Stefano Moia[1, 2, 3], Hao-Ting Wang[1, 4], Anibal S. Heinsfeld[5, 6], Dorota Jarecka[7], Yu-Fang Yang[8], Stephan Heunis[9], Michele Svanera[10], Benjamin De Leneer[1, 11, 12, 13], Andrea Gondova[1, 14, 15], Sin Kim[16], Arshitha Basavaraj[17], Johanna M. Bayer[1, 18, 19], Roza G. Bayrak[1, 20], Pierre-Louis Bazin[21, 22], Isil Poyraz Bilgin[4, 23], Steffen Bollmann[24, 25], Daniel Borek[26], Valentina Borghesani[1, 27], Trang Cao[28], Gang Chen[29], Alejandro De La Vega[30], Sebastian Dresbach[31], Philipp Ehses[32], Jan Ernsting[33, 34, 35], Inês Esteves[36], Oscar Ferrante[37], Kelly G. Garner[38], Rémi Gau[39], Elodie Germani[40], Tara Ghafari[37], Satrajit S. Ghosh[7, 41], Sarah Elizabeth Goodale[42], Cassandra D. Gould van Praag[43, 44], Samuel Guay[45], Omer Faruk Gulban[46, 31], Yaroslav O. Halchenko[47], Michael Hanke[9, 48], Peer Herholz[49], Katja Heuer[50], Felix Hoffstaedter[9, 51], Ruoqi Huang[52], Renzo Huber[53], Ole Jensen[37], Kan Keeratimahat[54], Julian Q. Kosciessa[55], Sladjana Lukic[56], Neville Magielse[9, 57, 58], Christopher J. Markiewicz[59], Caroline G. Martin[60], Camille Maumet[61], Anna Menacher[62], Jeff Mentch[7, 63], Christian Mönch[9], Shammi More[9, 48], Leonardo Muller-Rodriguez[64, 65], Samuel A. Nastase[66], Eliana Nicolaisen-Sobesky[9], Dylan M. Nielson[67], Christopher R. Nolan[38], François Paugam[68, 69], Pedro Pinheiro-Chagas[70], Ana Luísa Pinho[71, 72], Alessandra Pizzuti[53, 46], Benjamin Poldrack[64], Benedikt A. Poser[53], Roberta Rocca[73], Jacob Sanz-Robinson[74], Kelvin Sarink[34], Kevin R. Sitek[75], Nadine Spychala[76], Rüdiger Stirnberg[32], Michał Szczepanik[9], Mohammad Torabi[77], Roberto Roberto Toro[50], Sebastian GW Urchs[78], Sofie L. Valk[22, 9, 81], Adina S. Wagner[9, 80], Laura K. Waite[9], Alexander Q. Waite[9], Lea Waller[81], Tyler James Wishard[82, 83], Jianxiao Wu[9, 48], Yuchen Zhou[84], Janine D. Bijsterbosch[1, 85], The physiopy community[86],

1. OHBM Open Science Special Interest Group (SIG); 2. Neuro-X Institute, École polytechnique fédérale de Lausanne, Geneva, Switzerland; 3. Department of Radiology and Medical Informatics (DRIM), Faculty of Medicine, University of Geneva, Geneva, Switzerland; 3. Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Montréal, QC, Canada; 4. Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Montreal, Quebec, Canada; 5. Department of Psychology, Center for Perceptual Systems, The University of Texas at Austin, Austin, TX, USA; 6. Department of Computer Science, The University of Texas at Austin, Austin, TX, USA; 7. McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA; 8. Division of Experimental Psychology and Neuropsychology, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany; 9. Institute of Neuroscience and Medicine, Research Centre Jülich, Jülich, Germany; 10. Centre for Cognitive Neuroimaging, School of Psychology Neuroscience University of Glasgow, UK; 11. Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC,

CA; 12. CHU Sainte-Justine Research Center, University of Montreal, Montreal, QC, CA; 13. Department of Computer Engineering and Software Engineering, Polytechnique Montreal, Montreal, CA; 14. UNIACT, NeuroSpin, CEA, Université Paris-Saclay, Paris, France; 15. NeuroDiderot, Inserm, Université Paris Cité, Paris, France; 16. Brain and Cognitive Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea; 17. Data Science and Sharing Team (DSST), Intramural Research Program, NIMH, Bethesda, USA; 18. Centre for Youth Mental Health, the University of Melbourne, Australia; 19. Orygen, Melbourne, Australia; 20. Department of Computer Science, Vanderbilt University, Nashville TN, USA; 21. University of Amsterdam, Amsterdam, Netherlands; 22. Max Planck Institute for Human Cognitive and Brain Science, Leipzig, Germany; 23. Machine Learning Department and the Neuroscience Institute, Carnegie Mellon University, Pittsburgh, US; 24. School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia; 25. Centre for Innovation in Biomedical Imaging Technology, University of Queensland, Brisbane, Australia; 26. Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium; 27. Department of Psychology, Faculty of Psychology and Science of Education, University of Geneva, Geneva, Switzerland; 28. School of Psychological Sciences, Monash University, Clayton, VIC, Australia; 29. Scientific and Statistical Computing Core, NIMH, NIH, Department of Health and Human Services, USA; 30. Department of Psychology, University of Texas at Austin, Austin, TX, USA; 31. Maastricht University, Maastricht, Netherlands; 32. German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; 33. Institute for Geoinformatics, University of Münster, Germany; 34. University of Münster, Institute for Translational Psychiatry, Münster, Germany; 35. Faculty of Mathematics and Computer Science, University of Münster, Germany; 36. ISR-Lisboa and Department of Bioengineering, Instituto Superior Técnico – Universidade de Lisboa, Lisbon, Portugal; 37. Centre for Human Brain Health, School of Psychology, University of Birmingham, UK; 38. School of Psychology, The University of New South Wales, NSW, Australia; 39. Institut de recherche en sciences psychologique, Université catholique de Louvain. Louvain la neuve, Belgique; 40. Univ Rennes, Inria, CNRS, Inserm, Rennes, France; 41. Department of Otolaryngology - Head and Neck Surgery, Harvard Medical School, Boston, MA, USA; 42. Department of Biomedical Engineering, Vanderbilt University; 43. Alan Turing Institute, London, UK; 44. Department of Psychiatry, Oxford University, Oxford, UK; 45. Department of Psychology, University of Montreal, Canada; 46. Brain Innovation, Maastricht, The Netherlands; 47. Department of Psychological and Brain Sciences, Dartmouth college, New Hampshire, USA; 48. Medical Faculty, Heinrich Heine University Düsseldorf; 49. McGill, Montreal Neurological Institute - Hospital, Montreal, Quebec, Canada; 50. Institut Pasteur, Université Paris Cité, Unité de Neuroanatomie Appliquée et Théorique, F-75015 Paris, France; 51. Institute of Systems Neuroscience, Medical Faculty University Hospital Düsseldorf, Heinrich-Heine- University, Düsseldorf, Germany; 52. University of Toronto, Toronto, Ontario, Canada; 53. Department of

Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands; 54. Nuffield Department of Population Health, University of Oxford, Oxford, UK; 55. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; 56. Ruth S. Ammon College of Education and Health Sciences, Department of Communication Sciences and Disorders, Adelphi University, San Francisco, CA, USA; 57. Otto Hahn Cognitive Neurogenetics Group, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1A, 04103, Leipzig, Germany; 58. Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Moorenstraße 5 40225 Düsseldorf, Germany; 59. Stanford University, CA, USA; 60. Vanderbilt University, Nashville, TN USA; 61. Inria, Univ Rennes, CNRS, Inserm, Rennes, France; 62. Department of Statistics, University of Oxford, UK; 63. Program in Speech and Hearing Bioscience and Technology, Harvard University, Cambridge, MA, USA; 64. Psychoinformatics Lab, Forschungszentrum Julich, Julich, Germany; 65. Cambridge Institute for Music Therapy Research (CIMTR), Anglia Ruskin University, Cambridge, UK; 66. Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA; 67. Machine Learning Team, National Institute of Mental Health, Bethesda, MD, USA; 68. University of Montreal, Montreal, Quebec, Canada; 69. Mila - Quebec Artificial Intelligence Institute, University of Montreal, Montreal, Quebec, Canada; 70. Department of Neurology, Memory and Aging Center, University of California, San Francisco, San Francisco, CA, USA; 71. Western Institute for Neuroscience, Western University, London, Ontario, Canada; 72. Department of Computer Science, Faculty of Science, Western University, London, Ontario, Canada; 73. Department of Culture, Cognition, and Computation, School of Culture and Society, Aarhus University, Aarhus, Denmark; 74. Neuroscience, McGill University, Quebec, Canada; 75. Department of Communication Science and Disorders, School of Health and Rehabilitation Sciences, University of Pittsburgh, Pittsburgh, PA, USA; 76. Department of Informatics, University of Sussex, UK; 77. Graduate Program in Biological and Biomedical Engineering, McGill University, Montreal, Canada; 78. NeuroDataSciene - ORIGAMI lab, Montreal Neurological Institute-Hospital, McGill University, Montreal, Canada; 79. Institute of Systems Neuroscience Heinrich Heine University Düsseldorf; 80. Department of Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany ; 81. Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Psychiatry and Neurosciences CCM, Berlin, Germany; 82. Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, California, United States; 83. Department of Neurology, Washington University in St. Louis, St. Louis, Missouri, United States; 84. Department of Psychology, Carnegie Mellon University, Pittsburgh, USA; 85. Department of Radiology, Washington University in St Louis, Saint Louis, Missouri, USA; 86. physiopy.github.io

Proceedings of the OHBM Brainhack 2022

Abstract

OHBM Brainhack 2022 took place in June 2022. The first hybrid OHBM hackathon, it had an in-person component taking place in Glasgow and three hubs around the globe to improve inclusivity and fit as many timezones as possible. In the buzzing setting of the Queen Margaret Union and of the virtual platform, 23 projects were presented after development. Following are the reports of 14 of those, as well as a recapitulation of the organisation of the event.

Introduction

The Organisation of Human Brain Mapping BrainHack (shortened to OHBM Brainhack herein) is a yearly satellite event of the main OHBM meeting, organised by the Open Science Special Interest Group following the model of Brainhack hackathons[1]. Where other hackathons set up a competitive environment based on outperforming other participants' projects, Brainhacks foster a collaborative environment in which participants can freely collaborate and exchange ideas within and between projects.

