




datashuttle: automated data management for experimental neuroscience

Joseph J. Ziminski ¹, Nikoloz Sirmopilatze ¹, Brandon D. Peri², Shrey Singh³, Sepiedeh Keshavarzi ², and Adam L. Tyson ¹✉

¹ Neuroinformatics Unit, Sainsbury Wellcome Centre & Gatsby Computational Neuroscience Unit, University College London, London, U.K ² Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom ³ Netaji Subhas University of Technology, New Delhi, India ✉ Corresponding author

DOI: [10.21105/joss.09642](https://doi.org/10.21105/joss.09642)

Software

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

Editor: [Jonny Saunders](#)  

Reviewers:

- [@adswa](#)
- [@danielmk](#)
- [@Julie-Fabre](#)

Submitted: 01 August 2025

Published: 16 June 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

Datashuttle is a Python package that facilitates data standardisation in neuroscience. Experimental data are often stored using custom folder structures and naming conventions, which hinders data sharing, reproducibility, and the development of community tools. Datashuttle addresses this by providing user-friendly tools to create, validate, and transfer standardised experimental data folders. The package can be used programmatically—integrated into existing Python scripts for data acquisition—or via a graphical user interface.

Statement of Need

The past decade has seen significant progress in the development of neuroscience data standards. Experimental datasets have become increasingly complex, with multiple modalities (e.g. behaviour, electrophysiology and imaging) often collected from a single subject. At its core, standardisation facilitates reproducibility by ensuring these complex datasets are well organised, accessible, machine-readable and sufficiently documented ([Martone, 2024](#)). This standardisation permits robust, automated project management including the transfer of experimental data between machines and validation of project contents. Detailed specifications covering folder, file and metadata naming and structural conventions are required to achieve this goal.

Development and dissemination of comprehensive standards have been driven by community organisations such as the International Neuroinformatics Coordinating Facility (INCF) ([Abrams et al., 2022](#)). This includes adoption of the FAIR principles ([Abrams et al., 2022](#); [Wilkinson et al., 2016](#)) ensuring data are findable, accessible, interoperable and reusable. Important standardisation initiatives include the Brain Imaging Data Structure (BIDS) ([Gorgolewski et al., 2016](#)), a file, folder and metadata standard widely used in neuroimaging, and the open file format Neurodata Without Borders (NWB) ([Rübel et al., 2022](#); [Teeters et al., 2015](#)). These initiatives aim to achieve ‘full’ standardisation that enables automated analysis of machine-readable experimental datasets.

There is a rich tooling ecosystem for working with standardised data. NWB provides packages for the reading, writing and validation of NWB files ([Baker et al., 2025](#); [Tritt et al., 2025](#)), alongside infrastructure for visualising and sharing ([Magland et al., 2025](#)). The BIDS community has developed packages for converting raw data to BIDS format, as well as reading, writing and validating BIDS formatted folders, files and metadata ([Gorgolewski et al., 2020](#); [Yarkoni et al., 2019](#)), while DataLad ([Halchenko et al., 2021](#)) supports the transfer of large datasets within similar workflows. ezBIDS ([Levitas et al., 2024](#)) offers a web-based GUI for BIDS conversion

and sharing, though it is tailored towards neuroimaging. While highly valuable, these tools generally require coding experience and a good understanding of the underlying data schemas, with functionality distributed across multiple packages.

The adoption of data standards in systems neuroscience is not yet widespread (Klingner et al., 2023). This is due in part to the inherent, and necessary, complexity required to achieve full standardisation (Pierré et al., 2024) and lack of tools to automate the full management of standardised projects without requiring coding experience. Further, not all systems neuroscience methods have a corresponding data specification (e.g. fibre photometry). Researchers often default to custom folder structures in lieu of full standardisation, leading to inconsistencies both across and within laboratories.

Datashuttle aims to bridge the gap between no standardisation and full standardisation by implementing a simple specification called 'NeuroBlueprint' (Ziminski et al., 2025) (Figure 1). NeuroBlueprint mandates only folder naming and structure conventions, while recommending file and metadata-naming schemes. It is designed to be easy to adopt, meaning it is suitable for the busy data-acquisition stages of a project in which applying full standardisation is often too onerous.

