The Journal of Open Source Software

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 Bollmann**[2], **Aswin Narayanan**[2], **David White**[4], **Oren Civier**[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 ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Rachel Kurchin

Submitted: 06 August 2025
Published: unpublished

Summary

Electroencephalography (EEG) data are used in many fields, from neuroscience to clinical research, but it often comes in different formats and structures, making it hard to organize, share, or compare across studies. SOVABIDS is an open-source tool that helps researchers convert EEG data into the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016), a standard data structure that aligns with FAIR principles (Findability, Accessibility, Interoperability, and reusability). Specifically, BIDS ensures data consistency and interoperability, making it easier to analyze, share, and integrate EEG data across different tools, data repositories and research groups. Rather than manually renaming files or reorganizing folders, SOVABIDS allows users to define simple rules to automate the conversion (Figure 1 illustrates this conversion process). The tool is designed to be flexible and user-friendly, supporting customization without requiring advanced programming skills. It can be used as a Python package or as a command-line tool, and includes comprehensive documentation with tutorials to help users get started. SOVABIDS also integrates with other tools, making it suitable for both small research projects and large collaborative studies.

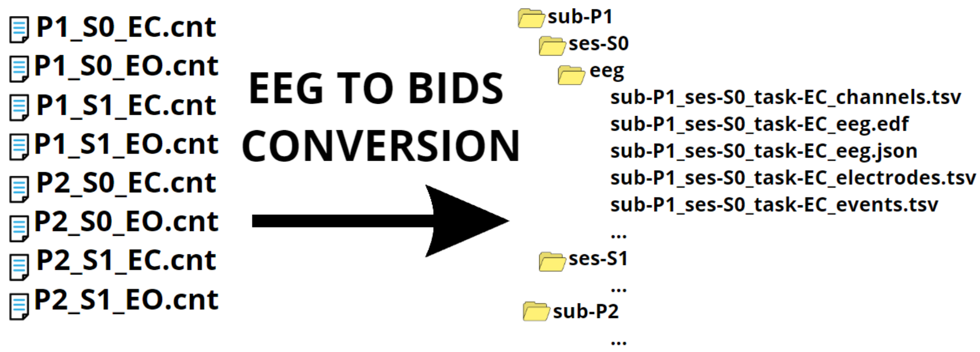

Figure 1: Illustration of the EEG to BIDS conversion. The left side shows raw EEG files with participant-specific naming conventions (for example, P1_S0_EC.cnt), where P1 and P2 represent participants, S0 and S1 indicate sessions, and EC (Eyes Closed) and EO (Eyes Open) refer to tasks. These raw files are converted into the BIDS format, shown on the right, where data are systematically organized into subject (sub-), session (ses-), and modality (eeg) folders. Each EEG recording is saved in standardized BIDS-compliant formats, including .edf for EEG signals and .tsv/.json for metadata.

Statement of need

Electroencephalography is a widely used neuroimaging technique that provides high temporal resolution for studying brain activity. Its applications span numerous fields, including cognitive neuroscience, clinical diagnostics, brain-computer interfaces, and neuroengineering. With the increasing volume and complexity of EEG data, ensuring reproducibility, standardization, and interoperability has become a growing priority in the field. The Brain Imaging Data Structure for EEG (EEG-BIDS) (Pernet et al., 2019) provides a consistent framework for organizing EEG datasets, facilitating data sharing (Markiewicz et al., 2021), large-scale collaborations, cross-study comparisons, and promoting FAIR data practices (Wilkinson et al., 2016) across a wide range of research applications from fundamental cognitive neuroscience to large-scale clinical neuroimaging.

