

SOVABIDS: EEG-to-BIDS conversion software focused on automation, reproducibility and interoperability

Yorguin-José Mantilla-Ramos^{1,5,6,7,8,9}, Brayan-Andrés Hoyos-Madera^{1,5}, Steffen Bollmann², Aswin Narayanan^{2,10}, David White⁴, Oren Cavier^{3,4}, and Tom Johnstone^{3,4}

1 Grupo Neuropsicología y Conducta (GRUNECO), Universidad de Antioquia, Medellín, Colombia **2** The University of Queensland, Brisbane, Queensland, Australia **3** Australian National Imaging Facility, Australia **4** Swinburne University of Technology, Melbourne, Victoria, Australia **5** Semillero de Investigación Neurociencias Computacionales (NeuroCo), Universidad de Antioquia, Medellín, Colombia **6** Cognitive and Computational Neuroscience Laboratory (CoCo Lab), Psychology Department, Université de Montréal, Montréal, Canada **7** Mila (Quebec AI Institute), Montréal, Canada **8** Grupo Sistemas Embebidos e Inteligencia Computacional (SISTEMIC), Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia **9** Semillero de Investigación Machine Learning and Robotics, Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia **10** Australian National Imaging Facility, The University of Queensland, Brisbane, Australia  Corresponding author

DOI: [10.21105/joss.10146](https://doi.org/10.21105/joss.10146)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Rachel Kurchin](#)  

Reviewers:

- [@qian-chu](#)
- [@wmvanvliet](#)

Submitted: 06 August 2025

Published: 21 October 2026

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Electroencephalography (EEG) recordings are stored in many vendor-specific formats and lab-specific folder structures, making them difficult to organize, share, or compare across studies. SOVABIDS is an open-source tool that converts EEG data into the Brain Imaging Data Structure (BIDS) ([Gorgolewski et al., 2016](#)), a standard that promotes FAIR data practices (Findability, Accessibility, Interoperability, and Reusability). Rather than manually renaming files or writing dataset-specific scripts, users define conversion rules in human-readable YAML configuration files, which SOVABIDS applies across the dataset ([Figure 1](#)). It can be used as a Python package, a command-line tool, or an experimental terminal user interface (TUI), and its API supports integration with graphical frontends. Documentation and tutorials are available at sovabids.readthedocs.io.

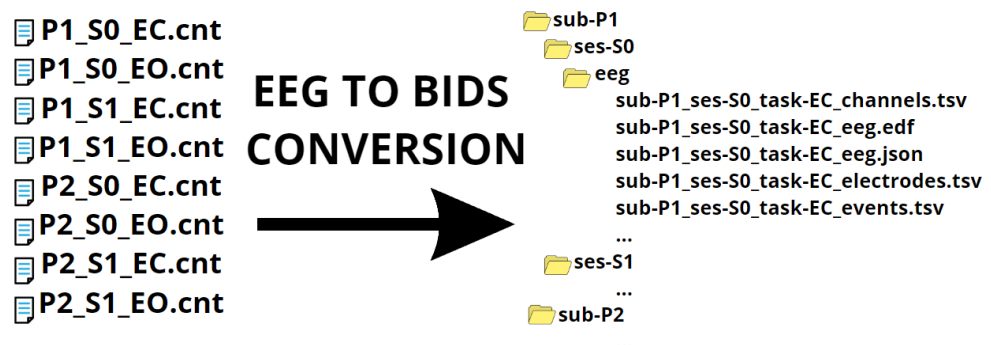


Figure 1: EEG-to-BIDS conversion. Left: raw EEG files following a lab-specific naming convention, e.g. P1_S0_EC.cnt for participant P1, session S0, task EC (Eyes Closed; EO, Eyes Open). Right: the converted BIDS dataset, organized into subject (sub-), session (ses-), and modality (eeg) folders, with signal files (here .edf) and .tsv/.json metadata.

Statement of need

EEG is a widely used neuroimaging technique, with applications in cognitive neuroscience, clinical diagnostics, brain-computer interfaces, and neuroengineering. The EEG extension of BIDS (EEG-BIDS) ([Pernet et al., 2019](#)) standardizes the organization of EEG datasets, facilitating data sharing ([Markiewicz et al., 2021](#)), cross-study comparison, and FAIR data practices ([Wilkinson et al., 2016](#)).

