Oct 15 15:38:40 2019 -0400

    Added subdataset for stimuli

commit e93be6b469e7e664abc2dd6c0790002a9d0171c0
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 18:45:51 2019 -0400

    Added subdatasets for code and derivatives

commit bee66631b402f4a048244c8c3cb1ebda2ad33c3a
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 18:35:07 2019 -0400

    Populated raw NIfTIs and metadata (BIDS valid)

commit daf3b89009b9bde93fabffdb2f11d2788ea44688
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 17:39:19 2019 -0400

    Instruct annex to add text files to Git

commit bc03620e50865195228e46c1ab3da3a2fa87e6cf
Author: Sam Nastase <sam.nastase@gmail.com>
Date:   Mon Oct 14 17:39:17 2019 -0400

    [DATALAD] new dataset
```

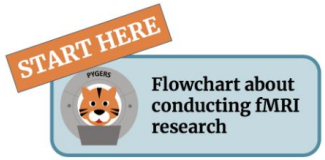

Brainhack Princeton 2020
Princeton, NJ, USA
December 9-11, 2020
Register Now
Schedule
Submit a Project
Code of Conduct
Join Princeton Brainhack on Mattermost
Photo by Smallbones

Brainhack: developing a culture of open, inclusive, community-driven neuroscience

This is a version for community review on psyarxiv and does not constitute the final version of the manuscript.

Rémi Gau*[1], Stephanie Noble*[2], Katja Heuer*[3,4], Katherine L. Bottenhorn*[5], Isil P. Bilgin*[6,7], Yu-Fang Yang*[8], Julia M. Huntenburg*[9], Johanna Bayer*[10,11], Richard A.I. Bethlehem*[12,13], Shawn A. Rhoads[14], Christoph Vogelbacher[15], Valentina Borghesani[16], Elizabeth Levitis[17,18], Hao-Ting Wang[19,20,21], Sofie Van Den Bossche[22], Xenia Kobeleva[23,24], Jon Haitz Legarreta[25], Samuel Guay[26], Selim Melvin Atay[27], Gael P. Varoquaux[28,29], Dorien C. Huijser[30,31], Malin S. Sandström[32], Peer Herholz[33], Samuel A. Nastase[34], AmanPreet Badhwar[35,16,36], Guillaume Dumas[37,38], Simon Schwab[39], Stefano Moia[40,41], Michael Dayan[42], Yasmine Bassil[43], Paula P. Brooks[34], Matteo Mancini[20,44,45], James M. Shine[46], David O'Connor[47], Xihe Xie[48], Davide Poggiali[49], Patrick Friedrich[50], Lydia Riedl[51], Roberto Toro[52,53], Anibal S. Heinsfeld[54,55], César Caballero-Gaudes[40], Anders Eklund[56,57,58], Kelly G. Garner[59,60,61], Christopher R. Nolan[62], Damion V. Demeter[63], Fernando A. Barrios[64], Junaid S. Merchant[65,66], Elizabeth A. McDevitt[34], Robert Oostenveld[67,68], R. Cameron Craddock[69], Ariel Rokem[70], Andrew Doyle[71], Satrajit S. Ghosh[72,73], Aki Nikolaidis[74], Olivia W. Stanley[75,76], Eneko Uruñuela[40,41], The Brainhack Community

The Princeton Handbook for Reproducible Neuroimaging will provide you with steps and best practices about how to collect and analyze fMRI data.



The Princeton Handbook for Reproducible Neuroimaging

Important:

Welcome to the Princeton Handbook for Reproducible Neuroimaging! The handbook is currently under active development—expect substantial changes in the near future. The goal of this handbook is to provide a reference for best practices in reproducible fMRI research. There's no single "right" answer for many questions in fMRI, but here we try to provide helpful references and recommendations. Many elements of the handbook are specific to the Princeton Neuroscience Institute computing infrastructure, but the principles are widely applicable. This document will be updated over time as best practices evolve. We hope that this handbook will be useful as you embark on your own journey doing fMRI experiments!

Please contact us if you have any questions, feedback, or suggestions!

Brand new to neuroimaging?

Familiarize yourself with important background reading material here. If you're unsure about any terms, check out the glossary.

Brooks et al, *Zenodo*, 2021

Developing community resources for reproducible neuroimaging

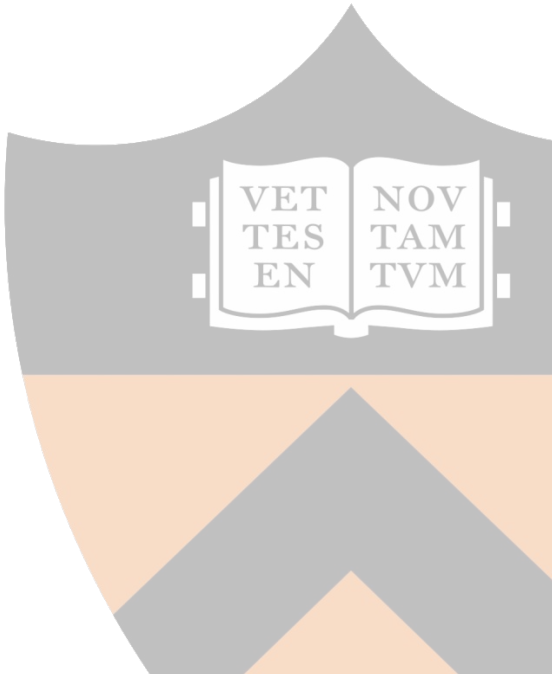
TL;DR

—data standardization and sharing are critical for democratizing research

—follow best practices from the start—not as an afterthought!

—incentive maintaining and expanding existing infrastructure (don't reinvent the wheel!)

—teaching students how to conduct transparent, reproducible research is probably more important than the actual content of their research (⊙ _ ⊙)



Developing community resources for reproducible neuroimaging

TL;DR

—data standardization and sharing are critical for democratizing research

—follow best practices from the start—not as an afterthought!

—incentive maintaining and expanding existing infrastructure (don't reinvent the wheel!)

—teaching students how to conduct transparent, reproducible research is probably more important than the actual content of their research (⚆ _ ⚆)

Thanks for listening!

NIH National Institute of Mental Health

ReproNim