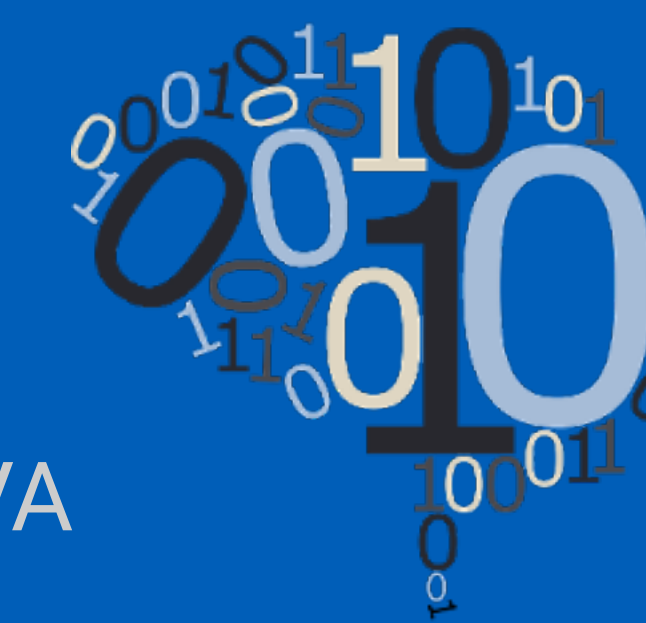# ndio: user-facing tools toward scalable neuroscience

J. MATELSKY[12], S. BERG[3], A. EUSMAN[1], K. LILLANEY[1], J. T. VOGELSTEIN[1], G. D. HAGER[1], W. R. GRAY RONCAL[12]

1. Johns Hopkins University, Baltimore, MD; 2. JHU Applied Physics Laboratory, Laurel, MD; 3. Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA

Abstract

Neuroscience datasets continue to grow in size and complexity, but many state-of-the-art data-access and manipulation tools remain too simple to easily handle large-scale interaction. Existing solutions may be too complex for easy use in a research environment and can limit analysis to scientists with a strong computational background.

ndio exposes a simple but robust Python API for:

- retrieval and posting of large-scale data from neuroscience datastores
- interactions with datastores such as:
  - NeuroData (Johns Hopkins University)
  - DVID (Howard Hughes Janelia Research Campus)
  - the BOSS (the Applied Physics Laboratory)
  - Local disk

ndio also provides multiple resources to aid the computationally savvy user with complex tasks such as batch-processing or parallel manipulations across several cores or threads, improving the accessibility of big-data, open, reproducible neuroscience.

Design and Implementation

ndio abstracts big-data repositories into *Remotes*, Python objects that hold no data but handle downloads and uploads, relevant metadata, and other housekeeping functionality.

Remotes expose functions to interact with a particular datastore's unique API, but also implement consistent interfaces for volumetric data-access:

- Download volumetric data
- Upload volumetric data
- Download labels and metadata
- Upload labels and metadata
- For RAMON datastores: Storage/retrieval of the full RAMON spec