Despite these advantages, converting EEG datasets to BIDS remains challenging. EEG formats vary across hardware vendors, and datasets are often organized by lab-specific or acquisition-driven conventions rather than a consistent structure. For researchers without strong programming backgrounds, particularly at less well-resourced institutions, conversion typically requires substantial manual effort or dataset-specific scripts, both error-prone and difficult to reproduce or scale to large multi-participant studies. SOVABIDS addresses this gap with a rule-based, semi-automated workflow that is accessible to users with limited programming experience yet flexible enough for the heterogeneity of real-world EEG datasets.

State of the Field

Several tools support EEG conversion to BIDS, including MNE-BIDS ([Appelhoff et al., 2019](#)), data2bids in FieldTrip ([Oostenveld et al., 2011](#)), EEG-BIDS in EEGLAB ([Delorme & Makeig, 2004](#)), EEG2BIDS ([Rogers et al., 2022](#)), and general-purpose, multi-modality converters such as Bidsme ([Beliy et al., 2023](#)).

MNE-BIDS provides a programmatic interface within the MNE ecosystem with fine-grained control over metadata, but typically requires dataset-specific scripting, which limits accessibility for users without programming experience.

FieldTrip and EEGLAB integrate conversion into their analysis environments, convenient for existing users, but these workflows often still require manual interaction or scripting for each dataset.

EEG2BIDS offers a more guided workflow but relies on detailed file-level user input, which becomes impractical for large heterogeneous datasets.

Bidsme is a general-purpose, YAML-configured converter whose preparation step derives subject and session labels from folder names or file metadata; other naming conventions require user-written Python plugins.

A natural question is whether SOVABIDS' goals could have been achieved by contributing to an existing tool, particularly MNE-BIDS. We argue this was impractical. In MNE-BIDS, conversion logic is written in Python, and adding rule-based, configuration-driven automation as a non-breaking extension would require architectural changes that diverge from its scripting-oriented design. FieldTrip and EEGLAB face the same barrier, as conversion is tightly coupled to their analysis environments.

SOVABIDS instead introduces an explicit two-tier separation between dataset-level rules and file-level mappings, encoded in human-readable YAML configuration files. Like other tools, it requires dataset-specific configuration, but this is defined once and expanded into file-level mappings, removing the need for per-file scripting. Combined with a semi-automatic API for extracting BIDS entities from file paths, this lets SOVABIDS handle the heterogeneous directory structures and vendor-specific naming conventions of real-world EEG datasets.

Software Design

SOVABIDS' central design tension is between expressiveness and accessibility: more expressive conversion logic typically requires programming skill, while simpler interfaces tend to sacrifice

flexibility. The five design principles below describe these trade-offs, and [Figure 2](#) summarizes the resulting architecture.