The datasets generated in studies using electroencephalography are not only large and complex, but also vary widely due to the multitude of formats used by different hardware vendors. Thus, despite the advantages of BIDS, the conversion of EEG datasets to the BIDS standard can be challenging, especially for researchers who are not well-versed in technical data manipulation, and those from smaller or less well-resourced institutions. Researchers are thus obliged to wrestle with converting their data to BIDS following data acquisition. This process is prone to error if done manually, which poses particular problems for large studies. Software solutions are available to assist the conversion, but they require either basic programming skills (e.g., MNE-BIDS (Appelhoff et al., 2019), data2bids in FieldTrip (Oostenveld et al., 2011) and EEG-BIDS in EEGLAB (Delorme & Makeig, 2004)), or detailed user input for each file being converted, again limiting practicality for large studies as happens in EEG2BIDS (Rogers et al., 2022).

SOVABIDS addresses this challenge by enabling reproducible semi-automatic and interoperable conversion of EEG datasets into the BIDS standard, even by non-technical users. This democratizes access to BIDS compliance, enabling more EEG data to be shared and analyzed within the broader scientific community. Moreover, SOVABIDS facilitates the development of user-friendly graphical frontends, which further enhances its accessibility to a wider audience, including those who may not be familiar with programming or data structuring concepts. SOVABIDS leverages established EEG analysis tools, in particular MNE (Gramfort et al., 2013) for reading EEG data formats and MNE-BIDS (Appelhoff et al., 2019) for BIDS compliant data saving. It also incorporates open-source best practices like automated testing and streamlined documentation that includes usage examples, ensuring continuous enhancement, and facilitating community usage and collaboration in maintaining and improving the software.

Research Impact Statement

SOVABIDS is currently available on the Neurodesk platform www.neurodesk.org (Renton et al., 2024). It has been used in both academic research and published scientific studies, including 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 focused on harmonizing EEG features across multiple recording sites (Jaramillo-Jimenez et al., 2024).

Software Design

Developing an EEG-to-BIDS conversion tool requires balancing usability, automation, reproducibility, and flexibility while ensuring compatibility with existing neuroimaging tools. SOVABIDS was designed with these challenges in mind, prioritizing accessibility for non-technical users, handling variations in EEG data structures, and enabling seamless integration with other software. The following five design principles guided its development:

1. Adoption by non-technical users

To decrease the need of programming skills, the conversion uses human-readable and writable YAML configuration files rather than a scripting language. This approach was inspired by Bidscoin (Zwiers et al., 2022), a BIDS converter for MRI data. To maximise software adoption, step-by-step guides for SOVABIDS are provided.

2. Automation that can accommodate outliers

The planned output of EEG experiments is usually multiple identically-organised data structures, typically one for each participant. Nevertheless, in practice data organisation often varies slightly between participants; for example, due to temporary technical issues, the data recorded for some participants might be partial or have repeated segments. Like other conversion tools such as Bidscoin (Zwiers et al., 2022) and HeuDiConv (Halchenko et al., 2023), we leverage the generally similar data organisation across participants, but we extend this by allowing for non-identical data structures. This is done by utilising two configuration files (which are illustrated in Figure 2):

- The Rules File, which encodes the general conversion rules for a multiple-participant EEG dataset.
- The Mappings File, which is derived from the Rules File, and holds specific conversion rules for every individual participant.