NeuroBlueprint is primarily a folder-organisation standard and is therefore not, on its own, sufficient to ensure data are FAIR. Consistent folder structures and naming conventions improve the findability of datasets and lower the barrier to depositing data in shared repositories, facilitating accessibility. However, interoperability and reusability require standardised file formats and metadata. NeuroBlueprint is heavily inspired by BIDS in order to minimise redundancy between specifications and facilitate later transition to this more comprehensive schema. As a result, NeuroBlueprint provides a low-overhead starting point for standardised data management while remaining compatible with a future extension toward fully FAIR-compliant workflows.

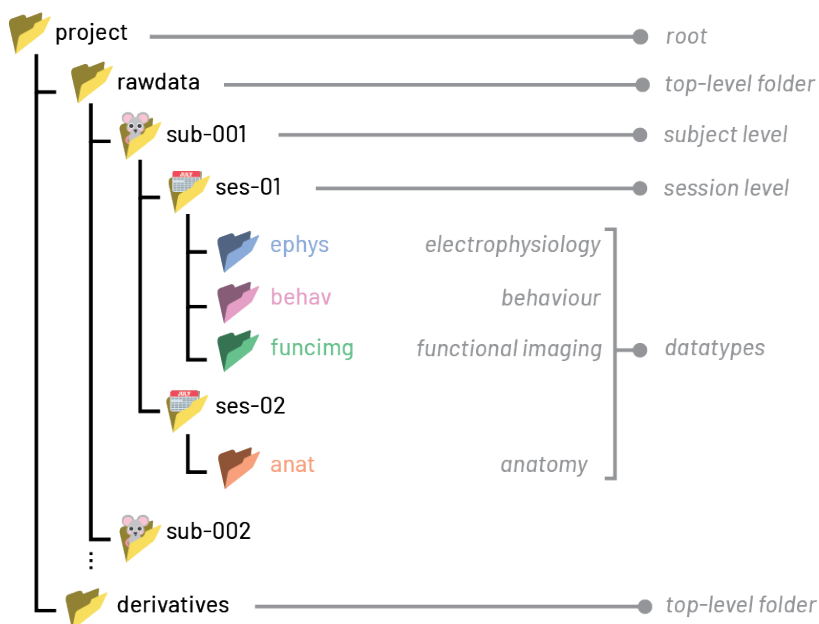


Figure 1: The NeuroBlueprint specification. Raw data (i.e. as collected from acquisition machines) are organised hierarchically by subject, session, and datatype. Subject and session names consist of key-value pairs. Only the sub- and ses-keys are required and others are optional. Acquired data are placed in a datatype folder, with valid datatype names defined in the specification. Derived data are stored in the top-level derivatives folder, and while not mandated, it is advised to organise these similarly to the rawdata directory.

Datashuttle automates the creation, validation and transfer of experimental folders in NeuroBlueprint standard. It is designed to drop into existing scripted or manual data-acquisition pipelines, ensuring standardisation at the point of data collection. Datashuttle offers flexible data transfer capabilities that make standardisation practical and convenient, rather than an added burden. With minimal dependencies and no lock-in, it provides a lightweight, adaptable solution for managing neuroscience project workflows.

Features

Datashuttle can be installed via the package manager conda. While pip installation is supported, the non-Python dependency RClone (used to manage data transfers) must be installed separately. The cross-platform terminal user interface (TUI) is built with Textual (McGugan, 2021) and can be used in the system terminal.

The typical workflow begins with researchers creating standardised folders at the start of each experimental session. Data generated during acquisition (e.g. from cameras, behaviour-monitoring devices or electrophysiology probes) are saved into the created folders. Real-time validation features ensure that common errors such as duplicate subject or session IDs are caught immediately. At the end of the experimental session, data are transferred to the laboratory's central storage. Transfers can be made to a remote server either via a mounted drive or SSH, while cloud services such as Google Drive and AWS S3 Buckets are also supported.

Folder Creation

NeuroBlueprint-formatted folder trees can be created for a given subject, session and datatype (e.g. 'funcimg' for functional imaging), with online validation to reduce the likelihood of errors in user input (Figure 2).