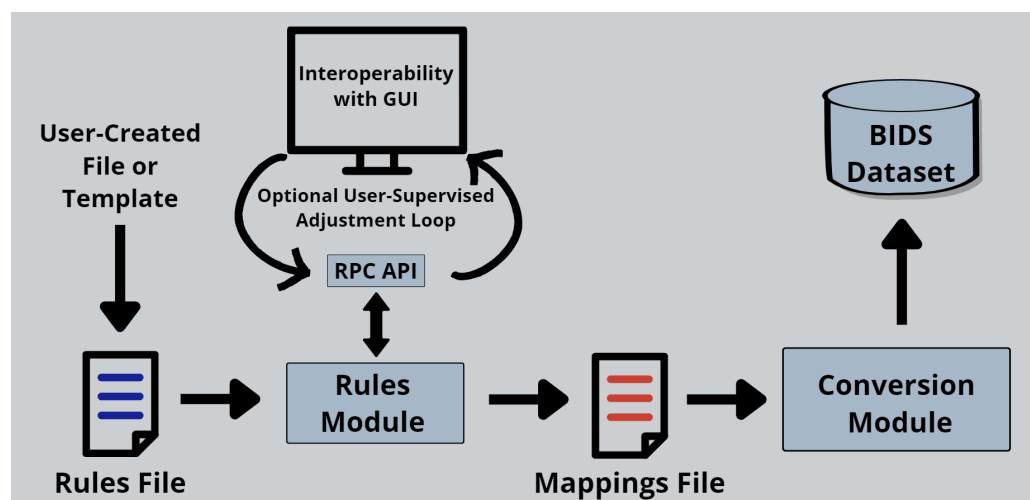


Figure 2: SOVABIDS architecture. The Rules Module applies a user-created Rules File across the dataset to produce a Mappings File, which the Conversion Module uses to convert EEG files to BIDS via MNE and MNE-BIDS. An RPC API lets external tools and GUIs inspect or adjust the configuration.

1. Accessibility for non-technical users

A scripting-based interface, as used by tools like MNE-BIDS, offers greater expressiveness but requires users to write and maintain dataset-specific code. SOVABIDS trades this expressiveness for accessibility by using YAML configuration files, an approach inspired by the MRI converter Bidscoin ([Zwiers et al., 2022](#)). YAML configurations are reusable across similar datasets and auditable without programming knowledge. Highly unusual conversion scenarios may require more verbose configuration, but for many EEG datasets the YAML-based approach is sufficient and more approachable than writing a custom script.

2. Automation that can accommodate outliers

EEG experiments typically record each participant in the same way, but in practice data organization often varies slightly between participants (technical issues, partial recordings, or repeated segments). A fully automated system that assumes identical structure can fail silently in these cases, while a fully manual one would not scale. SOVABIDS therefore separates conversion logic into two configuration files ([Figure 3](#)):

- The [Rules File](#), which encodes general conversion rules for the full dataset.
- The [Mappings File](#), populated from the Rules File, which holds specific conversion parameters for every individual file.

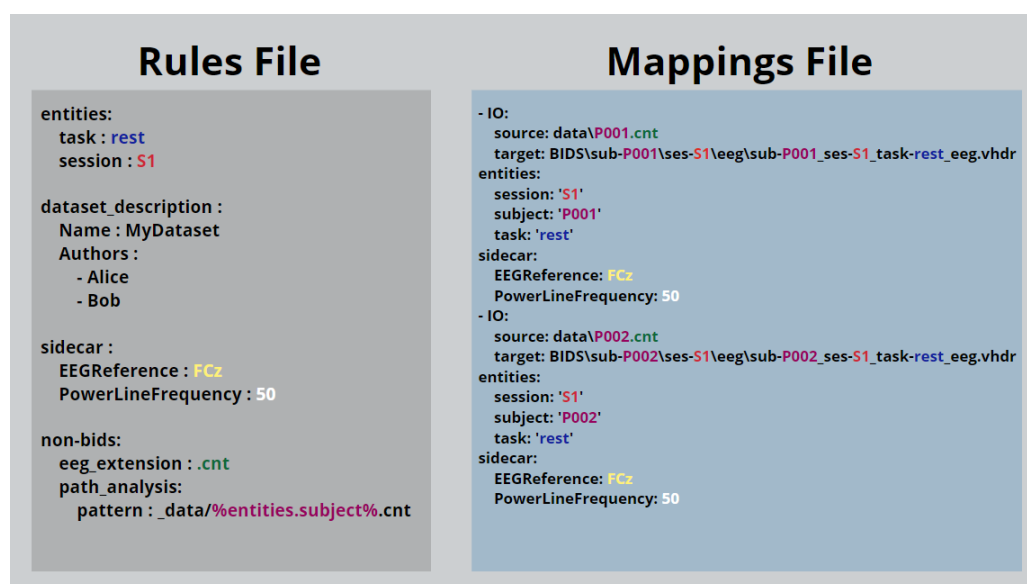


Figure 3: A Rules File (left) generates one mapping per file in the dataset, saved in the Mappings File (right); colors show how values in the two files correspond.

This two-tier approach is inspired by MRI tools such as Bidscoin (Zwiers et al., 2022) and HeuDiConv (Halchenko et al., 2023). It generates a separate mapping for each file, which users can review and edit, directly or through an external GUI connected via SOVABIDS' API, when a participant's data does not follow the general structure. Users can also start from an existing Rules File shared within a lab or community.

