Utilizing the Elephant and NetworkUnit frameworks within the Collaboratory for an HPC-enabled workflow

Alper Yeğenoğlu[1], Robin Gutzen[1], Michael Denker[1], Sonja Grün[1,2]

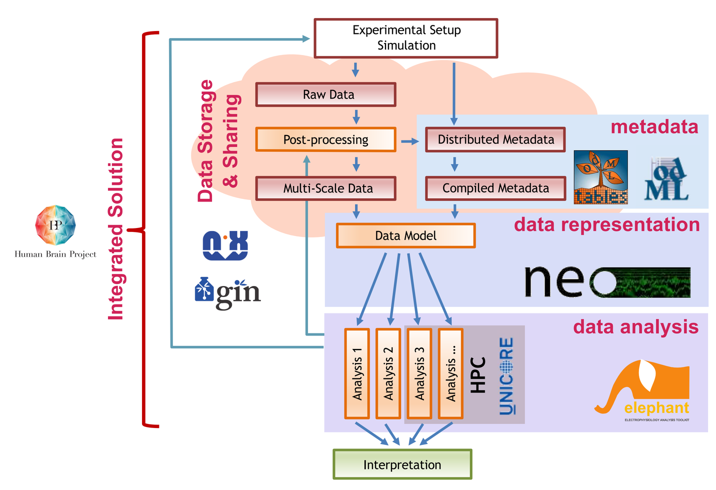
[1] Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany

[2] Theoretical Systems Neurobiology, RWTH Aachen University, Germany

Contact: a.yegenoglu@fz-juelich.de | contact@python-elephant.org

Introduction

The need for **reproducible research** is a topic of intense discussion in the neurosciences. In the context of **data analysis**, we develop the Electrophysiology Analysis Toolkit (Elephant, [1]) as a central resource to provide tested and validated reference implementations of common analysis methods for activity data. However, reproducibility also requires such tools to be embedded in **collaborative, holistic workflows** [2] providing **clear, traceable** analysis steps from data acquisition to publication.

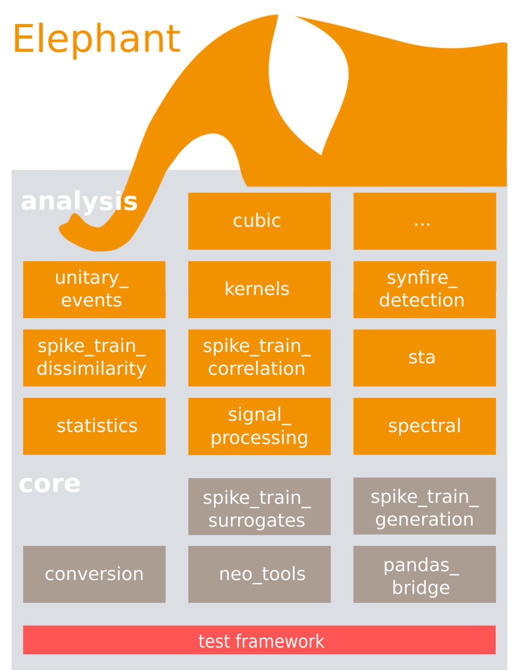


Here, we showcase how Elephant is integrated into an analysis workflow running on the the Collaboratory, reproducing work in [3]. The workflow consists of complementary open-source tools and services [4] for: metadata management (odML and odMLtables, [5,6,7]), data query (HBP Knowledgegraph, [8]), data versioning (gin, [9]), data storage (nix, [10]), data handling (Neo, [11]), and usage of high performance computing (HPC) using Unicore [12]. Finally, we outline how these building blocks, combined with generic tools (e.g., version control systems), can be assembled into formalized workflows to form a **blueprint for performing collaborative work** including access to high-performance computing.

Analysis using Elephant

Elephant is a community-centered, open-source software package that provides components for the analysis of multi-scale electrophysiological data (e.g., spike trains, local field potentials) from experiments and neuronal simulations, focusing on:

- methods for the analysis of parallel recordings
- correlative features of brain dynamics
- bridging different scales of observation



Summary

The presented analysis workflow. . .

- . . . **combines** several public, community-centered software tools to achieve a reproducible analysis.
- . . . is suitable for **collaborative work** between laboratories by use of the HBP Collaboratory.
- . . . provides a comprehensible data flow **independent of the data format** using the Neo library.
- . . . leads the way towards the implementation of **future analysis workflows** based on the Elephant library.