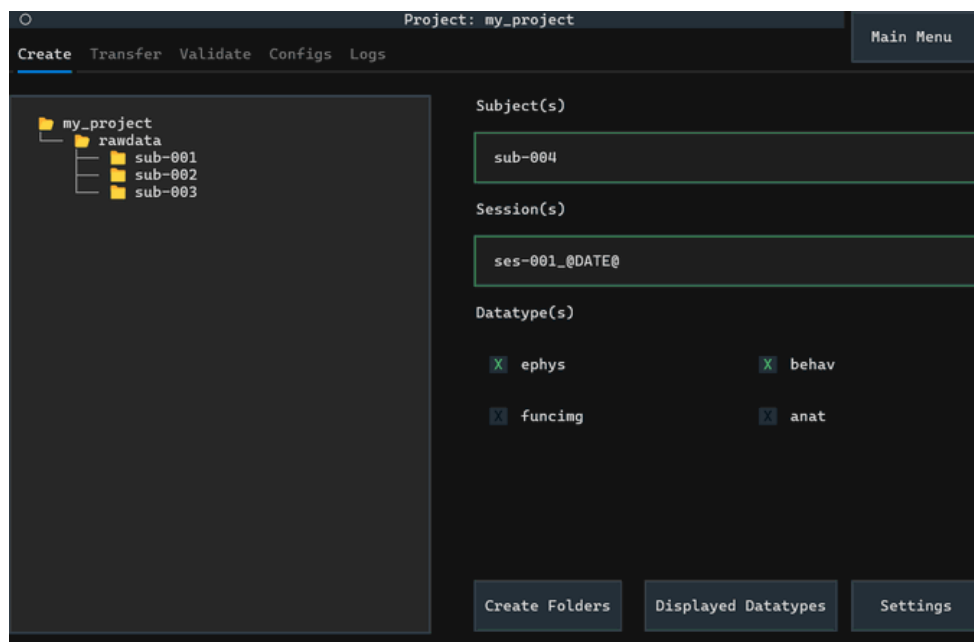


Figure 2: The Create Folders screen. Subject, session and datatype folders can be created through this interface. The text input border provides real-time validation results, while tags such as @DATE@ can be used to auto-format the system date. The current project is displayed on the left-hand directory tree, which can be used to copy file-paths and open the operating filesystem.

Validation

Validation catches issues such as duplicate subject or session IDs, inconsistent number of leading zeros, bad key-value pair formatting and other common typographical errors. Validation can be performed on an entire project, listing any formatting errors that are discovered. Additionally, real-time validation during folder creation ensures no non-NeuroBlueprint format folders can be made. Custom extensions to the validation can be added, with subject and session names validated against user-defined regular expression templates.

Data Transfer

Datashuttle uses the open source tool RClone ([Craig-Wood, 2014](#)) to perform data transfers. Experimental data can be 'uploaded' (from the local machine to central storage) or 'downloaded' (from the central storage to the local machine) ([Figure 3](#)). A benefit of standardisation is machine-readable folder names—meaning it is simple to select arbitrary subsets of data for transfer, e.g. only the first five subjects.

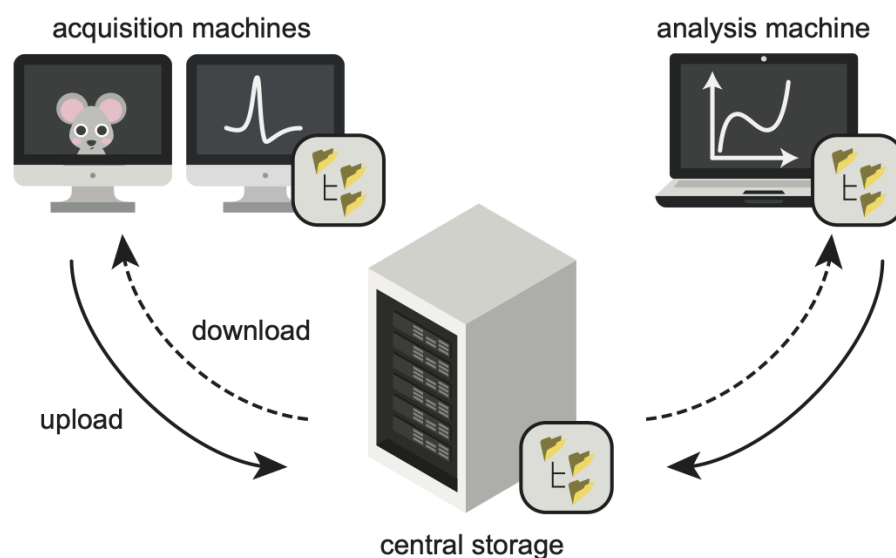


Figure 3: Data transfers in datashuttle. A typical workflow involves transferring data from an acquisition machine to a central laboratory storage. Later, the entire dataset or subsets of it (e.g. only electrophysiology data) may be downloaded to an analysis machine for processing.