ndio-downloaded data are represented in industry-standard Python types, such as numpy, to allow trivial import and export to common filetypes (npz, hdf5, png, multipage-tiff, ply, obj, etc).
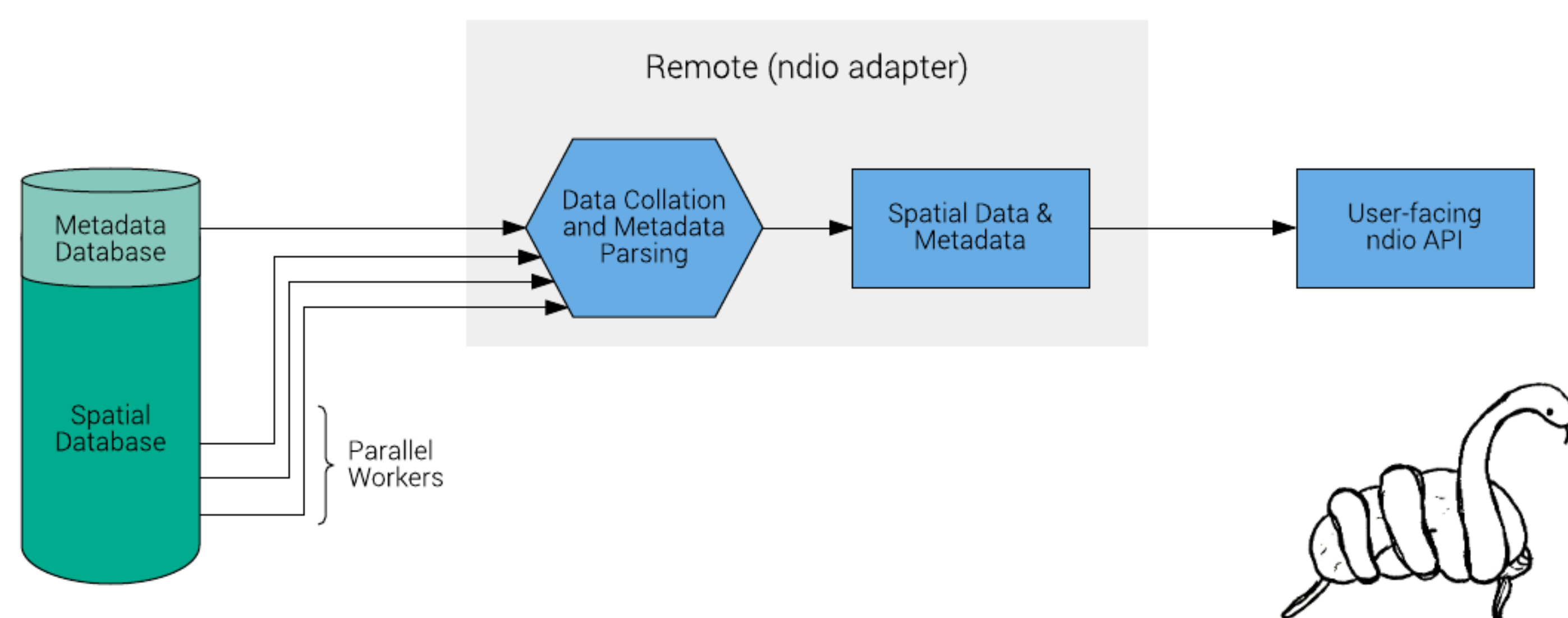
Install and Pull Data in 2 Minutes

pip install ndio or clone from https://github.com/neurodata/ndio

```python
import ndio.remote.neurodata as neurodata
nd = neurodata()
query = {
    "token": "kasthuri11cc",
    "channel": "image",
    "x_start": 694,   "x_stop": 1794,
    "y_start": 1750,  "y_stop": 2460,
    "z_start": 1004,  "z_stop": 1379,
    "resolution": 3
}

shiny_new_numpy_array = n.get_cutout(**query)
```

This example uses data from the Kasthuri, et al. 2015 *Cell* paper, *"Saturated Reconstruction of a Volume of Neocortex"* [1], hosted in the cloud by NeuroData [2].

Use the above code to download a sample image data volume. Example example slices are shown (at right). NeuroData stores annotations according to the RAMON data standard, which contains labels (paint) and metadata about each object. These annotations can be downloaded with token/channel ac3ac4/ac3_neuron_truth.

ndio.utils.mesh submodule (or any other 3D mesh-building packages): generate 3D models of neuron reconstructions (at right).