SOVABIDS infers subject, session, task, and other BIDS entities from file paths through a semi-automatic API. Rather than requiring data to be pre-organized into a standard folder hierarchy, as Bidscoin does, it uses patterns defined once at the dataset level and applied across selected files. Patterns take three increasingly technical forms: paired source-target examples, placeholder-based templates, or full regular expressions.

3. Reproducible conversion

Reproducibility requires that a conversion be specified by its saved configuration, not by undocumented manual steps. The rules and per-file mappings are saved and each run is logged, so users can audit, correct, and re-run conversions when a BIDS validator flags issues or downstream analysis reveals incorrect metadata.

4. Accessible interfaces and interoperability

Beyond the CLI and Python API, SOVABIDS provides two further access paths. The first is an experimental, optional terminal user interface (TUI). It guides users through four steps (Setup, Rules, Mappings, and Convert) without requiring any code. A TUI was chosen over web-based and native GUIs for its portability, small dependency footprint, and lower maintenance burden. A [walkthrough](#) is available.

The second is an RPC-based API that lets external applications drive SOVABIDS' conversion logic. RPC was chosen over REST because its action-oriented calls map naturally onto the procedural steps of a conversion. An [experimental reference web GUI in Flask \(demo\)](#) illustrates how platforms could embed SOVABIDS in a richer desktop or web frontend.

5. Format support through MNE delegation

Rather than supporting every EEG hardware format itself, SOVABIDS delegates file reading to MNE-Python ([Gramfort et al., 2013](#)) and BIDS-compliant saving to MNE-BIDS ([Appelhoff et al., 2019](#)), so its input coverage is bounded by MNE's broad and actively maintained format support. SOVABIDS is tested with BrainVision (.vhdr), EDF (.edf), EEGLAB (.set), and FIF (.fif) inputs, with BrainVision output by default (FIF for MEG). Other MNE-readable formats (Neuroscan .cnt and BDF) are handled through the same delegation but are untested in continuous integration. Basic MEG datatype routing is implemented, but MEG-specific BIDS requirements (empty-room recordings, manufacturer calibration files, and digitization coordinate systems) are not exposed through the rule system.

Research Impact Statement

SOVABIDS is listed in the [official BIDS converter registry](#) under EEG/MEEG/iEEG and is available on the [Neurodesk](#) platform ([Dao et al., 2025](#); [Renton et al., 2024](#)), a community-maintained open neuroimaging environment. Its use is documented in a Master's thesis on EEG-based Alzheimer's risk classification ([Henao Isaza, 2023](#)), a Bachelor's thesis on web-based EEG processing tools ([Zapata Saldarriaga, 2022](#)), and a peer-reviewed study on harmonizing EEG features across multiple recording sites ([Jaramillo-Jimenez et al., 2024](#)).

Acknowledgements

The authors acknowledge support from the 2021 Google Summer of Code program under the International Neuroinformatics Coordinating Facility (INCF), and funding from the Australian Research Data Commons (ARDC) for the Australian Electrophysiology Data Analytics Platform (AEDAPT). The authors also acknowledge the facilities and the scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability, at Swinburne Neuroimaging, Swinburne University of Technology, and at the Centre for Advanced Imaging, The University of Queensland.

AI Usage Disclosure

The core architecture and functionality were completed before December 2023. Generative AI was not used in the project's conceptual design, methodological decisions, or scientific development.

Since March 2025, generative AI tools, mainly OpenAI Codex and Claude Code (models from GPT-4.1 and Claude 4.6 onward), were used for maintenance and supporting work: the experimental TUI and its tests, continuous-integration workflows, documentation, small utilities, fixes during the JOSS review, and editing of this paper. The authors reviewed, tested, and validated all AI-assisted outputs and made all design decisions and scientific judgments.