Logging

In order to track full provenance of the project, datashuttle operations are logged to file with fancylog ([Ziminski & Tyson, 2025](#)). Logs can be accessed directly from disk or displayed in the TUI.

Future Directions

Datashuttle will continue to evolve alongside the NeuroBlueprint specification, implementing upcoming extensions as they emerge. While datashuttle does not currently support a metadata standard, implementation of the simple, lightweight metadata schema recommended by NeuroBlueprint for basic project documentation will be a key focus for future development. While full FAIR compliance is out of scope for NeuroBlueprint itself, datashuttle plans to

facilitate transition toward more comprehensive schemas by supporting automated conversion of NeuroBlueprint projects to BIDS, including metadata conversion, alongside raw data file conversion to open formats such as NWB.

Currently, NeuroBlueprint is designed for experiments in which subjects go through the experimental procedures individually. However, multi-animal experiments investigating social behaviours, in which animals interact during experimental sessions, are a growing area of neuroscience research. Future updates to both NeuroBlueprint and datashuttle will aim to support this use case.

Availability

Datashuttle's source code is available at <https://github.com/neuroinformatics-unit/datashuttle> and documentation published at <https://datashuttle.neuroinformatics.dev>.

Acknowledgements

J.J.Z., N.S. and A.L.T. were funded by the core grant to the Sainsbury Wellcome Centre (Wellcome - 219627/Z/19/Z, Gatsby Charitable Foundation - GAT3755). A.L.T. was funded by the core grant to the Gatsby Computational Neuroscience Unit (Gatsby Charitable Foundation - GAT3850). B.P. and S.K. are funded by a Wellcome Trust Career Development Award (226039/Z/22/Z to S.K.). We also thank Laura Porta, Alessandro Felder and Igor Tatarnikov for comments on the manuscript and codebase and all contributors to the datashuttle project.

References

- Abrams, M. B., Bjaalie, J. G., Das, S., Egan, G. F., Ghosh, S. S., Goscinski, W. J., Grethe, J. S., Kotaleski, J. H., Ho, E. T. W., Kennedy, D. N., Lanyon, L. J., Leergaard, T. B., Mayberg, H. S., Milanese, L., Mouček, R., Poline, J. B., Roy, P. K., Strother, S. C., Tang, T. B., ... Martone, M. E. (2022). A standards organization for open and FAIR neuroscience: The international neuroinformatics coordinating facility. *Neuroinformatics*, 20. <https://doi.org/10.1007/s12021-020-09509-0>
- Baker, C., Dichter, B., Ly, R., Prince, S., Weigl, S., Trapani, A., Mayorquin, H., Ruebel, O., Halchenko, Y., tom, Wodder II, J. T., daphnedequatebarbes, Buccino, A., Sprague, Y., Elpy, Adkisson, P., & McKenzie, Z. (2025). *NeurodataWithoutBorders/nwbinspector: v0.6.4* (Version 0.6.4). Zenodo. <https://doi.org/10.5281/zenodo.16415807>
- Craig-Wood, N. (2014). *RClone [software]*. <https://rclone.org/>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.44>
- Gorgolewski, K. J., Hardcastle, N., Hobson-Lowther, T., Nishikawa, D., Blair, R., Appelhoff, S., Suyash, constellates, Jas, M., Holdgraf, C., Jones, A., Goyal, R., Oostenveld, R., Markiewicz, C., noack, G., Zito, M., Durnez, J., Traut, N., Naveau, M., ... Thomas, A. (2020). *Bids-standard/bids-validator: 1.4.3* (Version 1.4.3). Zenodo. <https://doi.org/10.5281/zenodo.3688707>
- Halchenko, Y. O., Meyer, K., Poldrack, B. A., Solanky, D. S., Wagner, A. S., Gors, J., MacFarlane, D., Pustina, D., Sochat, V., Ghosh, S. S., Mönch, C., Markiewicz, C. J., Waite, L., Shlyakhter, I., Vega, A. de la, Hayashi, S., Häusler, C. O., Poline, J.-B., Kadelka, T., ... Hanke, M. (2021). DataLad: Distributed system for joint management

