



HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

Neuroimage. 2017 January ; 144(Pt B): 259–261. doi:10.1016/j.neuroimage.2015.05.073.

OpenfMRI: Open sharing of task fMRI data

Russell A. Poldrack and Krzysztof J. Gorgolewski

Department of Psychology, Stanford University, Stanford CA, USA

Abstract

OpenfMRI is a repository for the open sharing of task-based fMRI data. Here we outline its goals, architecture, and current status of the repository, as well as outlining future plans for the project.

1. Introduction

Task-based fMRI has become one of the primary tools of cognitive neuroscience, providing the ability to interrogate the neural basis of mental functions and representations. Most task fMRI studies involve relatively small samples (usually less than 50 subjects), and it is rare that exactly the same tasks are performed across many different studies (given that task development is the primary source of conceptual novelty in these studies). In addition, a significant amount of metadata (including description of task events and their timing) is required to analyze a raw task fMRI dataset. For these reasons, the sharing and combination of task-based fMRI data is significantly more challenging than structural MRI and resting state fMRI, for which data can be relatively easily combined across studies. The sharing of task-based fMRI data got an early start with the fMRI Data Center [12], but this early repository came at a time when the field was not yet ready for widespread data sharing, either technically or socially. Nonetheless, it lit the way for later data sharing efforts, and showed how shared data could be used to make new discoveries [13].

The OpenfMRI database (<http://www.openfmri.org>) [8] was designed as an open repository for task fMRI data. The inception of this project came about when one of us (RP) moved from UCLA to the University of Texas in 2009. In earlier work, he and colleagues had begun to apply novel analyses across multiple datasets with the goal of decoding mental tasks as well as characterizing the large-scale neural networks underlying task performance [9]. With his move came the need to deidentify all of the data used in these previous analyses, so he decided to go ahead and make the data publicly available, in the hope that others would also contribute data to the collection, thus providing an even more powerful foundation for task-based decoding. With funding from the National Science Foundation's program on Collaborative Research in Cognitive Neuroscience and support from the Texas Advanced Computing Center (TACC), we developed the OpenfMRI web site and database. We also

poldrack@stanford.edu Telephone: +1-650-497-8488 +1-FAX: 650-725-5699.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

teamed with several other labs to provide additional datasets, and published a set of analyses that highlighted the utility of this newly-grown database for the characterization of neural systems underlying task performance [8]. Subsequently the data have been used to address a number of novel questions, both neuroscientific [e.g. 3] and methodological [e.g. 7, 4, 10].

2. What is our goal?

The OpenfMRI database was designed to serve as a repository for the open sharing and dissemination of task-based fMRI data. As it has grown, it has broadened to encompass other datatypes as well, including EEG, MEG, resting fMRI, and diffusion MRI acquired on both healthy and clinical populations, and we now accept datasets that do not necessarily include task fMRI. We do, however, require of each dataset that it must be publicly distributable without requiring any signed usage agreements. Our database has been recommended as a repository for neuroimaging data by data centric journals such as Scientific Data and F1000Research. Although most of the datasets in the database are associated with publications, this is not a requirement. The database is currently open to submissions from everyone and continues to grow.

The initial goal of the project was to serve highly curated datasets that are organized in a very specific manner [8], to allow automated processing of the data. We quickly found that this curation process is labor-intensive and time-consuming, also requiring substantial interactions with the dataset providers. This need for curation has limited the growth of the database. In order to allow sharing of data on a more rapid timescale, we now allow the sharing of uncurated datasets (clearly labeled as such) while they are in the process of curation.

In recent years, there has been movement towards the sharing of large, well-curated, and highly general datasets such as the Human Connectome Project. The goal of OpenfMRI is to provide a complementary venue for every fMRI researcher to publish their task fMRI data, regardless of the size or generality of the dataset. There is a growing set of motivations that lead researchers to share their data, including data papers[6], journal requirements[11], grant requirements[1], ethical concerns[2], or genuine interest in the progress of science, but in addition to motivation researchers need a resource that provides support for open sharing. OpenfMRI provides such platform and thus access to a variety of experimental designs and population characteristics. This helps to capture the “long tail” of data sharing [5].

3. What is available?

Currently the OpenfMRI database contains 37 studies comprising a total of 1411 subjects; of these 22 have been curated, with a total of 499 subjects. All datasets include task-based functional and basic structural MRI, which are the only requirements for inclusion. In addition, there are studies with diffusion MRI, simultaneous EEG, and MEG. The current datasets come from 22 different laboratories, with a significant number (13) coming from author RP’s laboratory. The vast majority of the datasets have been collected at 3T, with four including data collected at 1.5T and one including data collected at 7T. Image resolutions vary from 1.4 – 4.0 mm in plane and 1.54 – 6.0 mm through plane. The studies vary in terms

of the nature of demographic and other phenotypic information available. For those with demographic information available, we find an age range of 11–49 (mean 24.0 \pm 5.64), and 57% male participants. All MRI data are shared in NIFTI format. Raw data are directly available from the web site, in the form of compressed TAR archives. For most datasets a single archive is presented for the entire study, but for some datasets the data must be split into multiple files in order to allow successful downloading. The dataset sizes range from 0.4 to 329 gigabytes (median = 2.51 gigabytes). Each dataset is associated with a specific accession number and has a permanent citable URI based on this accession number. Upon receipt, data are initially posted to the site in uncurated form (clearly specified as such). The data are then curated by a member of the OpenfMRI team in order to ensure their completeness, accuracy, and adherence to the OpenfMRI file organization and metadata standards. Each dataset is run through a manual curation process assuring that the accompanying metadata are accurate.