This edition of the OHBM Brainhack, that ran across the world over four days, was particularly special for two reasons: it celebrated the tenth year anniversary of Brainhack, and, like the main OHBM conference, it was the first edition to feature an in-person event after two years of virtual events. For this reasons, the whole organisation rotated around five main principles:

1. Providing a hybrid event incorporating the positive aspects of in-person and virtual events alike,

2. Celebrating the 10^{th} anniversary of the Brainhack by bringing back newcomer-friendly hands-on hacking and learning experience, enhancing the Hacktrack and

formatting the Traintrack as a collection of materials to consult beforehand and as spontaneous meetings of the participants aimed to learn together,

3. Bridging the gap between the Brainhack community and the main neuroimaging software developer groups, e.g. AFNI, FSL, SPM,

4. Due to amount of work required to meet the previous three principles, incorporating from the beginning a team of core organisers with a democratic approach to organisation, with a member in charge of an aspect of the event,

5. Brainhack event organisation should always be experimental, trying different solutions and formats to find a way to improve Brainhack events overall.

After a quick explanation of each main contribution of the core team, the next pages are dedicated to the summaries of the projects that were developed during the four days of hacking.

1 Hacktrack

Dorota Jarecka, Yu-Fang Yang, Hao-Ting Wang, Stefano Moia

The key component of each Brainhack is hacking. The hacking part, known as hacktrack, is where attendees collaborate on projects and explore their own ideas. There are 4 elements of hacktrack that were organised: project submission, project pitch, hacking period and project summary. For the project submission, we used the GitHub issue submission process that was used during recent years. We updated and simplified a project template from previous years and asked project leaders to open an issue for each project. Each issue after quick check was approved by the moderators and automatic workflows written by the team were responsible for sending project descriptions to the Brainhack page and setting Discord's channels. We received 38 projects that were submitted using this system. The project pitch was set for the morning of the first day and everyone had 2 minutes to talk about the suggested project and possible collaborations. After the pitches people had a chance to talk to each other and join the projects they were interested in. This year, we tried to maximise the time for hacking by providing a sparse schedule for talks. The closing ceremony of the Brainhack featured 23 project reports, in which teams talked about their experiences and described the work they accomplished.

This edition we allowed remote attendance from other locations. We organised three hubs aiming to cover all time zones, including 1) Asia-Pacific, 2) Glasgow, Europe, Middle East, and Africa, and 3) the Americas, to foster inclusiveness in the hybrid conference format. We also ensured that each hub had one live streamed session with the physical hub in Glasgow.

2 Traintrack

Yu-Fang Yang, Dorota Jarecka

Traintrack is the educational component of Brainhack events. The aim is to introduce tools and skills for attendees to start hacking. Unlike conventional scientific educational workshops centred around lectures and talks, data science skills are better learned through hands-on experience than lectures. With the Brainhack community growing mature, the community has developed their own curated educational material. *Brainhack School* has supplied high-quality content for independent study on a variety of themes.

This year, we combine the collaborative nature of brainhack projects and educational content to reimagine the format of traintack. Thus, we replaced tutorial lectures in the previous editions with curated online educational contents, released them prior to the main event, and attempted to integrate them with the hacktrack projects. This format also provides more time (i.e. schedule) and space (i.e. minimising large space not used for hacking) for attendees to self-organise. Participants were encouraged to form study groups on five suggested topics: 1) setting up your system for analysis 2) python for data analysis, 3) machine learning for neuroimaging, 4) version control systems, 5) cloud resource. The curated content was advertised on the main hackathon website. One dedicated channel was created on the hackathon Discord server. Individuals could determine the nature of their experiences and the skills they liked to acquire. Participants could form their own study group and on any selected topic. We would like to continue the experimentation on this format in the coming year.

3 Platforms, website, and IT

Anibal Solon Heinsfeld

Trying to bring a positive experience for both virtual and in-person attendees, we implemented several integrated solutions to ease communication in the different phases of the Hackathon, focusing on a single platform for the main event.

The first solution was the project's advertisement, in which the community promotes their projects, the goals for the Hackathon, and relevant information to get people interested and set to collaborate. To do so, we used the Github Issues feature in the Hackathon repository as the entrance for projects. Github Issues has been proven to be accepted by the community that relies on Github for code versioning, and was a successful approach in past hackathons.

In this edition, we were able to use Github Issue forms, a beta feature in Github. Past use of issues for project registration relies on Markdown code to specify which information the hacker needs to provide. However, the code can be easily broken and changed, which makes it harder to parse the information in automated setups. Towards this issue, the Issue Form can lower the barrier when submitting a project. By specifying form fields for the participants to fill, they faced a common web form instead of a Markdown editor, bringing more structure to their inputs and not requiring them to write code. After the organisers' quick validation, the project information was provided to the rest of the system. Per an automated pipeline, this information was compiled into the website.

The second solution was the central platform for real-time communication, namely Discord. For the first time using the platform for an OHBM Hackathon, Discord showed potential in bringing an all-in-one solution. Its track record with different communities and their formats was an essential prospect for the success of a hybrid hackathon, together with the different ways of communicating provided by the platform. Specifically, Discord offered chat and audio/video channels, with fine-tuned controls on permissions to see a channel, speak and use the camera, and send messages. With these features, we were able to create experiences for the attendants, such as text channels for consolidating information about the hackathon, main stages controlled by the hub hosts, a channel to join projects and hubs, and integrated text & voice channels for each project. The main stage was connected to a laptop in the venue, providing synchronous streaming for announcements, project pitches and progress reports for those participating virtually. The project channels were automatically created together with the Github Issues. However, given the thriving number of projects, the Discord server was replete with project channels. Such a scenario was overwhelming for the attendants, especially for those approaching Discord for the first time. To ameliorate this issue, a main projects channel was created, so attendants could automatically join projects via related emoji reactions. The project channels were of public access; however, only displayed upon joining the project. Besides initial technical hiccups, the platform proved a good alternative for such an event format.

These integrated solutions smoothed the organisation of the event, the virtual platform provided great support for the on-line participants. However, there was not a lot of interaction between in-person and online participants, and projects were mainly either virtual or in-person (with few exceptions). This is probably because hybrid hacking provides challenges for organisation and attendants alike, even just in the physical limitations of being able to have a video conference with a split team. It is important to consider, however, that this was also the first in-person event for many participants, who preferred in-person interaction and collaboration rather than the same on-line interaction that characterised such events in the previous two years.

4 Project Reports

The peculiar nature of a Brainhack[1] reflects in the nature of the projects developed during the event, that can span very different types of tasks. While most projects feature more 'hackathon-style' software development, in the form of improving software integration (Section 4.4), API refactoring (Section 4.11), or creation of new toolboxes and platforms (Sections 4.9, 4.10 and 4.13), the inclusion of newcomers and participants with less strong software development skills can foster projects oriented to user testing (Sections 4.3 and 4.9) or documentation compilation (Section 4.12). The scientific scopes of Brainhacks were reflected in projects revolving around data exploration (Sections 4.1 and 4.7) or model development (Section 4.13), or adding aspects of open science practices (namely, the Brain Imaging Data Structure) to toolboxes (Sections 4.6 and 4.14). Finally, fostering a collaborative environment and avoiding pitching projects against each others not only opens up the possibility for participants to fluidly move between different groups, but also to have projects which sole aim is supporting other projects (Section 4.2), learning new skills by having fun (Section 4.5), or fostering discussions and conversations among participants to improve the adoption of open science practices (Section 4.8).

Following are the 14 submitted reports of the 23 projects presented at the OHBM Brainhack.

4.1 Exploring the AHEAD brains together

Alessandra Pizzuti, Sebastian Dresbach, Satrajit Ghosh, Katja Heuer, Roberto Toro, Pierre-Louis Bazin

4.1.1 Introduction

One of the long-standing goals of neuroanatomy is to compare the cyto- and myeloarchitecture of the human brain. The recently made available 3D whole-brain post-mortem data set provided by Alkemade and colleagues[2] includes multiple microscopy contrasts and 7-T quantitative multi-parameter MRI reconstructed at 200µm from two human brains. Through the co-registration across MRI and microscopy modalities, this data set provides a unique direct comparison between histological markers and quantitative MRI parameters for the same human brain. In this BrainHack project, we explored this dataset, focusing on: (i) data visualization in online open science platforms, (ii) data integration of quantitative MRI with microscopy, (iii) data analysis of cortical profiles from a selected region of interest.

4.1.2 Results

Visualization and annotation of large neuroimaging data sets can be challenging, in particular for collaborative data exploration. Here we tested two different infrastructures: BrainBox `https://brainbox.pasteur.fr/`, a web-based visualization and annotation tool for collaborative manual delineation of brain MRI data, see e.g. Heuer and colleagues[3], and Dandi Archive `https://dandiarchive.org/`, an online repository of microscopy data with links to Neuroglancer `https://github.com/google/neuroglancer`. While Brainbox could not handle the high resolution data well, Neuroglancer visualization was successful after conversion to the Zarr microscopy format (Figure 1A).

To help users explore the original high-resolution microscopy sections, we also built a python notebook to automatically query the stains around a given MNI coordinate using the Nighres toolbox[4] (Figure 1B).

For the cortical profile analysis we restricted our analysis on S1 (BA3b) as a part of the somato-motor area from one hemisphere of an individual human brain. S1 is rather thin (∼2mm) and it has a highly myelinated layer 4 (see arrow Figure 1C). In a future step, we are aiming to characterize differences between S1 (BA3b) and M1 (BA4). For now, we used the MRI-quantitative-R1 contrast to define, segment the region of interest and compute cortical depth measurement. In ITK-SNAP[5] we defined the somato-motor area by creating a spherical mask (radius 16.35mm) around the 'hand knob' in M1. To improve the intensity homogeneity of the qMRI-R1 images, we ran a bias field correction (N4BiasFieldCorrection,[6]). Tissue segmentation was restricted to S1 and was obtained by combining four approaches: (i) fsl-fast[7] for initial tissues probability map, (ii) semi-automatic histogram fitting in ITK-SNAP, (iii) Segmentator[8], and (iv) manual editing. We used the LN2_LAYERS program from LAYNII open source software[9] to compute the equi-volume cortical depth measurements for the gray matter. Finally, we evaluated cortical depth profiles for three quantitative MRI contrasts (R1, R2, proton density) and three microscopy contrasts (thionin, bieloschowsky, parvalbumin) by computing a voxel-wise 2D histogram of image intensity (Figure 1C). Some challenges are indicated by arrows 2 and 3 in the lower part of Figure 1C.