The Journal of Open Source Software

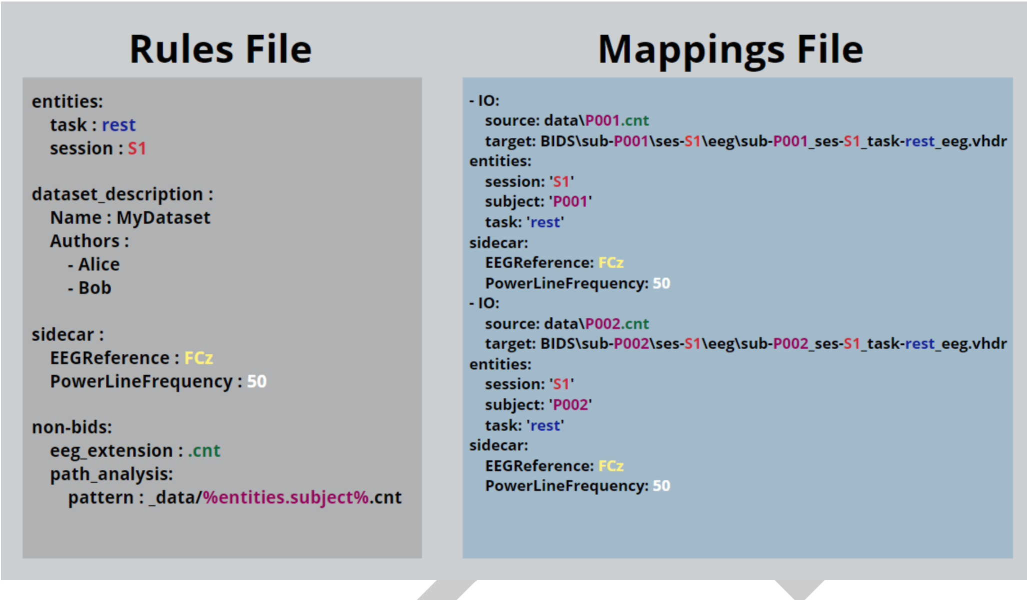

Figure 2: From a Rules File, a mapping for each file in the dataset can be generated and saved in the Mappings File. The colors illustrate how the information in both files is related.

In addition, the user can derive the initial Rules File from a community or institutional template, further decreasing manual input. For more fine-tuning of special cases, a user-supervised adjustment loop can be set up through SOVABIDS' interoperable API (Application Programming Interface) to connect with an external graphical user interface (GUI), allowing for fast manual inputs when a fully automated generation is not possible.

To support increased automation, SOVABIDS implements heuristics that take advantage of common file path patterns found in EEG research. Unlike tools that rely on strict prefix-based identification, such as the "–sub-prefix" and "–ses-prefix" options in Bidscoin (Zwiers et al., 2022) and Bidsme (Beliy et al., 2023), SOVABIDS enables flexible metadata extraction directly from file paths. It can extract not only subject and session information, but also other BIDS-relevant properties, including task labels. This flexibility is supported through multiple approaches:

- Regular expressions, for advanced users who require precise control.
- Placeholder-based templates, which offer an intuitive way to define rules (as shown in the "path_analysis" section of Figure 2).
- Paired source-target examples, designed for users without experience in regular expressions or placeholders.

3. Reproducible conversion

All the parameters needed to fully replicate the conversion are saved in the configuration files along with provenance information. This allows the user to evaluate, correct and repeat the conversion in case the BIDS-organised dataset has an invalid structure (as detected using a BIDS validator) or incorrect content (usually discovered during the analysis stage).

4. Interoperability

To increase maintainability, SOVABIDS does not include a graphical user interface, but its API allows interoperability with other applications, primarily GUI front ends (desktop or web-based). We used a remote procedure call (RPC) protocol, as its action-oriented design naturally fits the process-driven nature of data conversion workflows. Additionally, interoperability is enhanced through the YAML-based configuration files, which allow users to define conversion rules in a structured, human-readable format. This enables seamless integration with the many tools that support it without sacrificing the ease of managing the configuration files. To demonstrate the usability of the API, a basic GUI was developed in Flask.

5. Broad support of formats

SOVABIDS is designed to convert EEG datasets into the BIDS standard while accommodating diverse data formats. Since it relies on MNE-Python for reading electrophysiology files, any EEG format supported by MNE can be processed and converted. As of now, however, the software has been specifically tested with BrainVision (.vhdr) and Neuroscan (.cnt) filesonly.