Find further resouces:

http://python-elephant.org
https://github.com/NeuralEnsemble/elephant

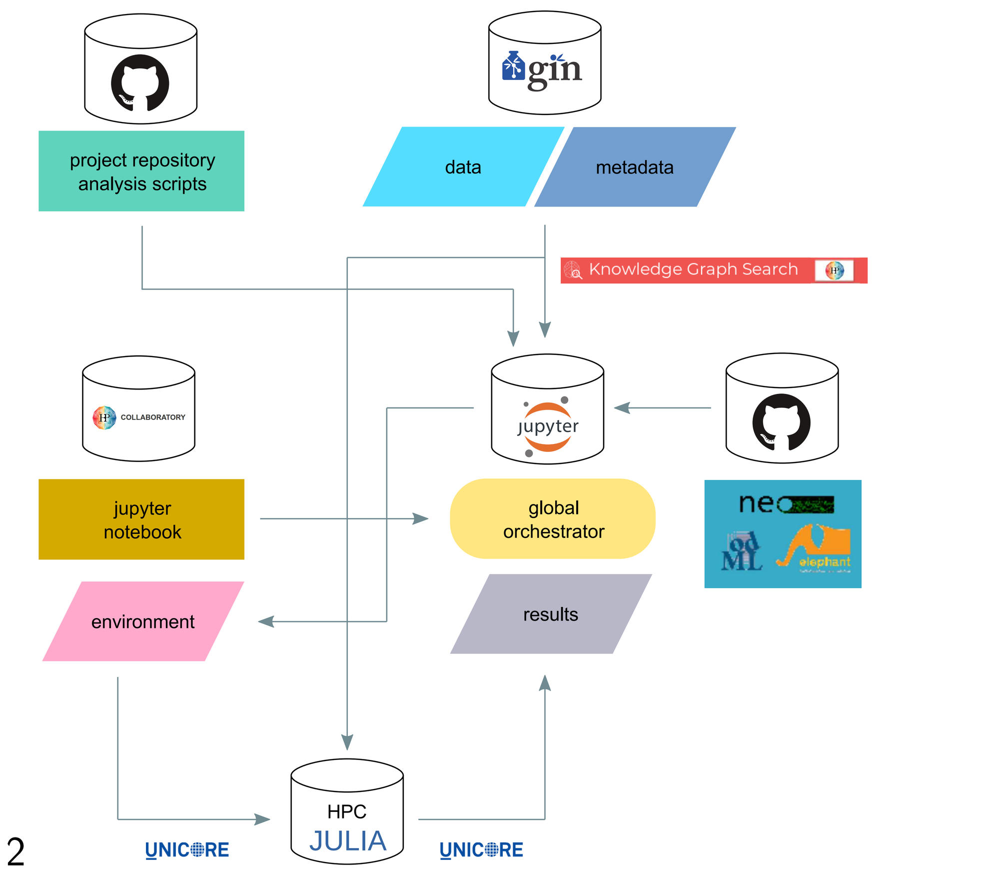
Analysis workflow on the HBP Collaboratory

The Collaboratory infrastructure of the Human Brain Project hosts the overall workflow of the project, required tools, central data management, data search, HPC access, and the ability for interactive work.

The Collab shown here [13] reproduces main findings of [3] using a Collab-centric workflow. The work concerns the analysis of spatio-temporal beta activity in motor cortex of macaque during a reach-to-grasp task [14].

The core of the project consists of a **Jupyter notebook** in the Collab, an associated **git repository** at github, and the datasets stored in an external **git-annex** based data store.

- **Jupyter notebook** in the Collaboratory:
  - Central control instance to orchestrate and display the analysis workflow
  - Construction of the environment
  - Visualization of the results
  - Documentation of the analysis process
- **Git repository**:
  - Analysis scripts are developed off-line and re-tained, version-controlled and shared in a git repository
  - Jupyter notebook synchronizes the repository to the Jupyter notebook container
  - Repository contains dependency file for software installations and and data requirements
- Unicore:
  - Delegates analysis to an HPC resource (JULIA)
  - Stage-in installs all software requirements
  - Transfers environment to JULIA via Collab storage
  - Executes the corresponding scripts on JULIA
  - Transfers results back to the Jupyter notebook
- Elephant/Neo:
  - Base libraries for performing analysis work
  - Analysis functions are chained together in the project analysis scripts
  - Data is read via the blackrockIO file connector provided by the Neo library



2
Data process workflow incorporating several facilities and tools, controlled via a single Jupyter notebook within the HBP Collaboratory.