From this Brainhack project, we conclude that the richness of the data set must be exploited from multiple points of view, from enhancing the integration of MRI with microscopy data in visualization software to providing optimized multi-contrast and multi-modality data analysis pipeline for high-resolution brain regions.

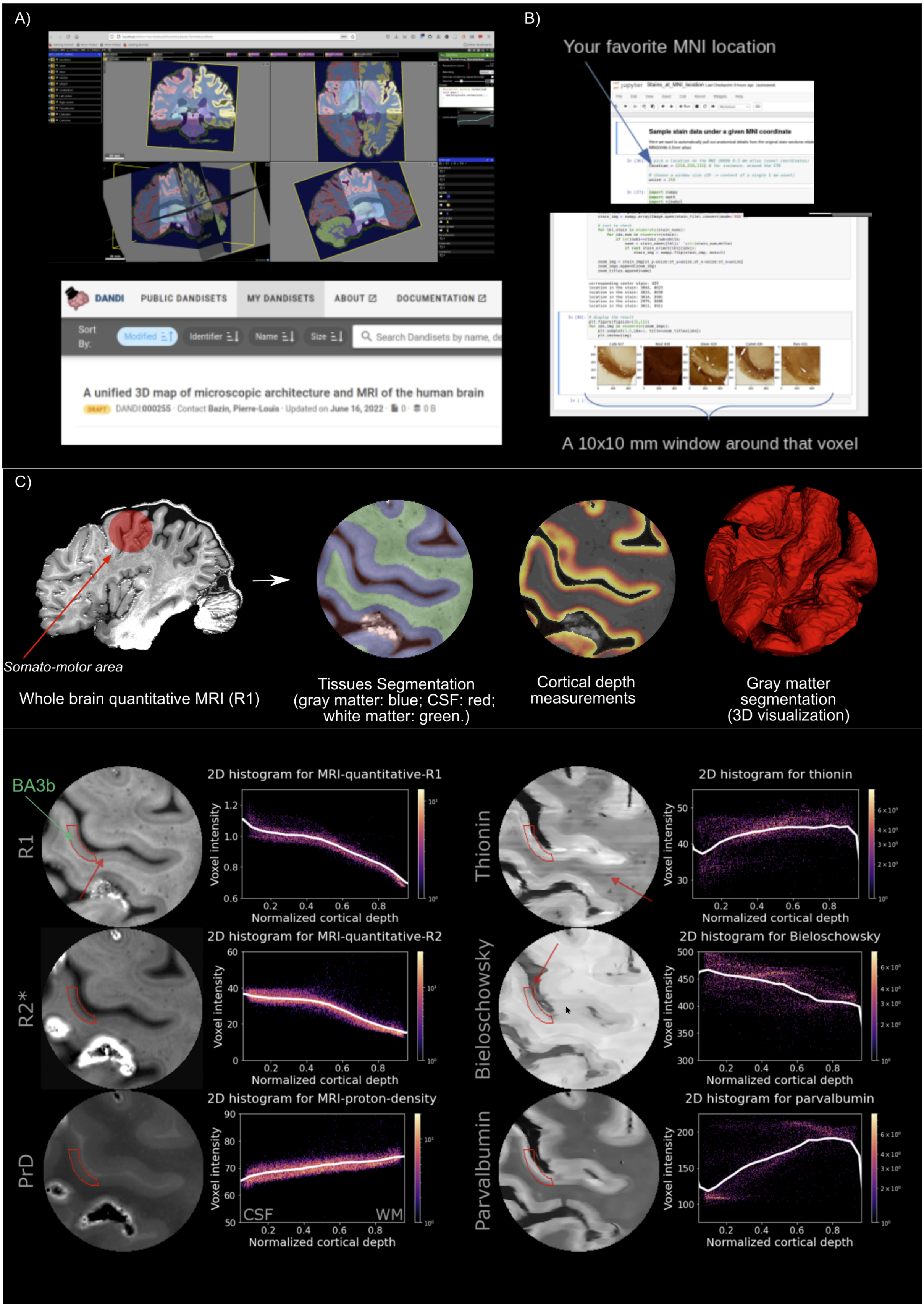


Figure 1: A) Neuroglancer visualization, B) section query notebook, C) Cortical ROI and corresponding depth histograms extracted from the different contrasts available.

4.2 Brainhack Cloud

Steffen Bollmann, Isil Poyraz Bilgin, Peer Herholz, Rémi Gau, Samuel Guay, Johanna Bayer

Today's neuroscientific research deals with vast amounts of electrophysiological, neuroimaging and behavioural data. The progress in the field is enabled by the widespread availability of powerful computing and storage resources. Cloud computing in particular offers the opportunity to flexibly scale resources and it enables global collaboration across institutions. However, cloud computing is currently not widely used in the neuroscience field, although it could provide important scientific, economical, and environmental gains considering its effect in collaboration and sustainability[10,11]. One problem is the availability of cloud resources for researchers, because Universities commonly only provide on-premise high performance computing resources. The second problem is that many researchers lack the knowledge on how to efficiently use cloud resources. This project aims to address both problems by providing free access to cloud resources for the brain imaging community and by providing targeted training and support.

A team of brainhack volunteers (`https://brainhack.org/brainhack_cloud/admins/team/`) applied for Oracle Cloud Credits to support open-source projects in and around brainhack with cloud resources. The project was generously funded by Oracle Cloud for Research[12] with $230,000.00 AUD from the 29th of January 2022 until the 28th of January 2024. To facilitate the uptake of cloud computing in the field, the team built several resources (`https://brainhack.org/brainhack_cloud/tutorials/`) to lower the entry barriers for members of the Brainhack community.

During the OHBM 2022 Brainhack, the team gave a presentation to share the capabilities that cloud computing offers to the Brainhack community, how they can place their resource requests and where they can get help. In total 11 projects were onboarded to the cloud and supported in their specific use cases: One team utilised the latest GPU architecture to take part in the Anatomical Tracings of Lesions After Stroke Grand Challenge. Others developed continuous integration tests for their tools using for example a full Slurm HPC cluster in the cloud to test how their tool behaves in such an environ-

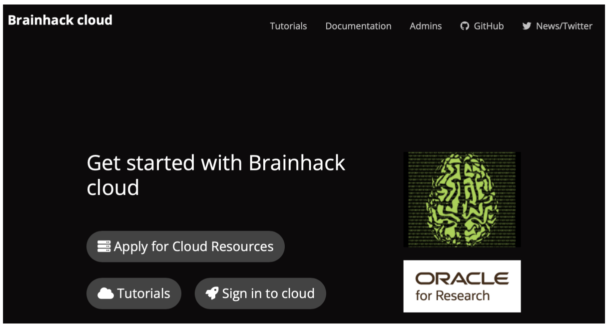


Figure 2: A team of brainhack volunteers, applied for Oracle Cloud Credits to support open source projects in and around brainhack with powerful cloud resources on the Oracle Cloud: `https://brainhack.org/brainhack_cloud/`

ment. Another group deployed the Neurodesk.org[13] project on a Kubernetes cluster to make it available for a student cohort to learn about neuroimage processing and to get access to all neuroimaging tools via the browser. All projects will have access to these cloud resources until 2024 and we are continuously onboarding new projects onto the cloud (`https://brainhack.org/brainhack_cloud/docs/request/`).

The Brainhack Cloud team plans to run a series of training modules in various Brainhack events throughout the year to reach researchers from various backgrounds and increase their familiarity with the resources provided for the community while providing free and fair access to the computational resources. The training modules will cover how to use and access computing and storage resources (e.g., generating SSH keys), to more advanced levels covering the use of cloud native technology like software containers (e.g., Docker/Singularity), container orchestration (e.g., Kubernetes), object storage (e.g, S3), and infrastructure as code (e.g., Terraform).

4.3 DataLad Catalog

Stephan Heunis, Adina S. Wagner, Alexander Q. Waite, Benjamin Poldrack, Christian Mönch, Julian Kosciessa, Laura Waite, Leonardo Muller-Rodriguez, Michael Hanke, Michał Szczepanik, Remi Gau, Yaroslav O. Halchenko

The importance and benefits of making research data Findable, Accessible, Interoperable, and Reusable are clear[14]. But of equal importance is our ethical and legal obligations to protect the personal data privacy of research participants. So we are struck with this apparent contradiction: how can we share our data openly... yet keep it secure and protected?

To address this challenge: structured, linked, and machine-readable metadata presents a powerful opportunity. Metadata provides not only high-level information about our research data (such as study and data acquisition parameters) but also the descriptive aspects of each file in the dataset: such as file paths, sizes, and formats. With this metadata, we can create an abstract representation of the full dataset that is separate from the actual data content. This means that the content can be stored securely, while we openly share the metadata to make our work more FAIR.

In practice, the distributed data management system DataLad[15] and its extensions for metadata handling and catalog generation are capable of delivering such solutions. `datalad` (github.com/datalad/datalad) can be used for decentralised management of data as lightweight, portable and extensible representations. `datalad-metalad` (github.com/datalad/datalad-metalad) can extract structured high- and low-level metadata and associate it with these datasets or with individual files. And at the end of the workflow, `datalad-catalog` (`github.com/datalad/datalad-catalog`) can turn the structured metadata into a user-friendly data browser.

This hackathon project focused on the first round of user testing of the alpha version of `datalad-catalog`, by creating the first ever user-generated catalog (`https://jkosciessa.github.io/datalad_cat_test`). Further results included a string of new issues focusing on improving user experience, detailed notes on how to generate a catalog from scratch, and code additions to allow the loading of local web-assets so that any generated catalog can also be viewed offline.

