



Chapter 1

Introduction to Methods for Analyzing Large Neuroimaging Datasets

Robert Whelan and Hervé Lemaître

Abstract

There is a recognition in the field of neuroimaging that sample size must drastically increase to achieve adequate statistical power and reproducibility. Several large neuroimaging studies and databases, such as OpenNeuro and the Adolescent Brain and Cognitive Development project, have emerged, offering open access to vast amounts of data. However, there is a dearth of practical guidance for working with large neuroimaging datasets, a deficit that this book seeks to address. With the emphasis on providing hands-on instruction, chapters contain worked examples using open-access data.

Key words Neuroimaging, Electroencephalography, Standardization, Toolboxes, Machine learning, Artificial intelligence

1 Structure of the Book

This book on methods for analyzing large neuroimaging datasets is organized as follows. In Subheading 4.1, the reader is shown how to access and download large datasets, and how to compute at scale. In Subheading 4.2, chapters cover best practices for working with large data, including how to build reproducible pipelines, the use of Git for collaboration, and how to make electroencephalographic and functional magnetic resonance imaging data sharable and standardized. In Subheading 4.3, chapters describe how to do structural and functional preprocessing data at scale, incorporating practical advice on potential trade-offs of standardization. In Subheading 4.4, chapters describe various toolboxes for interrogating large neuroimaging datasets, including those based on machine learning and deep learning approaches. These methods can be applied to connectomic and region-of-interest data. Finally, the book contains a glossary of useful terms.

2 Why This Book Is Needed: Neuroimaging Datasets Are Getting (Much) Bigger

Human brain imaging is in a period of profound change. The sample sizes of the first neuroimaging studies were relatively low, perhaps with dozen subjects [1] or even single-subject analyses [2]. Such studies were incredibly valuable, facilitating non-invasive exploration of human brain structure and function. Those early studies paved the way for a new generation of neuroscientists. A few decades later, we have more complete and increasingly detailed maps of the human brain, but there is a growing recognition that sample size must drastically increase to achieve adequate statistical power and reproducibility. Indeed, it has been suggested that brain-wide association studies using neuroimaging may be unreliable without very large samples [3]. Fortunately—with added impetus from the fields of imaging genetics and neuroepidemiology—there has been a radical increase in the number of subjects within neuroimaging studies. For imaging genetics, the main reason was statistical power because the effects of common individual genetic variants are small and thus required large samples of subjects [4]. For neuroepidemiology, there was a paradigm shift from a goal of identifying robust diagnostic biomarkers to a prevention/prediction orientation, which necessarily requires large samples of healthy subjects, some of whom may subsequently develop a disease [5].

Given the scientific benefits of increased sample sizes, several large neuroimaging studies have been established. For example, the Adolescent Brain and Cognitive Development study (ABCD: <https://abcdstudy.org/>) is a 10-year neuroimaging project that will recruit over 10,000 people. Importantly, ABCD data are open access, and available with minimal restrictions. The UK Biobank (<https://www.ukbiobank.ac.uk/>) has neuroimaging data from 100,000 people, and these data are available to researchers for an access fee. Similar open-access magnetic resonance imaging (MRI) databases include the Alzheimer's Disease Neuroimaging Initiative (ADNI: <http://adni.loni.usc.edu/>) and Open Access Series of Imaging Studies (OASIS; <https://www.oasis-brains.org/>). Open access electroencephalographic (EEG) databases include Two Decades-Brainclinics Research Archive for Insights in Neurophysiology (TDBRAIN) [6] and the Child Mind Institute Multimodal Resource for Studying Information Processing in the Developing Brain (MIPDB). There are also datasets that belong to large consortia, such as IMAGEN (<https://imagen-europe.com>) [7] and EuroLADEEG [8].

As part of the effort to address poor reproducibility in neuroimaging research, the open science movement has fostered neuroimaging research to share data, codes, and publications [9]. In addition to the single studies with very large sample sizes described

above, advances in open science have provided the conditions to allow researchers to access large neuroimaging datasets, or to combine several datasets to conduct meta- or mega-analyses [10, 11]. OpenNeuro (<https://openneuro.org>) [12] is a popular resource that hosts a variety of brain data, shared according to FAIR principles. The NeuroElectroMagnetic data Archive and tools Resource (NEMAR; <https://nemar.org/>) [13], contains EEG, magnetoencephalography, and intracranial EEG from OpenNeuro.