References

- Appelhoff, S., Sanderson, M., Brooks, T. L., Vliet, M. van, Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., Tavabi, K., Höchenberger, R., Welke, D., Brunner, C., Rockhill, A. P., Larson, E., Gramfort, A., & Jas, M. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. *Journal of Open Source Software*, 4(44), 1896. <https://doi.org/10.21105/joss.01896>
- Beliy, N., Guillemain, C., Pommier, E., Hammad, G., & Phillips, C. (2023). Bidsme: Expandable BIDS-ifier of brain imagery datasets. *Journal of Open Source Software*, 8(92), 5575. <https://doi.org/10.21105/joss.05575>

- Dao, T. T., Masson-Trottier, M., Ribeiro, F. L., Scarsbrook, J., Stewart, A., Ye, X., Hamilton, E., Bollmann, S., & Narayanan, A. (2025). Democratizing open neuroimaging: Neurodesk's approach to open data accessibility and utilization. *Aperture Neuro*, 5(SI 2). <https://doi.org/10.52294/001c.144107>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Gorgolewski, K., Auer, T., Calhoun, V., Craddock, C., Das, S., Duff, E., Flandin, G., Ghosh, S., Glatard, T., Halchenko, Y., Handwerker, D., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B., Nichols, T., Pellman, J., & Poldrack, R. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3, 160044. <https://doi.org/10.1038/sdata.2016.44>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience*, 7. <https://doi.org/10.3389/fnins.2013.00267>
- Halchenko, Y., Goncalves, M., Velasco, P., Di Oleggio Castello, M. V., Satrajit Ghosh, Salo, T., Wodder, J. T., Hanke, M., Sadil, P., Christian, H., Michael, Dae, Tilley, S., Kent, J., To, I., Brett, M., Amlie, I., Gorgolewski, C., Markiewicz, C., ..., Michael. (2023). *Nipy/heudiconv: v1.0.0*. Zenodo. <https://doi.org/10.5281/ZENODO.1012598>
- Henao Isaza, V. (2023). *Machine learning model for the classification of individuals at risk of dementia type alzheimer from multimodal databases of EEG and clinical information* [Master's Thesis, Universidad de Antioquia]. <https://bibliotecadigital.udea.edu.co/handle/10495/37580>
- Jaramillo-Jimenez, A., Tovar-Rios, D. A., Mantilla-Ramos, Y.-J., Ochoa-Gomez, J.-F., Bonanni, L., & Brønck, K. (2024). ComBat models for harmonization of resting-state EEG features in multisite studies. *Clinical Neurophysiology*, 167, 241–253. <https://doi.org/10.1016/j.clinph.2024.09.019>
- Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., 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>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 1–9. <https://doi.org/10.1155/2011/156869>
- Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, 6(1). <https://doi.org/10.1038/s41597-019-0104-8>
- Renton, A. I., Dao, T. T., Johnstone, T., Civier, O., Sullivan, R. P., White, D. J., Lyons, P., Slade, B. M., Abbott, D. F., Amos, T. J., Bollmann, S., Botting, A., Campbell, M. E. J., Chang, J., Close, T. G., Dörig, M., Eckstein, K., Egan, G. F., Evas, S., ..., Bollmann, S. (2024). Neurodesk: An accessible, flexible and portable data analysis environment for reproducible neuroimaging. *Nature Methods*, 21(5), 804–808. <https://doi.org/10.1038/s41592-023-02145-x>
- Rogers, C., Wickenheiser, A., Fesselier, L., Callegaro, J., Madjar, C., Das, S., & Evans, A. C. (2022). EEG2BIDS wizard converter tool: Flexible open-source cross-platform tool for anonymization and BIDS standardization. In *GitHub repository*. GitHub. <https://github.com/aces/EEG2BIDS>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes,

- A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>
- Zapata Saldarriaga, L. M. (2022). *Desarrollo de aplicación de servicios web basado en estándares de informática médica para el preprocesamiento y visualización de registros EEG* [Bachelor's Thesis, Universidad de Antioquia]. <https://bibliotecadigital.udea.edu.co/handle/10495/30073>
- Zwiers, M. P., Moia, S., & Oostenveld, R. (2022). BIDScoin: A user-friendly application to convert source data to brain imaging data structure. *Frontiers in Neuroinformatics*, 15. <https://doi.org/10.3389/fninf.2021.770608>