4.4 DataLad-Dataverse integration

Benjamin Poldrack, Jianxiao Wu, Kelvin Sarink, Christopher J. Markiewicz , Alexander Q. Waite , Eliana Nicolaisen-Sobesky, Shammi More, Johanna Bayer, Jan Ernsting, Adina S. Wagner, Roza G. Bayrak , Laura K. Waite, Michael Hanke, Nadine Spychala

The FAIR principles[14] advocate to ensure and increase the Findability, Accessibility, Interoperability, and Reusability of research data in order to maximize their impact. Many open source software tools and services facilitate this aim. Among them is the Dataverse project[16]. Dataverse is open source software for storing and sharing research data, providing technical means for public distribution and archival of digital research data, and their annotation with structured metadata. It is employed by dozens of private or public institutions worldwide for research data management and data publication. DataLad[15], similarly, is an open source tool for data management and data publication. It provides Git- and git-annex based data versioning, provenance tracking, and decentral data distribution as its core features. One of its central development drivers is to provide streamlined interoperability with popular data hosting services to both simplify and robustify data publication and data consumption in a decentralized research data management system[17]. Past developments include integrations with the open science framework[18] or webdav-based services such as sciebo, nextcloud, or the European Open Science Cloud[19].

In this hackathon project, we created a proof-of-principle integration of DataLad with Dataverse in the form of the Python package `datalad-dataverse` (`github.com/datalad/datalad-dataverse`). From a technical perspective, main achievements include the implementation of a git-annex special remote protocol for communicating with Dataverse instances, a new `create-sibling-dataverse` command that is added to the DataLad command-line and Python API by the `datalad-dataverse` extension, and standard research software engineering aspects of scientific software such as unit tests, continuous integration, and documentation.

From a research data management and user perspective, this development equips DataLad users with the ability to programatically create Dataverse datasets (containers for research data and their metadata on Dataverse) from DataLad datasets (DataLad's Git-repository-based core data structure) in different usage modes. Subsequently, DataLad dataset contents, its version history, or both can be published to the Dataverse dataset via a 'datalad push' command. Furthermore, published DataLad datasets can be consumed from Dataverse with a `datalad clone` call. A mode parameter configures whether Git version history, version controlled file content, or both are published and determines which of several representations the Dataverse dataset takes. A proof-of-principle implementation for metadata annotation allows users to supply metadata in JSON format, but does not obstruct later or additional manual metadata annotation via Dataverse's web interface.

Overall, this project delivered the groundwork for further extending and streamlining data deposition and consumption in the DataLad ecosystem. With DataLad-Dataverse interoperability, users gain easy additional means for data publication, archival, distribution, and retrieval. Post-Brainhack development aims to mature the current alpha version of the software into an initial v0.1 release and distribute it via standard Python package indices.

4.5 Exploding brains in Julia

Ömer Faruk Gülban, Leonardo Muller-Rodriguez

Particle simulations are used to generate visual effects (in movies, games, etc.). In this project, we explore how we can use magnetic resonance imaging (MRI) data to generate interesting visual effects by using (2D) particle simulations. Aside from providing an entertaining avenue to the interested participants, our project has further educational utility. For instance, anatomical MRI data analysis is done in two major frameworks: (1) manipulating fixed regularly spaced points in space (also known as Eulerian point of view), and (2) manipulating moving irregularly spaced points in space (Lagrangian point of view). For instance, bias field correction is commonly done from Eulerian point of view (e.g. computing a bias field is similar to computing a particle velocity field in each frame of the explosions), whereas cortical surface inflation is commonly done from Lagrangian point of view of the MRI data (e.g. computing the inflated brain surface is similar to computing the new positions of particles in each frame of the explosion). Therefore, our project provides an educational opportunity for those who would like to peek into the deep computational and data structure manipulation aspects of MRI image analysis. We note that we already made two hackathon projects in 2020 (see below) and were first inspired by a blog post (`https://nialltl.neocities.org/articles/mpm_guide.html`) on the material point method[20,21,22]. Our additional aim in Brainhack 2022 is to convert our previous progress in Python programming language to Julia. The reason why we have moved to Julia language is because we wanted to explore this new programming language's potential for developing MRI image analysis methods as it has convenient parallelization methods that speeds-up the particle simulations (and any other advanced image manipulation algorithms).

Our previous efforts are documented at:

1. 2020 OpenMR Benelux: `https://github.com/OpenMRBenelux/openmrb2020-hackathon/issues/7`
2. 2020 OHBM Brainhack: `https://github.com/ohbm/hackathon2020/issues/124`
3. Available within the following github repository: `https://github.com/ofgulban/slowest-particle-simulator-on-earth`

As a result of this hackathon project, we delivered a video compilation of our animations (Figure 3) which can be seen at `https://youtu.be/_5ZDctWv5X4`. We highlight that in addition to its educational value, our project provided stress relief by means of entertaining the participants after the pandemic. We believe that our project provides a blueprint for the future brainhacks where MRI science, computation, and education can be disseminated within an engaging and entertaining context. Our future efforts will involve sophisticating the particle simulations, the initial simulation parameters to generate further variations of the visual effects, and potentially synchronizing the simulation effects with musical beats.

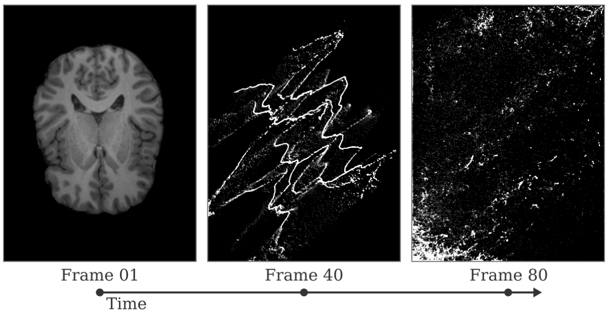


Figure 3: A video compilation of brain explosions can be seen at `https://youtu.be/_5ZDctWv5X4`.

4.6 FLUX: A pipeline for MEG analysis and beyond

Oscar Ferrante, Tara Ghafari, Ole Jensen

FLUX[23] is an open-source pipeline for analysing magnetoencephalography (MEG) data. There are several toolboxes developed by the community to analyse MEG data. While these toolboxes provide a wealth of options for analyses, the many degrees of freedom pose a challenge for reproducible research. The aim of FLUX is to make the analyses steps and setting explicit. For instance, FLUX includes the state-of-the-art suggestions for noise cancellation as well as source modelling including pre-whitening and handling of rank-deficient data.

So far, the FLUX pipeline has been developed for MNE-Python[24] and FieldTrip[25] with a focus on the MEGIN/Elekta system and it includes the associated documents as well as codes. The long-term plan for this pipeline is to make it more flexible and versatile to use. One key motivation for this is to facilitate open science with the larger aim of fostering the replicability of MEG research.

These goals can be achieved in mid-term objectives, such as making the FLUX pipeline fully BIDS compatible and more automated. Another mid-term goal is to containerize the FLUX pipeline and the associated dependencies making it easier to use. Moreover, expanding the applications of this pipeline to other systems like MEG CTF, Optically Pumped Magnetometer (OPM) and EEG will be another crucial step in making FLUX a more generalized neurophysiological data analysis pipeline.

During the 2022 Brainhack, the team focused on incorporating the BIDS standard into the analysis pipeline using MNE$_{B}IDS^{\mathbf{Appelhoff2019}}.Consequently, anupdatedversionofFLUXwasrelea$

4.7 Evaluating discrepancies in hippocampal segmentation protocols using automatic prediction of MRI quality (MRIQC)

Jacob Sanz-Robinson, Mohammad Torabi, Tyler James Wishard

4.7.1 Introduction

Neuroimaging study results can vary significantly depending on the processing pipelines utilized by researchers to run their analyses, contributing to reproducibility issues. Researchers in the field are often faced with multiple choices of pipelines featuring similar capabilities, which may yield different results when applied to the same data[26,27]. While these reproducibility issues are increasingly well-documented in the literature, there is little existing research explaining why this inter-pipeline variability occurs or the factors contributing to it. In this project, we set out to understand what data-related factors impact the discrepancy between popular neuroimaging processing pipelines.

4.7.2 Method

The hippocampus is a structure commonly associated with memory function and dementia, and the left hippocampus is proposed to have higher discriminative power for identifying the progression of Alzheimer's disease than the right hippocampus in multiple studies[28]. We obtained left hippocampal volumes using three widely-used neuroimaging pipelines: FSL 5.0.9[29], FreeSurfer 6.0.0[30], and ASHS 2.0.0 PMC-T1 atlas[31]. We ran the three pipelines on T1 images from 15 subjects from the Prevent-AD Alzheimer's dataset[32], composed of cognitively healthy participants between the ages of 55-88 years old that are at risk of developing Alzheimer's Disease. We ran MRIQC[33] - a tool for performing automatic quality control and extracting quality measures from MRI scans - on the 15 T1 scans and obtained Image Quality Metrics (IQMs) from them. We then found the correlations between the IQMs and the pairwise inter-pipeline discrepancy of the left hippocampal volumes for each T1 scan.

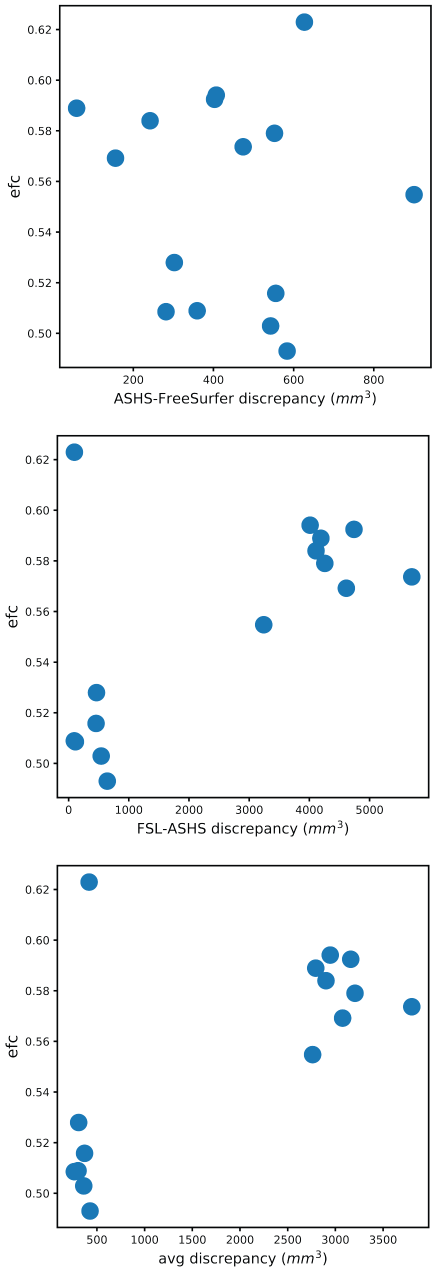


Figure 4: Plots showing the association between left hippocampal volume discrepancies and MRIQC's EFC quality measure for each of the pipeline pairings.