3 Why Focus Specifically on Analysis Methods for Large Neuroimaging Datasets?

We proposed to develop this book because several additional challenges arise when analyzing large—rather than small or medium—neuroimaging datasets. These challenges include the need for greater standardization, importance of good code management, use of scalable methods to process large volumes of data, and use of appropriate methods to uncover between-group or individual differences. The following chapters will bring the reader systematically through the essentials of working with large neuroimaging datasets, from downloading and storing data; to best practices for ensuring reproducibility; to preprocessing functional and structural data; to toolboxes for statistical analysis. Each chapter has comprehensive step-by-step instructions on a particular method (including examples of code where appropriate). We have not included any chapters in this book on neuroimaging acquisition. This is because, with respect to very large neuroimaging datasets, the data have either already been collected or will be collected according to a consensus protocol.

4 Overview of Chapters in This Book

4.1 Section 1: *Accessing and Computing at Scale*

Chapter 2 is titled “Getting Started, Getting Data” (Lemaître et al.) and illustrates different methods for downloading datasets using command lines (wget, curl), data management tools such as Datalad, Amazon Web Services, and graphical user interface options (e.g., Cyberduck). Chapter 2 demonstrates how to download data from OpenNeuro for a range of operating systems, which is important for using the worked examples later in the book. In general, after reading Chapter 2, researchers will be equipped with the knowledge and tools to download large neuroimaging datasets.

Analysis of large neuroimaging datasets requires scalable computing power and storage, plus methods for secure collaboration and for reproducibility. For example, data preprocessing is unlikely to be possible on a single computer. In Chapter 3 (Madhyastha)—“Neuroimaging Workflows in the Cloud”—the theory and practice of using cloud computing to address many of these requirements is

presented. Cloud computing offers a highly flexible model that is typically more cost-effective than a single laboratory investing in computer equipment to accommodate its peak demand. Chapter 3 describes the various considerations and options related to cloud-based neuroimaging analyses, including cost models and architectures. In this chapter, you will learn how to run a neuroimaging workflow in order to leverage cloud-computing capabilities. Using data from the AOMIC-PIOP2 project hosted on OpenNeuro, this chapter shows how to use Nextflow to create a very simple skull stripping and tissue segmentation workflow using FSL's *bet* and *fast* programs installed on a local computer. Nextflow allows scalability from a laptop to a cluster to cloud-native services with no code changes.

4.2 Section 2: Best Practices for Working with Large Data

Unlike neuroimaging datasets with ~30 participants, where it could be possible to manually or individually apply processing steps or statistical tests, a key part of working with large neuroimaging datasets involves controlling all steps with code. Therefore, we devote Chapter 4, “Establishing a Reproducible and Sustainable Analysis Workflow” (Ramduny et al.), to best practices for producing reproducible pipelines. In Chapter 4, you will also learn about FAIR principles (data should be Findable, Accessible, Interoperable, and Reusable). The BIDS (Brain Imaging Data Structure) format, which provides a common structure for data organization, is an extremely important tool for working with neuroimaging data (including M/EEG). Chapter 4 contains a worked example with Docker, using fMRIPrep on an open-access dataset. There is also a section on working with Python notebooks and invaluable advice on writing sharable and reusable code, including commenting and debugging tips. Finally, perhaps the most impactful advice in Ramduny et al. is to write code and process data in the most efficient way possible to minimize the energy burden.

Continuing the emphasis on efficient code management, Chapter 5—“Optimising Your Reproducible Neuroimaging Workflow with Git” (Garcia and Kelly)—demonstrates the use of Git, which is a very important tool for collaboration and scalable neuroimaging. In Chapter 5, you will learn via a worked example of a cluster analysis on open-access data: version control, branching (especially useful when collaborating), and conflict resolution. Garcia and Kelly also describe how to use GitHub, and again the benefits of collaboration are outlined.

