

DATA DESCRIPTOR

The In2PrimateBrains Data Descriptor complements data shared by the consortium on the data sharing platform GIN (<https://gin.g-node.org/In2PrimateBrains>). A rich data descriptor improves the comprehension of your data and increases the chances of reuse. **Please fill in all sections.** You may use tables and figures (with captions) for improved readability where you see fit. References should follow the standard Nature referencing style. This template is based on data descriptor template of the data repository EBRAINS (RRID:SCR_019260, <https://www.ebrains.eu/data/find-data/>) and the journal Scientific Data (ISSN 2052-4463, <https://www.nature.com/sdata/>).

TITLE

Maximum 110 characters including spaces. No colons and parentheses.

The title should be descriptive for the presented data. The use of acronyms and abbreviations should be avoided where possible. Please also provide a abbreviated title to be used. This could use the project numbers, host name or lab name.

[Add the title here]

AUTHORS

Forename Surname¹, Forename Surname¹, and Forename Surname²

[Add author names here]

AFFILIATIONS

Please provide the affiliations for all authors. Format examples:

1. Institute (working group) department, university,city, country
2. department, organization, city, country)

[List each institution here]

CORRESPONDING AUTHOR(S):

Please name the corresponding author(s) formatted as:

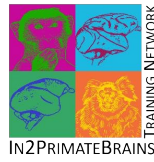
Forename Surname: email@address

[List corresponding author information here]

SUMMARY

Maximum 220 words, no references.

Shortly summarise the content, provenance and usage potential of the shared data: include information about subjects, methods and materials, relevant study targets and in which context the data could be (re)used. Do not describe the research results, but describe purely the shared data. The summary should be written so it is readily intelligible to any scientist. Specialised terms should be explained concisely.



Example structure: Write one sentence about the origin of the dataset. Write one or two sentences that give an overview of the data. This could for example include the content of the dataset, the number of subjects and sessions, and the original purpose of the data. Write one or two sentences describing the methods and materials used. Write one sentence about the reusability of the data.

[Add the summary here]

MATERIALS AND METHODS

*Describe the materials and methods used to generate the data. Include information on all specimen, acquisition techniques, processing steps (incl. validations and quality assessments), tools and workflows as well as information on the relevant anatomical entity and location for this study. Use **subheadings** for improved readability.*

Please note: It is accepted to cite previous methodological descriptions (cf. [guidelines](#) on how to cite your own work), but the descriptions should be complete enough for others to understand the materials and methods without reading the associated publications. Please ensure to closer define the location of the reference within your publication (e.g. under which subheading, on which page).

Anatomical information: *If applicable, please provide sufficient information on the entity of the data (e.g. cell type, histological tissue section, etc.) specific anatomical location the data were collected from (e.g. brain region name or exact coordinate points), and how the anatomical location was defined (e.g. used brain atlas version, cell type ontology, applied coordinate system/template space). If data represent anatomical images and/or coordinate(s) please always provide the used coordinate system/template space. This might be the subject's native space, or a common coordinate system (e.g., MNI Colin 27 v1998) if data were transformed.*

[Add the materials and methods here]

USAGE NOTES

*Provide suggestions for **how the shared data can be (re)used and warn about how they should not be (re)used** (e.g. “the used spike sorting and quality assurance make this dataset particularly interesting for precise spike time analysis”, “data are generated from adult subjects, making it unsuitable for developmental studies”, “the image magnification is too low for cell quantification”, “the spike data contain artefacts rendering it unsuitable for ...”).*

*Also provide suggestions for **what software/code can be used** with the shared data (e.g. loading routines, analyses, pipelines, visualisation tools), **including where that software/code can be found**.*

[Add the usage notes here]

DATA RECORDS

Describe how the data are stored, as well as their file types and content.

1. Provide a reduced generic layout of the file repository structure displaying the storage convention (see example below). Add a one sentence summary describing the content of each (example) file in square brackets. If a (meta)data standard was followed (e.g. In2PrimateBrains common localizer standard, or BIDS), include this information. E.g.:

The data are stored in the following structure (incl. info on file content):

/repository-root

/sub_info.tsv *[contains information on the subjects]*
 /sub-XXX
 /sub-XXX_slice-XXXX.tif *[brain slice scan of subject XXX; XXXX = physical slice position in μm , anterior to posterior]*
 /derived-data
 /analysis-X_info.json *[structured info on applied analysis]*
 /analysis-X
 /analysis-X_set-X.tsv *[resulting data of analysis X that is defined in 'analysis_info.json']*
 /code
 /analysis-X.py *[script that produces result data of analysis X]*

2. List all file formats and include which software created the respective files. If files were created manually, please explain their internal structure. E.g.

Format	Extension	Software used / file specification
Tab-Separated Value	tsv	self made sub_info.tsv: column headers: 'id', 'species', 'biological sex'; one row per subject analysis-X_set-X.tsv: column headers: x, y, z coordinates in μm , brain region; one row per point of interest including it's semantic location (brain region)
JavaScript Object Notation	json	generated by analysis-X.py analysis-X_info.json: keys: brain atlas, brain atlas version, certainty level used for coordinate location
Tagged Image Format File	tif	generated by Zen software v2.6, Blue edition / 100 μm thickness; 1x1 pixel dimension; unit in μm
Python Script	py	Python 3.0 script

[Add the data records here]

Acknowledgements

This section should contain brief acknowledgements of **non-author contributors and funding** (incl. grant name and number if possible). Anonymous referees and editors or effusive comments are not accepted. This **MUST** contain reference to the In2PrimateBrains consortium in addition to any additional funding.

Example use - "This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956669".

[Add acknowledgements here]

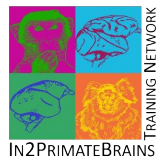
Author contributions [OPTIONAL]

This section should briefly state **each author's contribution** (on a separate line) in producing or maintaining the presented data as well as their role in publishing them. E.g.:

author X: data analysis...

author Y: image acquisition, data analysis...

[Add contributions here]



REFERENCES

*This section should list all bibliographic information for **all literature cited** in the above sections using the standard Nature referencing style.*

[Add references here]