4.7.3 Results

We found that for The FSL-FreeSurfer and FSL-ASHs discrepancies, MRIQC's EFC measure produced the highest correlation, of 0.69 and 0.64, respectively. The EFC "uses the Shannon entropy of voxel intensities as an indication of ghosting and blurring induced by head motion"[34]. No such correlations were found for the ASHS-FreeSurfer discrepancies. Figure 4 shows a scatter plot of the discrepancies in left hippocampal volume and EFC IQM for each pipeline pairing. The preliminary results suggest that FSL's hippocampal segmentation may be sensitive to head motion in T1 scans, leading to larger result discrepancies, but we require larger sample sizes to make meaningful conclusions. The code for our project can be found on GitHub at this link.

4.7.4 Conclusion and Next Steps

In this project, we investigated the correlation between MRIQC's IQMs and discrepancies in left hippocampal volume derived from three common neuroimaging pipelines on 15 subjects from the Prevent-AD study dataset. While our preliminary results indicate image ghosting and blurring induced by head motion may play a role in inter-pipeline result discrepancies, the next steps of the project will consist of computing the correlations on the full 308 subjects of the Prevent-AD dataset to investigate whether they persist with the full sample.

4.8 Accelerating adoption of metadata standards for dataset descriptors

Cassandra Gould van Praag, Felix Hoffstaedter, Sebastian Urchs

Thanks to efforts of the neuroimaging community, not least the brainhack community[1], datasets are increasingly shared on open data repositories like OpenNeuro[35] using standards like BIDS[36] for interoperability. As the amount of datasets and data repositories increases, we need to find better ways to search across them for samples that fit our research questions. In the same way that the wide adoption of BIDS makes data sharing and tool development easier, the wide adoption of consistent vocabulary for demographic, clinical and other sample metadata would make data search and integration easier. We imagine a future platform that allows cross dataset search and the pooling of data across studies. Efforts to establish such metadata standards have had some success in other communities[37,38], but adoption in the neuroscience community so far has been slow. We have used the space of the brainhack to discuss which challenges are hindering wide adoption of metadata standards in the neuroimaging community and what could be done to accelerate it.

We believe that an important social challenge for the wider adoption of metadata standards is that it is hard to demonstrate their value without a practical use case. We therefore think that rather than focusing on building better standards, in the short term we need to prioritize small, but functional demonstrations that help convey the value of these standards and focus on usability and ease of adoption. Having consistent names and format for even a few metadata variables like age, sex, and diagnosis already allows for interoperability and search across datasets. Selecting a single vocabulary that must be used for annotating e.g. diagnosis necessarily lacks some precision but avoids the need to align slightly different versions of the same terms. Accessible tools can be built to facilitate the annotation process of such a basic metadata standard. The best standard will be poorly adopted if there are no easy to use tools that implement it. Efforts like the neurobagel project (`neurobagel.org/`) are trying to implement this approach to demonstrate a simple working use case for cross dataset integration and search. Our goal is to use such simpler demonstrations to build awareness and create a community around the goal of consistent metadata adoption.

Our long term goal is to use the awareness of the value of shared metadata standards to build a community to curate the vocabularies used for annotation. The initially small number of metadata variables will have to be iteratively extended through a community driven process to determine what fields should be standardized to serve concrete use cases. Rather than creating new vocabularies the goal should be to curate a list of existing ones that can be contributed to where terms are inaccurate or missing. The overall goal of such a community should be to build consensus on and maintain shared standards for the annotation of neuroimaging metadata that support search and integration of data for an ever more reproducible and generalizable neuroscience.

4.9 The NARPS Open Pipelines Project

Elodie Germani, Arshitha Basavaraj, Trang Cao, Rémi Gau, Anna Menacher, Camille Maumet

The goal of the NARPS Open Pipelines Project is to provide a public codebase that reproduces the 70 pipelines chosen by the 70 teams of the NARPS study[39]. The project is public and the code hosted on GitHub at `https://github.com/Inria-Empenn/narps_open_pipelines`.

This project initially emerged from the idea of creating an open repository of fMRI data analysis pipelines (as used by researchers in the field) with the broader goal to study and better understand the impact of analytical variability. NARPS – a many-analyst study in which 70 research teams were asked to analyze the same fMRI dataset with their favorite pipeline – was identified as an ideal usecase as it provides a large array of pipelines created by different labs. In addition, all teams in NARPS provided extensive (textual) description of their pipelines using the COBIDAS[40] guidelines. All resulting statistic maps were shared on NeuroVault[41] and can be used to assess the success of the reproductions.

At the OHBM Brainhack 2022, our goal was to improve the accessibility and reusability of the database, to facilitate new contributions and to reproduce more pipelines. We focused our efforts on the first two goals. By trying to install the computing environment of the database, contributors provided feedback on the instructions and on specific issues they faced during the installation. Two major improvements were made for the download of the necessary data: the original fMRI dataset and the original results (statistic maps stored in NeuroVault) were added as submodules to the GitHub repository. Finally, propositions were made to facilitate contributions: the possibility to use of the Giraffe toolbox[42] for contributors that are not familiar with NiPype[43] and the creation of a standard template to reproduce a new pipeline.

With these improvements, we hope that it will be easier for new people to contribute to reproduction of new pipelines. We hope to continue growing the codebase in the future.

4.10 NeuroCausal: Development of an Open Source Platform for the Storage, Sharing, Synthesis, and Meta-Analysis of Neuropsychological Data

Isil Poyraz Bilgin, Francois Paugam, Ruoqi Huang, Ana Luísa Pinho, Yuchen Zhou, Sladjana Lukic, Pedro Pinheiro-Chagas, Valentina Borghesani

Cognitive neuroscience has witnessed great progress since modern neuroimaging embraced an open science framework, with the adoption of shared principles[14], standards[36], and ontologies[44], as well as practices of meta-analysis[45,46] and data sharing[41]. However, while functional neuroimaging data provide correlational maps between cognitive functions and activated brain regions, its usefulness in determining causal link between specific brain regions and given behaviors or functions is disputed[47,48]. On the contrary, neuropsychological data enable causal inference, highlighting critical neural substrates and opening a unique window into the inner workings of the brain[49]. Unfortunately, the adoption of Open Science practices in clinical settings is hampered by several ethical, technical, economic, and political barriers, and as a result, open platforms enabling access to and sharing clinical (meta)data are scarce[50].

With our project, NeuroCausal (`https://neurocausal.github.io/`), we aim to build an online platform and community that allows open sharing, storage, and synthesis of clinical (meta) data crucial for the development of modern, transdiagnostic, accessible, and replicable (i.e., FAIR: Findability, Accessibility, Interoperability, and Reusability) neuropsychology. The project is organized into two infrastructural stages: first, published peer-reviewed papers will be scrapped to collect already available (meta)data; second, our platform will allow direct uploading of clinical (de-identified) brain maps and their corresponding metadata.

The meta-analysis pipeline developed for the first stage of the project is inspired by and built upon the functionalities of NeuroQuery[45], a successful large-scale neuroimaging meta-analytic platform. The first stage is the development of the code base allowing (1) downloading and filtering of neuropsychological papers, (2) extraction of reported brain lesion locations and their conversion into a common reference space (3) extraction of clinical and behavioral symptoms and their translation into a common annotation scheme, (4) learning the causal mapping between the neural and neuropsychological information gathered.

The second stage of the study aims at creating an online platform that allows for the direct uploading of clinical brain maps and their corresponding metadata. The platform will provide a basic automated preprocessing and a data-quality check pipeline, ensuring that all the ethical norms regarding patient privacy are met. The platform will automatically extract and synthesize key data to ultimately create probabilistic maps synthesizing transdiagnostic information on symptom-structure mapping, which will be dynamically updated as more data are gathered.

The nature of the project requires expertise in different fields (from clinical neuroscience to computer science) in order to overcome both technical and theoretical challenges. The OHBM Brainhack 2022 gave us the opportunity to set the first stones. In small subteams, we worked on developing three key building blocks: (1) the input filtering pipeline to ensure the downloaded papers are neuropsychological in nature and offer causal symptom-structure mapping; (2) the extraction of key terms occurrences in the text as to assess which neural space is reported (as they will have to be converted to a common one), (3) the curation of clinical ontology mapping specific neuropsychological batteries and tasks to the cognitive term(s) they touch upon. During the hackahton we worked on developing three key building blocks in small subteams. First, we prepared an input filtering pipeline to ensure that the downloaded papers are neuropsychological in nature (and thus offer causal symptom-structure mapping): we count the occurrences of clinically relevant terms, and papers are included only if they pass an arbitrary threshold. Second, we coded a script automatically returning for each paper information on the neural spaced used (e.g., which atlas? MNI coordinates?), a crucial step to enable future conversion to a common reference space. Finally, we curated a list of clinically relevant terms and constructs (a clinical ontology) that maps specific neuropsychological batteries

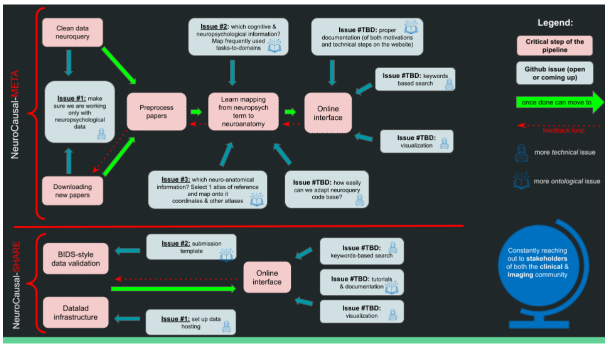


Figure 5: NeuroCausal: The future of neuropsychology, i.e. brain lesions-symptom mapping, will be transdiagnostic, open, and FAIR: we set out to provide the field with an open-source platform fostering storage, sharing, synthesis, and meta-analysis of clinical data.

and tasks to the cognitive term(s) they touch upon.

As we keep tackling our roadmap (Figure 1), we believe our efforts will help promote open science practices in clinical neuroscience to the benefit of both the neuroscientific and the clinical communities.

Acknowledgments **:** The authors would like to thank Eric Earl, Samuel Guay, Jerome Dockès, Bertrand Thirion, Jean Baptiste Poline, Yaroslav Halchenko, Sara El-Gebali and the whole Open Life Science team for their help and support.