In addition to very large, centrally coordinated studies (e.g., ABCD), data aggregation is an efficient way to build large datasets. “Mega” analyses can provide insights that are otherwise not afforded by smaller studies. Making data sharable and standardized is therefore crucially important, and we include chapters here relevant to both EEG and functional MRI (fMRI). With respect to EEG, in Chapter 6—“End-to-End Processing of M/EEG Data

with BIDS, HED, and EEGLAB”—Trong and colleagues introduce a combined BIDS and Hierarchical Event Descriptors (HED) approach that addresses a notable gap in the methods landscaper: namely, a standardized approach to characterize events during time series data. HED is a vocabulary designed to describe experiment events in a structured human-readable and machine-actionable way and HED metadata can enable intelligent combining of event-related data from different recordings and studies. HED can be accessed in several ways: as online tools, Python-based command line scripts and notebooks, and MATLAB scripts and plug-in tools for EEGLAB. Chapter 6 is focused on neuroelectric approaches and, as with all chapters, there is a worked example of end-to-end processing of EEG data using standardized BIDS and HED format to organize and describe information about the dataset. With respect to fMRI, in Chapter 7—“Actionable Event Annotation and Analysis in fMRI: A Practical Guide to Event Handling”—Denissen et al., again building on BIDS, describe tools for efficiently generating event files from experimental logs. These event restructuring tools (remodelers) allow users to modify a dataset’s event files by specifying a series of operations in a JSON text file, improving reproducibility, and reducing the need for bespoke coding solutions. An example of using HED remodeling tools is given via a simple analysis of two datasets, working through the required event restructuring.

4.3 Section 3: Preprocessing

Preprocessing of large MRI and M/EEG datasets can be computationally expensive. Here, we describe methods for preprocessing data derived from functional MRI (Chapter 8), from structural MRI data—both gray (Chapter 9) and white matter (Chapter 10)—and from EEG (Chapter 11).

Chapter 8 (Esteban), using NiPreps as a foundation, focuses on the preprocessing stage of neuroimaging pipelines, exploring the rationale, benefits, and potential tradeoffs of standardization. It explores dimensions such as standardizing inputs and outputs, and modularization using tools such as NiPreps and NiReports. Emphasis is placed on version control, software engineering practices, and the use of TemplateFlow for standardizing spatial mappings. Challenges in implementation choice are discussed. Finally, Chapter 8 delves into the integration of artificial intelligence, including the importance of developing transparent, interpretable deep-learning models, trained on openly available data.

Voxel-based morphometry (VBM) is a widely used method for structural MRI analysis, quantifying local gray matter volume (GMV) by segmenting whole brain scans into tissue classes. Applying VBM to detect structural brain-behavior associations in moderate-to-large-sized samples faces challenges, with findings prone to overestimation and limited replicability. Chapter 9 (Hoffstaedter, Antonopoulos, and Gaser)—“Structural MRI and Computational

Neuroanatomy”—is a demonstration of the fully automatic processing of a public dataset with CAT12 in a fully reproducible workflow [14]. Chapter 9 emphasizes the importance of methodological transparency, public data sharing, and the availability of analysis code to enhance reproducibility and facilitate replications. A practical demonstration of fully automated processing using CAT12 on a public dataset is presented, showcasing a reproducible workflow.

Chapter 10—(Connaughton et al.) begins with an introduction to the concepts and techniques of diffusion MRI data processing used in the field and a step-by-step guide for processing diffusion imaging data and for generating tractography. Chapter 10 demonstrates the usage of the popular diffusion imaging toolbox, *ExploreDTI* [15]. Working with BIDS-formatted data, Chapter 10 contains advice specific to very large datasets, plus several helpful recommendations (especially for the novice), and identifies common pitfalls. Chapter 10 describes steps that can be taken to reduce processing time through resource optimization, and options are given to find the optimal balance between reconstruction accuracy and processing time.

In Chapter 11 (Sainz-Ballesteros et al.), the *ConneEEGtome* toolbox is introduced. Although this toolbox also contains methods for between-group comparisons, we included it in this section because users will encounter the preprocessing features first. Relative to MRI, multicentric high-density EEG studies are less common, even though EEG is much more scalable than MRI [16]. EEG presents extra challenges with data harmonization, not least because there are many different hardware configurations and montages. *ConneEEGtome* offers an elegant and open-access solution to these challenges, including the option of a graphical user interface. Using data in EEG-BIDS format, the authors bring the reader step-by-step through from preprocessing to classification. Notably, an automatic artifact rejection approach based on independent components is included, as is bad channel interpolation. Recommendations are provided for optimizing storage needs.