ndio.convert.hdf5: Export the native numpy data to hdf5 for storage.
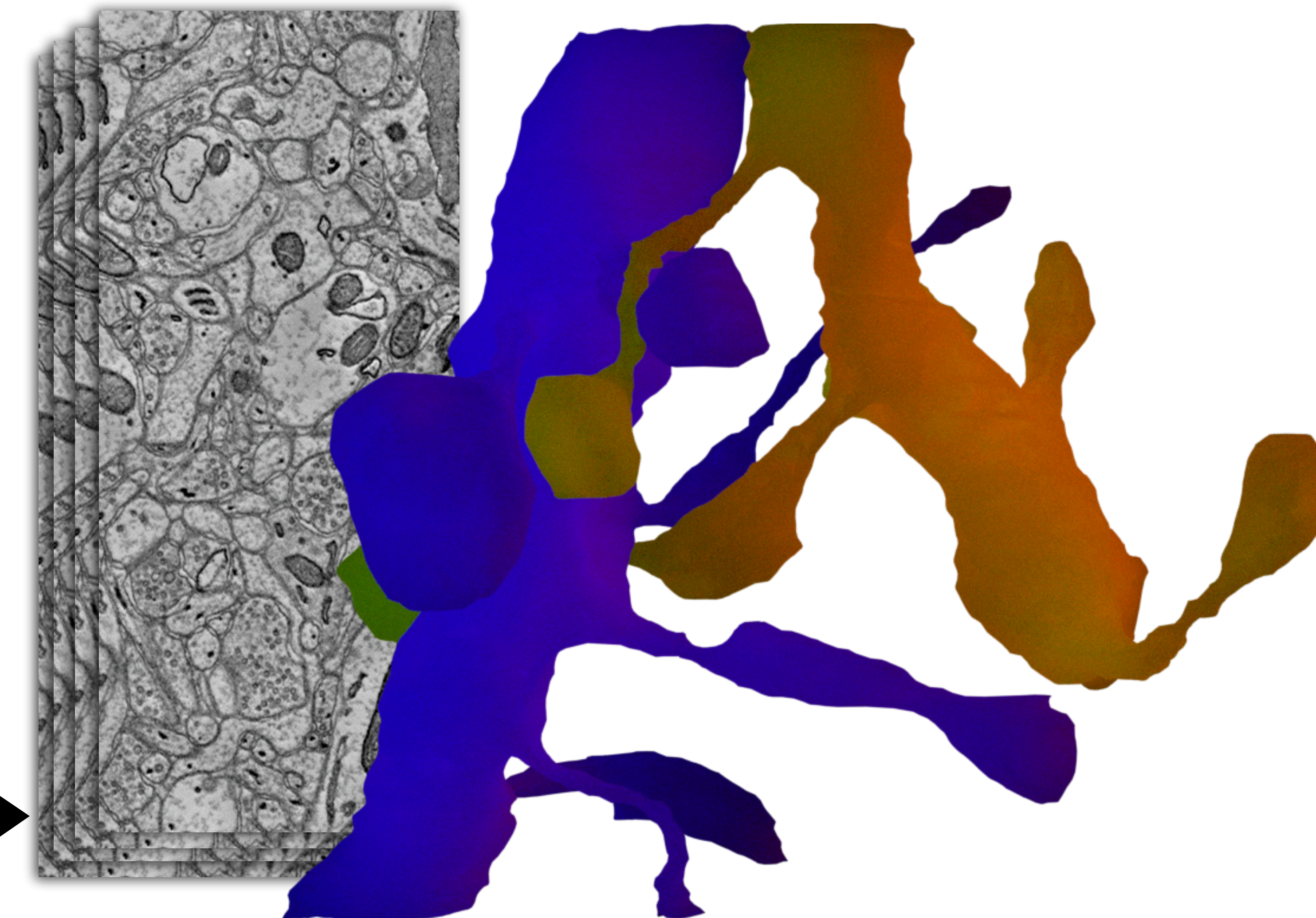
Links and References

References

[1]. N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. A. Conchello, S. Knowles-Barley, D. Lee, A. Vázquez-Reina, V. Kaynig, T. R. Jones, M. Roberts, J. L. Morgan, J. C. Tapia, H. S. Seung, W. Gray Roncal, J. T. Vogelstein, R. Burns, D. L. Sussman, C. E. Priebe, H. Pfister, and J. W. Lichtman, "Saturated Reconstruction of a Volume of Neocortex," *Cell*, vol. 162, no. 3, pp. 648–661, 2015.

[2]. R. Burns, W. Gray Roncal, D. Kleissas, K. Lillaney, P. Manavalan, E. Perlman, D. R. Berger, D. D. Bock, K. Chung, L. Grosenick, N. Kasthuri, N. C. Weiler, K. Deisseroth, M. Kazhdan, J. Lichtman, R. C. Reid, S. J. Smith, A. S. Szalay, J. T. Vogelstein, and R. J. Vogelstein, "The Open Connectome Project Data Cluster: Scalable Analysis and Vision for High-Throughput Neuroscience," *Proc. 25th Int. Conf. Sci. Stat. Database Management,* 2013.

Links

GitHub Repository: https://github.com/neurodata/ndio
Documentation: https://docs.neurodata.io/nddocs/ndio

Acknowledgements

ndio Use Cases

ndio provides a framework for a variety of big-data neuroscience interactions:

- Storing, organizing, and retrieving large data (locally or in the cloud) for a research project
- Counting cells in a tissue sample, using ndio's implementation of the RAMON neuroscience data standard.
- Generating 3D visualization of isolated neurons or subvolumes (at right)
- Batch processing across cores and threads