4.11 Neuroscout: A platform for fast and flexible re-analysis of (naturalistic) fMRI studies

Alejandro De La Vega, Roberta Rocca, Sam Nastase, Peer Horholz, Jeff Mentch, Kevin Sitek, Caroline Martin, Leonardo Muller-Rodriguez, Kan Keeratimahat, Dylan Nielson

Neuroscout is an end-to-end platform for analysis of naturalistic fMRI data designed to facilitate the adoption of robust and generalizable research practices. Neuroscout's goal is to make it easy to analyze complex naturalistic fMRI datasets by providing an integrated platform for model specification and automated statistical modeling, reducing technical barriers. Importantly, Neuroscout is at its core a platform for reproducible analysis of fMRI data in general, and builds upon a set of open standards and specifications to ensure analyses are Findable, Accessible, Interoperable, and Reusable (FAIR).

In the OHBM Hackathon, we iterated on several important projects that substantially improved the general usability of the Neuroscout platform. First, we launched a revamped and unified documentation which links together all of the subcomponents of the Neuroscout platform (`https://neuroscout.github.io/neuroscout-docs/`). Second, we facilitated access to Neuroscout's data sources by simplifying the design of Python API, and providing high-level utility functions for easy programmatic data queries. Third, we updated a list of candidate naturalistic and non-naturalistic datasets amenable for indexing by the Neuroscout platform, ensuring the platform stays up to date with the latest public datasets.

In addition, important work was done to expand the types of analyses that can be performed with naturalistic data in the Neuroscout platform. Notably, progress was made in integrating Neuroscout with Himalaya, a library for efficient voxel wide encoding modeling with support for banded penalized regression. In addition, a custom image-on-scalar analysis was prototyped on naturalistic stimuli via the publicly available naturalistic features available in the Neuroscout API. Finally, we also worked to improve documentation and validation for BIDS StatsModels, a specification for neuroimaging statistical models which underlies Neuroscout's automated model fitting pipeline.

4.12 Physiopy - Documentation of Physiological Signal Best Practices

Sarah E. Goodale, Ines Esteves, Roza G. Bayrak, Neville Magielse, Stefano Moia, Yu-Fang Yang, The Physiopy Community

Physiological data provides a representation of a subject's internal state with respect to peripheral measures (i.e., heart rate, respiratory rate, etc.). Recording physiological measures is key to gain understanding of sources of signal variance in neuroimaging data that arise from outside of the brain[51]. This has been particularly useful for functional magnetic resonance imaging (fMRI) research, improving fMRI time-series model accuracy, while also improving real-time methods to monitor subjects during scanning[52,53].

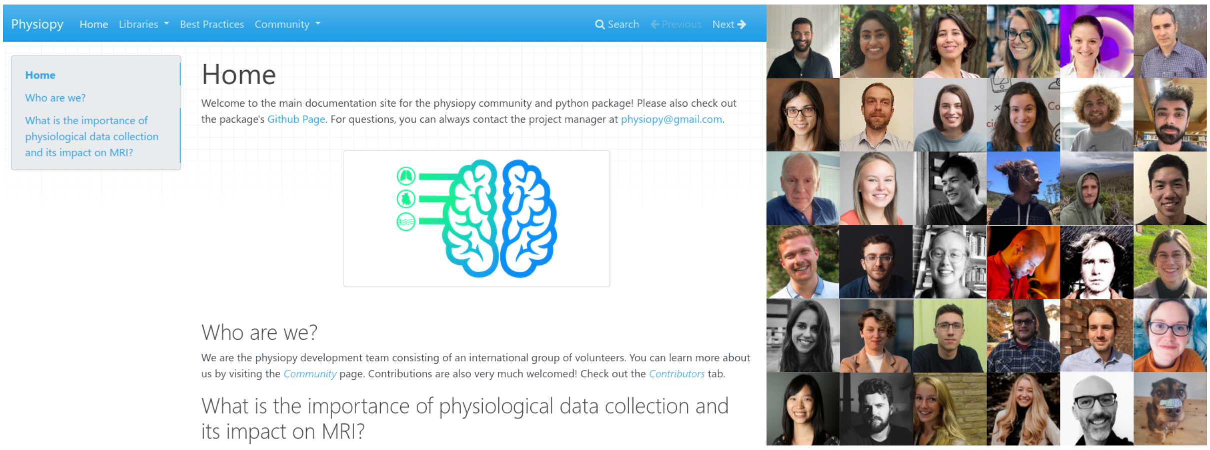


Figure 6: Left: Current version of the documentation homepage; Right: Physiopy Contributors

Physiopy (`https://github.com/physiopy`) is an open and collaborative community, formed around the promotion of physiological data collection and incorporation in neuroimaging studies. Physiopy is focused on two main objectives. The first is the community-based development of tools for fMRI-based physiological processing. At the moment, there are three toolboxes: *phys2bids* (physiological data storage and conversion to BIDS format[54], *peakdet* (physiological data processing), and *phys2denoise* (fMRI denoising). The second objective is advancing the general knowledge of physiological data collection in fMRI by hosting open sessions to discuss best practices of physiological data acquisition, preprocessing, and analysis, and promoting community involvement. Physiopy maintains documentation with best practices guidelines stemming from these joint discussions and recent literature.

At the OHBM 2022 Brainhack, we aimed to improve our community documentation by expanding on best practices documentation, and gathering libraries of complementary open source software. This provides new users resources for learning about the process of physiological collection as well as links to already available resources.The short-term goal for the Brainhack was to prepare a common platform (and home) for our documentation and repositories. We prioritised fundamental upkeep and content expansion, adopting Markdown documents and GitHub hosting to minimise barriers for new contributors. Over the course of the Brainhack, and with the joint effort within three hubs (Glasgow, EMEA and Americas), we were able to improve the current community website by rethinking its structure and adding fundamental content relative to who we are, contributions, and updated best practices, such as creating home pages, easy to find and navigate contribution tabs, adding new information from community best practices discussions as well as links to relevant software and datasets. Additionally, we aggregated the information scattered across different repositories, allowing important information for both the community and new collaborators to be accessible in a single location.

The long-term goals of the community are to develop and sustain knowledge and instruments for physiological signal adoption in f/MRI settings. Our aim is to facilitate the coming-together of researchers that are just starting to include physiological measures and experienced users. This community will then provide consensus guidelines for standardised data collection and preprocessing. Building on what we have already achieved, we will continue to promote and document best practices. Further development is ongoing and anyone that is interested in physiological signal collection for f/MRI data, independently of their level and type of expertise, is highly encouraged to check Physiopy out, to join the community, or to contribute in any way.

4.13 Handling multiple testing problem through effect calibration: implementation using PyMC

Lea Waller, Kelly Garner, Christopher R. Nolan, Daniel Borek, Gang Chen

4.13.1 Introduction

Human brain imaging data is massively multidimensional, yet current approaches to modelling functional brain responses entail the application of univariate inferences to each voxel separately. This leads to the multiple testing problem and unrealistic assumptions about the data such as artificial dichotomization (statistically significant or not) in result reporting. The traditional approach of massively univariate analysis assumes that no information is shared across the brain, effectively making a strong prior assumption of a uniform distribution of effect sizes, which is unrealistic given the connectivity of the human brain. The consequent requirement for multiple testing adjustments results in the *calibration of statistical evidence* without considering the estimation of effect, leading to substantial information loss and an unnecessarily heavy penalty.

A more efficient approach to handling multiplicity focuses on the *calibration of effect estimation* under a Bayesian multilevel modeling framework with a prior assumption of, for example, normality across space[55]. The methodology has previously been implemented at the region level into the `AFNI` program `RBA`[56] using `Stan` through the `R` package `brms`[57]. We intend to achieve two goals in this project:

(i) To re-implement the methodology using PyMC improve the performance and flexibility of the modeling approach.

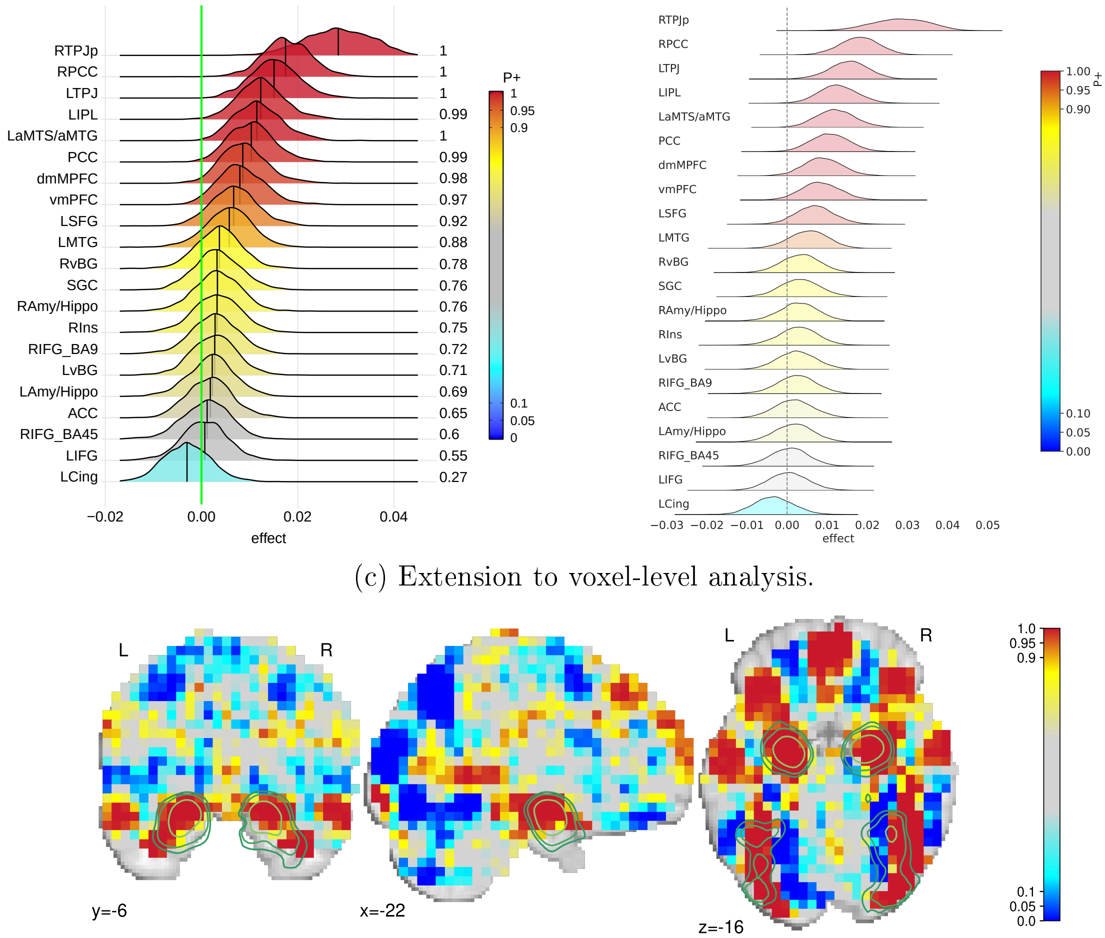
(ii) To explore the possibility of analyzing voxel-level data using the multilevel modeling approach