4.4 Section 4: Toolboxes for Statistical Analysis

In the next section, we introduce several toolboxes for statistical analysis of large datasets. The *ConneEEGtome* described in Chapter 11 is a toolbox that includes a classifier, with feature selection followed by Gradient Boosting Machines and a feature importance report. Chapter 12 (Hahn et al.) describes the *Brain Predictability toolbox* (BPt), which is a Python-based, cross-platform, toolbox. BPt can run analyses on single personal computers, with the option to scale up to be used in a cloud-computing environment, and has many user-friendly features, with inbuilt safeguards to prevent the many common errors that beginner users make. BPt provides support for several common data preparation steps: data organization, exploratory data visualization, transformations such as k-binning and binarization, automatic outlier

detection, information on missing data, and other summary measures. A very useful aspect is the ability to correctly impute missing data (i.e., without data leakage). Recommendations are included for when to tune model hyperparameters. Understanding your model is made easier because feature importances from BPt can be easily visualized through the related Python package. As with the other chapters, a step-by-step example is included.

Chapter 13 (Serin et al.) describes the *NBS-Predict* toolbox, which builds on network-based statistics (NBS) to produce connectome-based predictions. In this way, connected graph components are used as features, thus incorporating the topological structure of features into account. NBS-Predict is a particularly user-friendly tool that allows the user to easily create models via a graphical user interface and enables automatic generation of training and test datasets for cross-validation purposes. There are two worked examples in Chapter 13: a linear regression and a classification.

The Predictive Clinical Neuroscience (PCN) Toolkit is described in Chapter 14 (Rutherford & Marquand), and offers an easy way to apply the powerful tool that is Normative Modeling. With Normative Modeling, the overarching aim is to define a reference range for a certain given structural or functional brain measurement in a certain sample and to create a reference standard. This reference standard can then be used as a comparator for individuals living with neurological or psychiatric conditions. The PCN toolbox is Python based. In the step-by-step tutorial, you will make predictions for a multi-site transfer dataset, derived from open-access data. The PCN toolkit contains a very useful facility to create an “adaptation dataset” to account for confounding variables such as site effects. The PCN toolkit outputs several evaluation metrics, which can be saved for later plotting/interrogation, and advice and recommendations for further post hoc analyses are included.

In Chapter 15 (Boyle and Weng) presents a flexible method, optimized for large datasets, for implementing *connectomic predictive modeling* (a popular approach for predicting phenotypes from fMRI connectivity data). Notably, this chapter includes an option to use leave-site-out validation. A schematic overview is provided plus all code necessary to conduct an analysis: only a beginner level of coding is needed. A helpful schematic of all decision points is included. Recommendations—such as doing global signal regression and for handling missing data—are provided, as are resources for plotting the output and for implementing computational lesions.

Turning to toolboxes that can be used to interrogate structural MRI, in Chapter 16, Moguilner and Ibañez describe in detail their application of *DenseNet* (a convolutional neural network) to MRI images. Notably, these images had not been preprocessed: the

ability to utilize raw MR images expands the possibilities for the use of medical images, which are acquired with a variety of sequences from a range of manufacturers and include heterogeneous samples. All of the relevant code is open source, based on MONAI's PyTorch-based tools. A step-by-step example in Google Colab, using structural data from the AOMIC database, is provided including starting parameters for the DenseNet. Code is included for plotting model performance and metrics such as area under the curve of the receiver operating characteristic from the test set are included, as is code for an "occlusion sensitivity" map, which shows the brain regions that contribute to the prediction.

At the end of this book, we have added a list of resources. These include a list of tools described in the book. We also refer the reader to an excellent compilation of ~300 resources by Niso et al. [17]. There is also a list of open-access neuroimaging databases, and we refer the reader to Madan [18] Table 1 for a comprehensive list. The list of resources also includes information on resources for learning coding languages such as Python. There is also a glossary at the back with explanations of technical terms used in this book.

5 Concluding Comments

In summary, as neuroimaging datasets continue to grow in size, the need for standardized methodologies, efficient code management, and scalable data processing becomes increasingly crucial. This book seeks to address these challenges, and it is our goal to equip researchers with the practical knowledge necessary for conducting robust and reproducible analyses of large neuroimaging datasets.