Software Architecture

The software architecture of SOVABIDS is designed to streamline the process of converting EEG datasets to the BIDS format. As depicted in Figure 3, the system is built around two core modules that work in tandem to simplify this complex task:

- Rules Module: At the heart of SOVABIDS, the Rules Module is where the logic of conversion is enacted. Users define specific rules in the 'Rules File', which the module applies to target EEG files, extracting and compiling conversion parameters into a 'Mappings File'. This 'Mappings File' becomes the blueprint for the subsequent data transformation process, ensuring that the individualized nuances of each EEG file are accounted for. These nuances can be introduced through manual editing or through the interoperation with GUIs by leveraging the RPC API.
- Conversion Module: Acting upon the 'Mappings File', this module is responsible for the hands-on task of converting raw EEG data into the BIDS format. It's a crucial step that translates the preparatory work done by the Rules Module into a structured dataset aligned with the stringent requirements of the BIDS standard, allowing for better interoperability and reproducibility of research.

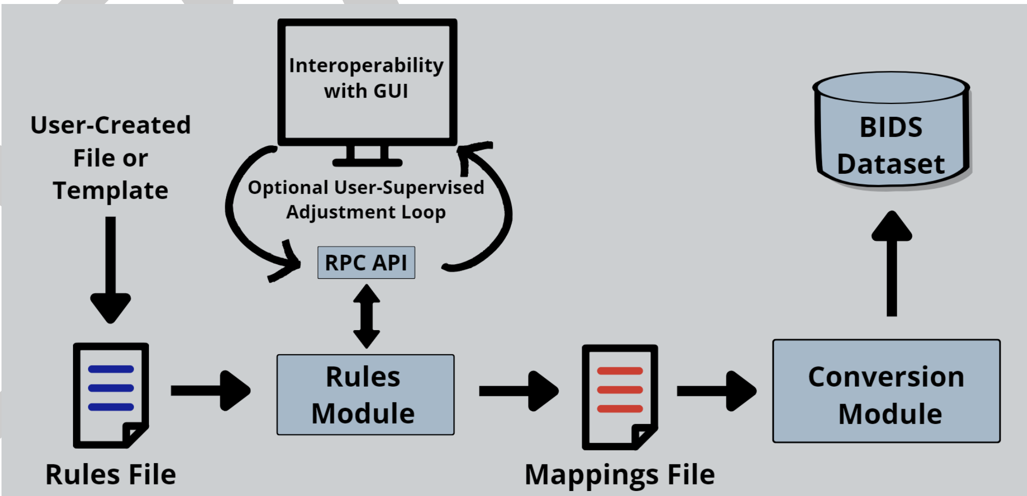

Figure 3: The architecture of SOVABIDS. The conversion process starts with a user-defined Rules File, which encodes general conversion rules (represented in blue inside the Rules File). The Rules Module processes these rules to generate a Mappings File, which contains specific configurations for all EEG files (each red line in the Mappings File represents the configuration of a different EEG file). The Conversion Module then applies these configurations to produce a BIDS-compliant dataset. Interoperability is enabled via an RPC API, allowing integration with external tools, including graphical user interfaces for optional user-supervised adjustments.

Acknowledgements

The authors acknowledge the support from the 2021 Google Summer of Code program under the International Neuroinformatics Coordinating Facility (INCF) organization, and the funding provided by the Australian Research Data Commons (ARDC) to support 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 primary architecture and core functionality of this software were completed prior to December 2023. Generative AI tools were not used in the conceptual design, methodological decisions, or scientific development of the project.

After March 2025, limited use of generative AI tools was made for maintenance and supporting tasks. The tools used were ChatGPT Codex (o4-mini) and ChatGPT (GPT-4.1).

AI assistance was used for: - Updating and improving GitHub Actions continuous integration workflows - Enhancing and clarifying documentation - Assisting in the implementation of a utility function for generating random 1/f signals - Minor code refactoring and formatting improvements - Minor improvements on the paper

All AI-assisted outputs were carefully reviewed, edited, tested, and validated by the authors. All core design decisions and scientific judgments were made by the human authors.

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., Guillemin, 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

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., Amlien, 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., & Brønnick, 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., Goncavles, 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, Ij. 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

DRAFT