4.13.2 Implementation using PyMC

We used the dataset from Chen and colleagues[55] to validate our PyMC implementation. The data contain the subject-level response variable y and a predictor of the behavioral measure x from $S = 124$ subjects at $R = 21$ regions. The modeling framework is formulated for the data y_{rs} of the sth subject at the rth region as below,

$$
\begin{aligned}
y_{rs} &\sim \mathcal{N}(\mu_{rs},\ \sigma^2) \\
\mu_{rs} &= \alpha_0 + \alpha_1 x_s + \theta_{0r} + \theta_{1r} x_s + \eta_s \\
\begin{bmatrix} \theta_{0r} \\ \theta_{1r} \end{bmatrix} &\sim \mathcal{N}(\mathbf{0}_{2\times1},\ \boldsymbol{S}_{2\times2}) \\
\eta_s &\sim \mathcal{N}(0,\ \tau^2) \\
&\text{where } r = 1, 2, \ldots, R \text{ and } s = 1, 2, \ldots, S
\end{aligned}
\tag{1}
$$

In the model, μ_{rs} and σ are the mean effect and standard deviation of the sth subject at the rth region, α_0 and α_1 are the overall mean and slope effect across all regions and



(a) Posterior distributions: `brms`

(b) Posterior distributions: PyMC

(c) Extension to voxel-level analysis.

Figure 7: Validation of implementation using PyMC. (A) Posterior distributions of region-level behavior effects using `brms`. (B) Posterior distributions of region-level behavior effects using PyMC. (C) Posterior probabilities of the voxel-level effects being positive or negative, obtained using PyMC (plotted using Nilearn and overlaid in green with the NeuroQuery[45] map for the term “emotional faces”).

subjects, θ_{0r} and θ_{1r} are the mean and slope effect at the rth region, η_s is the mean effect of the sth subject, $\boldsymbol{S}_{2\times 2}$ is the variance-covariance of the mean and slope effect at the rth region, and τ is the standard deviation of the sth subject's effect η_s.

We implemented this model using the PyMC probabilistic programming framework[58], and the BAyesian Model-Building Interface (BAMBI)[59]. The latter is a high-level interface that allows for specification of multilevel models using the formula notation that is also adopted by `brms`. A notebook describing the implementation is available here. Our PyMC implementation was successfully validated: as shown in Figure 7a and Figure 7b, the posterior distributions from the PyMC implementation matched very well with their counterparts from the `brms` output.

4.13.3 Extension of Bayesian multilevel modeling to voxel-level analysis

After exploring the model on the region level, we wanted to see if recent computational and algorithmic advances allow us to employ the multilevel modeling framework on the voxel level as well. We obtained the OpenNeuro dataset `ds000117`[60] from an experiment based on a face processing paradigm. Using `HALFpipe`[61], which is based on `fMRIPrep`[62], the functional images were preprocessed with default settings and z-statistic images were calculated for the contrast "famous faces + unfamiliar faces versus 2 · scrambled faces".

We applied the same modeling framework and PyMC code as for region-based analysis, but without the explanatory varaiable x in the model (Equation (1)). To reduce computational and memory complexity, the z-statistic images were downsampled to an isotropic resolution of 5mm. Using the GPU-based `nuts_numpyro` sampler[63] with default settings, we were able to draw 2,000 posterior samples of the mean effect parameter for each of the 14,752 voxels. Sampling four chains took 23 minutes on four Nvidia Tesla V100 GPUs.

The resulting posterior probabilities are shown in Figure 7c overlaid with the meta-analytic map for the term "emotional faces" obtained from NeuroQuery[45]. The posterior probability map is consistent with meta-analytic results, showing strong statistical evidence in visual cortex and amygdala voxels. The posterior probability maps also reveal numerous other clusters of strong statistical evidence for both positive and negative effects.

This implementation extension shows that large multilevel models are approaching feasibility, suggesting an exciting new avenue for statistical analysis of neuroimaging data. Next steps will be to investigate how to interpret and report these posterior maps, and to try more complex models that include additional model terms.

Acknowledgements

Computation has been performed on the HPC for Research cluster of the Berlin Institute of Health.

4.14 MOSAIC for VASO fMRI

Renzo (Laurentius) Huber, Remi Gau, Rüdiger Stirnberg, Philipp Ehses, Ömer Faruk Gülban, Benedikt A. Poser

Vascular Space Occupancy (VASO) is a functional magnetic resonance imaging (fMRI) method that is used for high-resolution cortical layer-specific imaging[64]. Currently, the most popular sequence for VASO at modern SIEMENS scanners is the one by Stirnberg and Stöcker[65] from the DZNE in Bonn, which is employed at more than 30 research labs worldwide. This sequence concomitantly acquires fMRI BOLD and blood volume signals. In the SIEMENS' reconstruction pipeline, these two complementary fMRI contrasts are mixed together within the same time series, making the outputs counter-intuitive for users. Specifically:

- The 'raw' NIfTI converted time-series are not BIDS compatible (see `https://github.com/bids-standard/bids-specification/issues/1001`).
- The order of odd and even BOLD and VASO image TRs is unprincipled, making the ordering dependent on the specific implementation of NIfTI converters.

Workarounds with 3D distortion correction, results in interpolation artifacts. Alternative workarounds without MOSAIC decorators result in unnecessarily large data sizes.

In the previous Brainhack[1], we extended the existing 3D-MOSAIC functor that was previously developed by Benedikt Poser and Philipp Ehses. This functor had been previously used to sort volumes of images by dimensions of echo-times, by RF-channels, and by magnitude and phase signals. In this Brainhack, we successfully extended and validated this functor to also support the dimensionality of SETs (that is representing BOLD and VASO contrast).

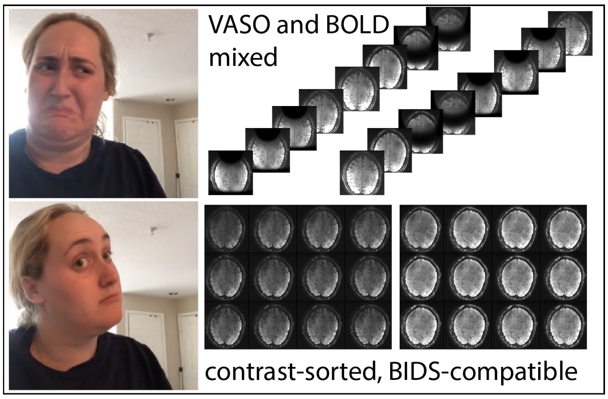


Figure 8: Previously, most VASO sequences provided unsorted image series of MRI contrasts. This was not BIDS compatible and could suffer from gradient non-linearity artifacts in the scanner's MR-reconstruction pipeline. In Brainhack 2022, we adapted the SIEMENS reconstruction and to sort volume series by fMRI contrasts. This is BIDS compatible and does not require non-linearity corrections.

We are happy to share the compiled SIEMENS ICE (Image Calculation Environment) functor that does this sorting. Current VASO users, who want to upgrade their reconstruction pipeline to get the MOSAIC sorting feature too, can reach out to Renzo Huber (RenzoHuber@gmail.com) or Rüdiger Stirnberg (Ruediger.Stirnberg@dzne.de).

Furthermore, Remi Gau, generated a template dataset that exemplifies how one could to store layer-fMRI VASO data. This includes all the meta data for 'raw and 'derivatives'. Link to this VASO fMRI BIDS demo: `https://gin.g-node.org/RemiGau/ds003216/src/bids_demo`.

Acknowledgements: We thank Chris Rodgers for instructions on how to overwrite existing reconstruction binaries on the SIEMENS scanner without rebooting. We thank David Feinberg, Alex Beckett and Samantha Ma for helping in testing the new reconstruction binaries at the Feinbergatron scanner in Berkeley via remote scanning. We thank Maastricht University Faculty of Psychology and Neuroscience for supporting this project with 2.5 hours of 'development scan time'.

5 Conclusion and future directions

Stefano Moia, Hao-Ting Wang

Approaching the organisation of an event as an experiment allows incredible freedom and dynamicity, albeit knowing that there will be risks and venues of improvement for the future.

The organisation managed to provide a positive onsite environment, aiming to allow participants to self-organise in the spirit of the Brainhack[1], with plenty of moral - and physical - support.

The technical setup, based on heavy automatisation flow to allow project submission to be streamlined, was a fundamental help to the organisation team, that would benefit even more from the improvement of such automatisation flows.

This year, representatives of AFNI, FSL, and SPM (among the major neuroscience software developers) took part in the event, and their presence was appreciated both by other participants and themselves. In the future, connecting to more developers, not only from the MRI community, might improve the quality of the Brainhack even more.

The most challenging element of the organisation was setting up an hybrid event. While this element did not go as smoothly as it could have, this experimental setup seemed to have worked, allowing the participation of about 70 participants online. However, there is still a lot to improve for a truly hybrid event. For instance, it is important to allow spaces (both in time and space) for participants on-site to interact with online participants, and more attention, time, volunteers, and equipment has to be put to achieve a smooth online participation. For this reason, the Open Science Special Interest Group instituted a position to have a dedicated person for the hybridisation process. The other challenge was to welcome newcomers into this heavily project-development-oriented event. While newcomers managed to collaborate with projects and self-organise to learn open science related skills, this integration of pre-event train track and beginner friendly process will benefit from more attention.