References

1. Szucs D, Ioannidis JPA (2020) Sample size evolution in neuroimaging research: an evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage* 221:117164. <https://doi.org/10.1016/j.neuroimage.2020.117164>
2. Boddaert N, Barthélémy C, Poline J-B, Samson Y, Brunelle F, Zilbovicius M (2005) Autism: functional brain mapping of exceptional calendar capacity. *Br J Psychiatry* 187(1):83–86. <https://doi.org/10.1192/bjp.187.1.83>
3. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, Donohue MR, Foran W, Miller RL, Hendrickson TJ, Malone SM, Kandala S, Feczko E, Miranda-Dominguez O, Graham AM, Earl EA, Perrone AJ, Cordova M, Doyle O, Moore LA, Conan GM, Uriarte J, Snider K, Lynch BJ, Wilgenbusch JC, Pengo T, Tam A, Chen J, Newbold DJ, Zheng A, Seider NA, Van AN, Metoki A, Chauvin RJ, Laumann TO, Greene DJ, Petersen SE, Garavan H, Thompson WK, Nichols TE, Yeo BTT, Barch DM, Luna B, Fair DA, Dosenbach NUF (2022) Reproducible brain-wide association studies require thousands of individuals. *Nature* 603(7902):654–660. <https://doi.org/10.1038/s41586-022-04492-9>
4. The Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguénay Youth Study (SYS) Group, Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Renteria ME, Toro R, Jahanshad N, Schumann G, Franke B, Wright MJ, Martin NG, Agartz I, Alda M, Alhusaini S, Almasy L, Almeida J, Alpert K, Andreasen NC, Andreassen OA, Apostolova LG, Appel K, Armstrong NJ, Aribisala B, Bastin ME,