The variety of tasks used in datasets deposited in OpenfMRI enabled new exciting research questions [10] which would not be possible with a more homogenous database. Additionally, the fact that the data is easily accessible made it very appealing in methodological validations. Up to date there are over 30 publications that use OpenfMRI to test, validate, and improve neuroimaging methods.

4. Access

The access system was designed in order to allow the widest possible access and usage of the data. The database is openly accessible to any user; no registration, approval, or data usage agreement is required. Licensing varies across datasets. The default license is a Public Domain Dedication and License v1.0 <http://www.opendatacommons.org/licenses/pddl/1.0/>. However, upon request by the data submitter the use of other licenses is allowed. In addition, we strongly encourage all data users to follow the ODC Attribution/Share-Alike Community Norms, which recommend (but do not legally require) attribution and sharing of any resulting products.

Each dataset includes a set of release notes that outline any changes made to the data. In cases when those changes are substantial, the previous version of the dataset will remain accessible so that researchers can compare results across versions if necessary. Major changes and additions to the database are announced via the @OpenfMRI Twitter feed.

5. Architecture

The data are currently stored both on a replicated storage system at the Texas Advanced Computing Center (TACC), and as part of the Public Data Sets project at Amazon. Initial processing and curation of the data takes place on the TACC system; upon completion of the analyses and curation, the entire raw and processed dataset is uploaded to Amazon's Simple Storage Service (S3) for distribution. This system has very high throughput and thus provides sufficient resources for large requests. The fact that data is stored on S3 makes it easier (and cheaper) to access it directly from Amazon Elastic Compute Cloud (EC2) virtual machines that provide flexible data processing capabilities.

6. Contribution

The database is open to all contributors. The requirements for contribution are that the data be deidentified (including removal of facial structures from structural MRI data, and removal of any unique identifiers from image filenames or metadata), and that the investigator have the appropriate IRB approval for public and unconstrained sharing of the deidentified data. In addition, we request that the data be formatted according to the published OpenfMRI data and metadata organization standards; failure to perform this formatting will result in delays in the curation process.

7. Long-term plans

In the short term, we have funding from the National Science Foundation and National Institute of Drug Abuse to maintain and develop the resource. In addition, we have obtained funding from the Laura and John Arnold Foundation to establish the next generation of the project, which will include development of new tools for automated data ingestion and processing of the data to assess reproducibility.

Acknowledgments

Development and population of the OpenfMRI resource has been supported by NSF Grant #OCI-1131441, NIDA Grant #R21DA034316, and the Laura and John Arnold Foundation.

References

1. [Accessed: 2014-12-3] NIH sharing policies and related guidance on NIH-Funded research resources. <http://grants.nih.gov/grants/sharing.htm>
2. Brakewood, Beth; Poldrack, Russell A. The ethics of secondary data analysis: considering the application of belmont principles to the sharing of neuroimaging data. *Neuroimage*. 2013 Nov 15;82:671–676. [PubMed: 23466937]
3. Cai, Weidong; Ryali, Srikanth; Chen, Tianwen; Li, Chiang-Shan R.; Menon, Vinod. Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: Evidence from intrinsic and task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. *J Neurosci*. 2014 Oct; 34(44):14652–14667. [PubMed: 25355218]
4. Joshua Carp. On the plurality of (methodological) worlds: estimating the analytic flexibility of fmri experiments. *Front Neurosci*. 2012; 6:149. [PubMed: 23087605]
5. Ferguson AR, Nielson JL, Cragin MH, et al. Big data from small data: data-sharing in the 'long tail' of neuroscience. *Neuroscience*. 2014
6. Gorgolewski, Krzysztof J.; Margulies, Daniel S.; Milham, Michael P. Making data sharing count: a publication-based solution. *Front. Neurosci*. 2013 Feb.7(9):6. [PubMed: 23386809]
7. Marquand, Andre F.; Brammer, Michael; Williams, Steven CR.; Doyle, Orla M. Bayesian multi-task learning for decoding multi-subject neuroimaging data. *Neuroimage*. 2014 May.92:298–311. [PubMed: 24531053]
8. Poldrack, Russell A.; Barch, Deanna M.; Mitchell, Jason P.; Wager, Tor D.; Wagner, Anthony D.; Devlin, Joseph T.; Cumba, Chad; Koyejo, Oluwasanmi; Milham, Michael P. Toward open sharing of task-based fmri data: the openfMRI project. *Front Neuroinform*. 2013; 7:12. [PubMed: 23847528]
9. Poldrack, Russell A.; Halchenko, Yaroslav O.; Hanson, Stephen José. Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychol Sci*. 2009 Nov; 20(11):1364–1372. [PubMed: 19883493]
10. Schwartz, Yannick; Thirion, Bertrand; Varoquaux, Gaël. Mapping paradigm ontologies to and from the brain. *NIPS*. 2013:1673–1681.

11. Silva, Liz. PLOS' new data policy: Public access to data. 2014 Feb 24. [Accessed: 2014-12-3] <http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/>.
12. Van Horn JD, Grethe JS, Kostelec P, Woodward JB, Aslam JA, Rus D, Rockmore D, Gazzaniga MS. The functional magnetic resonance imaging data center (fmridc): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philos Trans R Soc Lond B Biol Sci*. 2001 Aug; 356(1412):1323–1339. [PubMed: 11545705]
13. Darrell, John; Horn, Van; Gazzaniga, Michael S. Why share data? lessons learned from the fmridc. *Neuroimage*. 2012 Nov.:10.

Highlights

- OpenfMRI.org is an open database for sharing of task-based fMRI data.
- Data are shared with no restrictions on access or usage agreements.
- The database is open to all contributors.