Overall this OHBM Brainhack was a successful outcome for the organisation team experiment, and we hope that our findings will be helpful to future Brainhack events organisations.

References

1. Gau R, Noble S, Heuer K et al. Brainhack: Developing a culture of open, inclusive, community-driven neuroscience. *Neuron*. 2021; 109(11):1769–1775.
2. Alkemade A, Bazin PL, Balesar R et al. A unified 3D map of microscopic architecture and MRI of the human brain. *Science Advances*. 2022; 8(17):1–10.
3. Heuer K, Gulban OF, Bazin PL et al. Evolution of neocortical folding: A phylogenetic comparative analysis of MRI from 34 primate species. en. *Cortex*. 2019; 118():275–291.
4. Huntenburg JM, Steele CJ, Bazin PL. Nighres: processing tools for high-resolution neuroimaging. en. *GigaScience*. 2018; 7(7).
5. Yushkevich PA, Piven J, Hazlett HC et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*. 2006; 31(3):1116–1128.
6. Cox RW. *AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages*. Tech. rep. 1996:162–173.
7. Smith SM, Jenkinson M, Woolrich MW et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. Vol. 23. SUPPL. 1. 2004.
8. Gulban OF, Schneider M, Marquardt I, Haast RA, De Martino F. *A scalable method to improve gray matter segmentation at ultra high field MRI*. Vol. 13. 6. 2018:1–31.
9. Huber L(R, Poser BA, Bandettini PA et al. LayNii: A software suite for layer-fMRI. *NeuroImage*. 2021; 237(May):118091.
10. Apon A, Ngo L, Payne M, Wilson P. Assessing the effect of high performance computing capabilities on academic research output. *Empirical Economics*. 2014; 48().
11. *Oracle cloud sustainability*. `https://www.oracle.com/uk/sustainability/green-cloud/`. Accessed: 2022-07-19.

12. *Oracle for Research*. `https://docs.oracle.com/en/programs/research/`. Accessed: 2022-07-19.

13. *NeuroDesk*. `https://www.neurodesk.org/`. Accessed: 2022-07-19.

14. Wilkinson MD, Dumontier M, Aalbersberg IJ et al. The FAIR Guiding Principles for scientific data management and stewardship. en. *Sci. Data*. 2016; 3(1):160018.

15. Halchenko YO, Meyer K, Poldrack B et al. DataLad: distributed system for joint management of code, data, and their relationship. *Journal of Open Source Software*. 2021; 6(63):3262.

16. King G. *An introduction to the dataverse network as an infrastructure for data sharing*. 2007.

17. Hanke M, Pestilli F, Wagner AS, Markiewicz CJ, Poline JB, Halchenko YO. In defense of decentralized research data management. en. *Neuroforum*. 2021; 27(1). Publisher: De Gruyter Section: Neuroforum:17–25.

18. Hanke M, Poldrack B, Wagner AS et al. *datalad/datalad-osf: Cleanup*. Version 0.2.3. Mar. 2021.

19. Halchenko YO, Hanke M, Heunis S et al. *DataLad-next extension*.

20. Jiang C, Selle A, Teran J. The Affine Particle-In-Cell Method Chenfanfu. *ACM Transactions on Graphics*. 1965; 34(4):51.

21. Love E, Sulsky DL. An unconditionally stable, energy-momentum consistent implementation of the material-point method. *Computer Methods in Applied Mechanics and Engineering*. 2006; 195(33-36):3903–3925.

22. Stomakhin A, Schroeder C, Chai L, Teran J, Selle A. A material point method for snow simulation. *ACM Transactions on Graphics*. 2013; 32(4).

23. Ferrante O, Liu L, Minarik T et al. FLUX: A pipeline for MEG analysis. *NeuroImage*. 2022; 253():119047.

24. Gramfort A, Luessi M, Larson E et al. MNE software for processing MEG and EEG data. *NeuroImage*. 2014; 86():446–460.

25. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011; 2011().

26. Carp J. On the plurality of (methodological) worlds: estimating the analytic flexibility of FMRI experiments. *Frontiers in neuroscience.* 2012; 6():149.

27. Kennedy DN, Abraham SA, Bates JF et al. Everything matters: the ReproNim perspective on reproducible neuroimaging. *Frontiers in neuroinformatics.* 2019():1.

28. Schuff N, Woerner N, Boreta L et al. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain.* 2009; 132(4):1067–1077.

29. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage.* 2011; 56(3):907–922.

30. Fischl B. FreeSurfer. *Neuroimage.* 2012; 62(2):774–781.

31. Xie L, Wisse LE, Pluta J et al. Automated segmentation of medial temporal lobe subregions on in vivo T1-weighted MRI in early stages of Alzheimer's disease. *Human brain mapping.* 2019; 40(12):3431–3451.

32. Tremblay-mercier J, Madjar C, Das S et al. Open science datasets from PREVENT-AD, a longitudinal cohort of pre-symptomatic Alzheimer's disease. *NeuroImage: Clinical.* 2021; 31():102733.

33. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PloS one.* 2017; 12(9):e0184661.

34. *MRIQC Documentation.* 2020. URL: `https://mriqc.readthedocs.io/en/stable/iqms/t1w.html`.

35. Markiewicz CJ, Gorgolewski KJ, Feingold F et al. The OpenNeuro resource for sharing of neuroscience data. en. *Elife.* 2021; 10().

36. Gorgolewski KJ, Auer T, Calhoun VD et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. en. *Sci Data.* 2016; 3():160044.

37. Field D, Garrity G, Gray T et al. The minimum information about a genome sequence (MIGS) specification. en. *Nat. Biotechnol.* 2008; 26(5):541–547.

38. Stang PE, Ryan PB, Racoosin JA et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. en. *Ann. Intern. Med.* 2010; 153(9):600–606.

39. Botvinik-nezer R, Holzmeister F, Camerer CF et al. Variability in the analysis of a single neuroimaging dataset by many teams. en. 2020():26.

40. Nichols TE, Das S, Eickhoff SB et al. Best practices in data analysis and sharing in neuroimaging using MRI. en. *Nature Neuroscience.* 2017; 20(3):299–303.

41. Gorgolewski KJ, Varoquaux G, Rivera G et al. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. en. *Frontiers in Neuroinformatics.* 2015; 9().

42. Mourik T van. *GiraffeToolBox.* Accessed: 2022-06-16.

43. Gorgolewski K. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. en. *Frontiers in Neuroinformatics.* 2017():15.

44. Poldrack R, Kittur A, Kalar D et al. The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience. *Frontiers in Neuroinformatics.* 2011; 5().

45. Dockès J, Poldrack RA, Primet R et al. NeuroQuery, comprehensive meta-analysis of human brain mapping. *eLife.* 2020; 9(). Ed. by Büchel C, Yeo T, Wager TD. Publisher: eLife Sciences Publications, Ltd:e53385.

46. Yarkoni T, Poldrack RA, Nichols TE, Van essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. en. *Nat Methods.* 2011; 8(8). Number: 8 Publisher: Nature Publishing Group:665–670.

47. Weber MJ, Thompson-schill SL. Functional neuroimaging can support causal claims about brain function. *J Cogn Neurosci.* 2010; 22(11):10.1162/jocn.2010.21461.

48. Siddiqi SH, Kording KP, Parvizi J, Fox MD. Causal mapping of human brain function. en. *Nat Rev Neurosci.* 2022; 23(6). Number: 6 Publisher: Nature Publishing Group:361–375.

49. Price CJ. The evolution of cognitive models: From neuropsychology to neuroimaging and back. en. *Cortex.* In Memory of Professor Glyn Humphreys 2018; 107():37–49.

50. Larivière S, Paquola C, Park By et al. The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets. en. *Nat Methods.* 2021; 18(7). Number: 7 Publisher: Nature Publishing Group:698–700.

51. Chen JE, Lewis LD, Chang C et al. Resting-State "Physiological Networks". *NeuroImage.* 2020; 213():116707.

52. Bulte D, Wartolowska K. Monitoring Cardiac and Respiratory Physiology during FMRI. *NeuroImage.* 2017; 154():81–91.

53. Caballero-gaudes C, Reynolds RC. Methods for Cleaning the BOLD fMRI Signal. *NeuroImage.* 2017; 154():128–149.

54. Alcalá D, Ayyagari A, Bottenhorn K et al. *physiopy/phys2bids: BIDS formatting of physiological recordings.* June 2021.

55. Chen G, Xiao Y, Taylor PA et al. Handling Multiplicity in Neuroimaging through Bayesian Lenses with Multilevel Modeling. *Neuroinformatics.* 2019; 17(4):515–545.

56. Chen G, Taylor PA, Stoddard J, Cox RW, Bandettini PA, Pessoa L. Sources of Information Waste in Neuroimaging: Mishandling Structures, Thinking Dichotomously, and Over-Reducing Data. *Aperture Neuro.* 2022; 2021(5):46.

57. Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. en. *Journal of Statistical Software.* 2017; 80(1). Number: 1:1–28.

58. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science.* 2016; 2():e55.

59. Capretto T, Piho C, Kumar R, Westfall J, Yarkoni T, Martin OA. *Bambi: A simple interface for fitting Bayesian linear models in Python.* 2020. arXiv: 2012.10754 [stat.CO].

60. Wakeman DG, Henson RN. A multi-subject, multi-modal human neuroimaging dataset. en. *Scientific Data.* 2015; 2(1):150001.

61. Waller L, Erk S, Pozzi E et al. ENIGMA HALFpipe: Interactive, reproducible, and efficient analysis for resting-state and task-based fMRI data. en. *Human Brain Mapping.* 2022; 43(9):2727–2742.

62. Esteban O, Markiewicz CJ, Blair RW et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. en. *Nature Methods.* 2019; 16(1). Number: 1 Publisher: Nature Publishing Group:111–116.

63. Phan D, Pradhan N, Jankowiak M. *Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro.* arXiv:1912.11554 [cs, stat]. Dec. 2019.

64. Huber L, Finn ES, Chai Y et al. Layer-dependent functional connectivity methods. *Progress in Neurobiology.* 2021; 207().

65. Stirnberg R, Stöcker T. Segmented K-Space Blipped-Controlled Aliasing in Parallel Imaging (Skipped-CAIPI) for High Spatiotemporal Resolution Echo Planar Imaging. *Magnetic Resonance in Medicine.* 2021; 85(0):1540–1551.