- Bauer M, Bearden CE, Bergmann Ø, Binder EB, Blangero J, Bockholt HJ, Bøen E, Bois C, Boomsma DI, Booth T, Bowman IJ, Bralten J, Brouwer RM, Brunner HG, Brohawn DG, Buckner RL, Buitelaar J, Bulayeva K, Bustillo JR, Calhoun VD, Cannon DM, Cantor RM, Carless MA, Caseras X, Cavalleri GL, Chakravarty MM, Chang KD, CRK C, Christoforou A, Cichon S, Clark VP, Conrod P, Coppola G, Crespo-Facorro B, Curran JE, Czisch M, Deary IJ, EJC DG, Den Braber A, Delvecchio G, Depondt C, De Haan L, De Zubizaray GI, Dima D, Dimitrova R, Djurovic S, Dong H, Donohoe G, Duggirala R, Dyer TD, Ehrlich S, Ekman CJ, Elväsåhagen T, Emsell L, Erk S, Espeseth T, Fagerness J, Fears S, Fedko I, Fernández G, Fisher SE, Foroud T, Fox PT, Francks C, Frangou S, Frey EM, Frodl T, Frouin V, Garavan H, Giddaluru S, Glahn DC, Godlewska B, Goldstein RZ, Gollub RL, Grabe HJ, Grimm O, Gruber O, Guadalupe T, Gur RE, Gur RC, Göring HHH, Hagenaars S, Hajek T, Hall GB, Hall J, Hardy J, Hartman CA, Hass J, Hatton SN, Haukvik UK, Hegenscheid K, Heinz A, Hickie IB, Ho B-C, Hoehn D, Hoekstra PJ, Hollinshead M, Holmes AJ, Homuth G, Hoogman M, Hong LE, Hosten N, Hottenga J-J, Hulshoff Pol HE, Hwang KS, Jack CR, Jenkinson M, Johnston C, Jönsson EG, Kahn RS, Kasperaviciute D, Kelly S, Kim S, Kochunov P, Koenders L, Krämer B, JBJ K, Lagopoulos J, Laje G, Landen M, Landman BA, Lauriello J, Lawrie SM, Lee PH, Le Hellard S, Lemaître H, Leonardo CD, Li C, Liberg B, Liewald DC, Liu X, Lopez LM, Loth E, Lourdasamy A, Luciano M, Macciardi F, Machielsen MWJ, MacQueen GM, Malt UF, Mandl R, Manoach DS, Martinot J-L, Matarin M, Mather KA, Mattheisen M, Mattingsdal M, Meyer-Lindenberg A, McDonald C, McIntosh AM, McMahon FJ, McMahon KL, Meisenzahl E, Melle I, Milaneschi Y, Mohnke S, Montgomery GW, Morris DW, Moses EK, Mueller BA, Muñoz Maniega S, Mühleisen TW, Müller-Myhsok B, Mwangi B, Nauck M, Nho K, Nichols TE, Nilsson L-G, Nugent AC, Nyberg L, Olvera RL, Oosterlaan J, Ophoff RA, Pandolfo M, Papalampropoulou-Tsiridou M, Papmeyer M, Paus T, Pausova Z, Pearlson GD, Penninx BW, Peterson CP, Pfennig A, Phillips M, Pike GB, Poline J-B, Potkin SG, Pütz B, Ramasamy A, Rasmussen J, Rietschel M, Rijpkema M, Risacher SL, Roffman JL, Roiz-Santiañez R, Romanczuk-Seiferth N, Rose EJ, Royle NA, Rujescu D, Ryten M, Sachdev PS, Salami A, Satterthwaite TD, Savitz J, Saykin AJ, Scanlon C, Schmaal L, Schnack HG, Schork AJ, Schulz SC, Schür R, Seidman L, Shen L, Shoemaker JM, Simmons A, Sisodiya SM, Smith C, Smoller JW, Soares JC, Sponheim SR, Sprooten E, Starr JM, Steen VM, Strakowski S, Strike L, Sussmann J, Sämann PG, Teumer A, Toga AW, Tordesillas-Gutierrez D, Trabzuni D, Trost S, Turner J, Van Den Heuvel M, Van Der Wee NJ, Van Eijk K, Van Erp TGM, Van Haren NEM, Van 't Ent D, Van Tol M-J, Valdés Hernández MC, Veltman DJ, Versace A, Völzke H, Walker R, Walter H, Wang L, Wardlaw JM, Weale ME, Weiner MW, Wen W, Westlye LT, Whalley HC, Whelan CD, White T, Winkler AM, Wittfeld K, Woldehawariat G, Wolf C, Zilles D, Zwiers MP, Thalamuthu A, Schofield PR, Freimer NB, Lawrence NS, Drevets W (2014) The ENIGMA consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav* 8(2):153–182. <https://doi.org/10.1007/s11682-013-9269-5>
5. Vernooij MW, De Groot M, Bos D (2016) Population imaging in neuroepidemiology. In: *Handbook of clinical neurology*. Elsevier, pp 69–90
 6. Van Dijk H, Van Wingen G, Denys D, Olbrich S, Van Ruth R, Arns M (2022) The two decades brainclinics research archive for insights in neuropsychology (TDBRAIN) database. *Sci Data* 9(1):333. <https://doi.org/10.1038/s41597-022-01409-z>
 7. Mascarell Maričić L, Walter H, Rosenthal A, Ripke S, Quinlan EB, Banaschewski T, Barker GJ, Bokde ALW, Bromberg U, Büchel C, Desrivieres S, Flor H, Frouin V, Garavan H, Itterman B, Martinot J-L, Martinot M-LP, Nees F, Orfanos DP, Paus T, Poustka L, Hohmann S, Smolka MN, Fröhner JH, Whelan R, Kaminski J, Schumann G, Heinz A, IMAGEN consortium, Albrecht L, Andrew C, Arroyo M, Artiges E, Aydin S, Bach C, Banaschewski T, Barbot A, Barker G, Boddaert N, Bokde A, Bricaud Z, Bromberg U, Bruehl R, Büchel C, Cachia A, Cattrell A, Conrod P, Constant P, Dalley J, Decideur B, Desrivieres S, Fadaï T, Flor H, Frouin V, Gallinat J, Garavan H, Briand FG, Gowland P, Heinrichs B, Heinz A, Heym N, Hübner T, Ireland J, Ittermann B, Jia T, Lathrop M, Lanzerath D, Lawrence C, Lemaître H, Lüdemann K, Macare C, Mallik C, Mangin J-F, Mann K, Martinot J-L, Mennigen E, De Carvahlo FM, Mignon X, Miranda R, Müller K, Nees F, Nymberg C, Paillere M-L, Paus T, Pausova Z, Poline J-B, Poustka L, Rapp M, Robert G, Reuter J, Rietschel M, Ripke S, Robbins T, Rodehacke S, Rogers J, Romanowski A,