- of code, data, and their relationship. *Journal of Open Source Software*, 6(63), 3262. <https://doi.org/10.21105/joss.03262>
- Klingner, C. M., Denker, M., Grün, S., Hanke, M., Oeltze-Jafra, S., Ohl, F. W., Radny, J., Rotter, S., Scherberger, H., Stein, A., Wachtler, T., Witte, O. W., & Ritter, P. (2023). Results of a community survey of the German national research data infrastructure initiative neuroscience. *eNeuro*, 10, ENEURO.0215–22.2023. <https://doi.org/10.1523/ENEURO.0215-22.2023>
- Levitas, D., Hayashi, S., Vinci-Booher, S., Heinsfeld, A., Bhatia, D., Lee, N., Galassi, A., Niso, G., & Pestilli, F. (2024). ezBIDS: Guided standardization of neuroimaging data interoperable with major data archives and platforms. *Scientific Data*, 11, 179. <https://doi.org/10.1038/s41597-024-02959-0>
- Magland, J. F., Ly, R., Rübél, O., Dichter, B., Schaffer, E., Hanke, M., Holdgraf, C., Varoquaux, G., Poldrack, R. A., Halchenko, Y. O., Markiewicz, C. J., Poline, J.-B., Esteban, O., & Appelhoff, S. (2025). Facilitating analysis of open neurophysiology data on the DANDI archive using large language model tools. *Scientific Data*, 12, 1988. <https://doi.org/10.1038/s41597-025-06285-x>
- Martone, M. E. (2024). The past, present and future of neuroscience data sharing: A perspective on the state of practices and infrastructure for FAIR. *Frontiers in Neuroinformatics*, 17. <https://doi.org/10.3389/fninf.2023.1276407>
- McGugan, W. (2021). *Textual [software]*. <https://github.com/Textualize/textual>
- Pierré, A., Pham, T., Pearl, J., Datta, S. R., Ritt, J. T., & Fleischmann, A. (2024). A perspective on neuroscience data standardization with neurodata without borders. *The Journal of Neuroscience*, 44. <https://doi.org/10.1523/JNEUROSCI.0381-24.2024>
- Rübél, O., Tritt, A., Ly, R., Dichter, B. K., Ghosh, S., Niu, L., Baker, P., Soltesz, I., Ng, L., Svoboda, K., Frank, L., & Bouchard, K. E. (2022). The neurodata without borders ecosystem for neurophysiological data science. *eLife*, 11. <https://doi.org/10.7554/eLife.78362>
- Teeters, J. L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., Denisov, G., Siegle, J. H., Olsen, S. R., Martin, C., Chun, M., Tripathy, S. S., Blanche, T. J., Harris, K. D., Buzsáki, G., ... Sommer, F. T. (2015). Neurodata without borders: Creating a common data format for neurophysiology. *Neuron*, 88(4), 629–634. <https://doi.org/10.1016/j.neuron.2015.10.025>
- Tritt, A., Dichter, B., Rübél, O., Ly, R., Fillion-Robin, J.-C., Nicain, Braun, T., Prince, S., NileGraddis, Avaylon, M., Ozturk, D., Davidson, T., Mayorquin, H., Baker, C., Weigl, S., Bolton, J. R., Halchenko, Y., Melara, M., Tom, ... Magland, J. (2025). *NeurodataWithoutBorders/pynwb: 3.1.1* (Version 3.1.1). Zenodo. <https://doi.org/10.5281/zenodo.16347328>
- 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. <https://doi.org/10.1038/sdata.2016.18>
- Yarkoni, T., Markiewicz, C. J., Vega, A. de la, Gorgolewski, K. J., Salo, T., Halchenko, Y. O., McNamara, Q., DeStasio, K., Poline, J.-B., Petrov, D., Hayot-Sasson, V., Nielson, D. M., Carlin, J., Kiar, G., Whitaker, K., DuPre, E., Wagner, A., Tirrell, L. S., Jas, M., ... Blair, R. (2019). PyBIDS: Python tools for BIDS datasets. *Journal of Open Source Software*, 4(40), 1294. <https://doi.org/10.21105/joss.01294>
- Ziminski, J. J., Sirmipilatz, N., Porta, L., Plattner, V., Peri, B. D., Keshavarzi, S., & Tyson, A. L. (2025). *NeuroBlueprint*. Zenodo. <https://doi.org/10.5281/zenodo.15720970>

Ziminski, J. J., & Tyson, A. L. (2025). *Fancylog* [software]. <https://doi.org/10.5281/zenodo.15776028>