- Ruggeri B, Schmäl C, Schmidt D, Schneider S, Schumann M, Schubert F, Schwartz Y, Smolka M, Sommer W, Spanagel R, Speiser C, Spranger T, Stedman A, Steiner S, Stephens D, Strache N, Ströhle A, Struve M, Subramaniam N, Topper L, Walter H, Whelan R, Williams S, Yacubian J, Zilbovicius M, Wong CP, Lubbe S, Martinez-Medina L, Fernandes A, Tahmasebi A (2020) The IMAGEN study: a decade of imaging genetics in adolescents. *Mol Psychiatry* 25(11):2648–2671. <https://doi.org/10.1038/s41380-020-0822-5>
8. Parra-Rodriguez MA, Prado P, Moguilner S, Herzog RA, Birba A, Santamaria-García HA, Tagliazucchi E, Reyes PA, Cruzat JA, García AM, Whelan R, Lopera F, Ochoa JF, Anghinah R, Ibáñez A (2022) The EuroLaD-EEG consortium: towards a global EEG platform for dementia, for seeking to reduce the regional impact of dementia. *Alzheimers Dement* 18(S6):e059944. <https://doi.org/10.1002/alz.e059944>
 9. Gorgolewski KJ, Poldrack RA (2016) A practical guide for improving transparency and reproducibility in neuroimaging research. *PLoS Biol* 14(7):e1002506. <https://doi.org/10.1371/journal.pbio.1002506>
 10. Boccia M, Piccardi L, Guariglia P (2015) The meditative mind: a comprehensive meta-analysis of MRI studies. *Biomed Res Int* 2015:1–11. <https://doi.org/10.1155/2015/419808>
 11. Kochunov P, Jahanshad N, Sprooten E, Nichols TE, Mandl RC, Almasy L, Booth T, Brouwer RM, Curran JE, De Zubicaray GL, Dimitrova R, Duggirala R, Fox PT, Elliot Hong L, Landman BA, Lemaître H, Lopez LM, Martin NG, McMahon KL, Mitchell BD, Olvera RL, Peterson CP, Starr JM, Sussmann JE, Toga AW, Wardlaw JM, Wright MJ, Wright SN, Bastin ME, McIntosh AM, Boomsma DI, Kahn RS, Den Braber A, De Geus EJC, Deary IJ, Hulshoff Pol HE, Williamson DE, Blangero J, Van'T Ent D, Thompson PM, Glahn DC (2014) Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and mega-analytical approaches for data pooling. *NeuroImage* 95:136–150. <https://doi.org/10.1016/j.neuroimage.2014.03.033>
 12. Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, Hardcastle N, Wexler J, Esteban O, Goncalves M, Jwa A, Poldrack R (2021) The OpenNeuro resource for sharing of neuroscience data. *eLife* 10:e71774. <https://doi.org/10.7554/eLife.71774>
 13. Delorme A, Truong D, Youn C, Sivagnanam S, Stirn C, Yoshimoto K, Poldrack RA, Majumdar A, Makeig S (2022) NEMAR: an open access data, tools and compute resource operating on neuroelectromagnetic data. *Database* 2022:baac096. <https://doi.org/10.1093/database/baac096>
 14. Wagner AS, Waite LK, Wierzbza M, Hoffstaedter F, Waite AQ, Poldrack B, Eickhoff SB, Hanke M (2022) FAIRly big: a framework for computationally reproducible processing of large-scale data. *Sci Data* 9(1):80. <https://doi.org/10.1038/s41597-022-01163-2>
 15. Leemans A, Jeurissen B, Sijbers J, Jones DK (2009) ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. p 3537
 16. Whelan R, Barbey FM, Cominetti MR, Gillan CM, Rosická AM (2022) Developments in scalable strategies for detecting early markers of cognitive decline. *Transl Psychiatry* 12(1):473
 17. Niso G, Botvinik-Nezer R, Appelhoff S, De La Vega A, Esteban O, Etzel JA, Finc K, Ganz M, Gau R, Halchenko YO (2022a) Open and reproducible neuroimaging: from study inception to publication. *NeuroImage* 263:119623
 18. Madan CR (2022) Scan once, analyse many: using large open-access neuroimaging datasets to understand the brain. *Neuroinformatics* 20(1):109–137

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