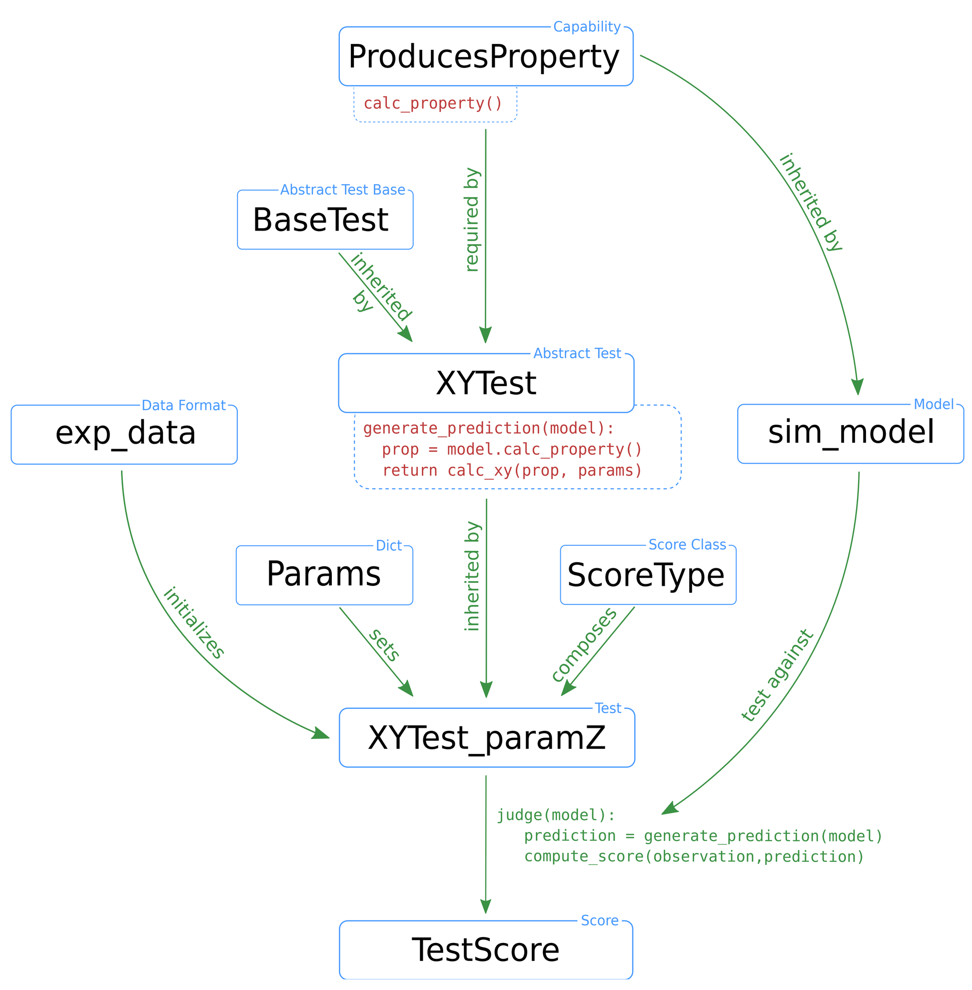
- odML/odMLtables:
  - Metadata is read from associated odML file
  - Selected metadata are used as annotations to the created Neo object
- gin:
  - Experimental data are published on the gin repository under version control
- Knowledgegraph:
  - Data are registered and located using the HBP Knowledgegraph service and the pyxus API

Validation with NetworkUnit

- **Validation** is the process of establishing confidence in a model by quantitatively testing whether its prediction accuracy is within an acceptable agreement to its system of interest.
- **Network-level validation** evaluates the model simulation on the level of the network activity as opposed to the complementary approach of validating on a single-cell level.
- **Model-to-model validation** compares models (or their implementations) for consistency, cross-validation, simulator evaluation, or quantification of model developments. [16]

The Python module NetworkUnit [17] is based on the SciUnit [18] and Elephant, and provides a formalized framework as well as a battery of standardized tests for network-level validation.

- Models are matched to appropriate tests via 'capabilities'.
- New or variation of tests can be easily derived from a range of base tests.
- Tests can be used to also compare multiple models.
- Test scores are annotated with their provenance.



Standardization → Reproducibilty

Modularization → Versatility

Formalization → Understandability

References

[1] http://python-elephant.org
[2] Badia, R., et al. "INCF Program on Standards for data sharing: new perspectives on workflows and data management for the analysis of electrophysiological data." Techn. Report, International Neuroinformatics Coordination Facility (INCF). 2015.
[3] Denker, M, et al. (2018) LFP beta amplitude is linked to mesoscopic spatio-temporal phase patterns. Scientific Reports 8.1: 5200. DOI:10.1038/s41598-018-22990-7
[4] Denker M. and Grün S (2016). Designing Workflows for the Reproducible Analysis of Electrophysiological Data. In Brain-Inspired Computing, K. Amunts, L. Grandinetti, T. Lippert, and N. Petkov, eds. (Cham: Springer International Publishing), pp. 58–72
[5] Grewe, J. et al. (2011). A bottom-up approach to data annotation in neurophysiology. Frontiers in Neuroinformatics, 5, p.16.
[6] Zehl, L. et al. (2016). Handling metadata in a neurophysiology laboratory. Frontiers in Neuroinformatics, 10, p.26.
[7] https://github.com/INM-6/python-odmltables
[8] https://www.humanbrainproject.eu/en/explore-the-brain/search
[9] https://web.gin.g-node.org
[10] https://github.com/G-Node/nix/wiki
[11] Garcia S, et al., (2014). Neo: an object model for handling electrophysiology data in multiple formats. Frontiers in Neuroinformatics, 8, 10. DOI:10.3389/fninf.2014.0001
[12] https://www.unicore.eu
[13] https://collab.humanbrainproject.eu/#/collab/5185
[14] Brochier, T, et al. "Massively parallel recordings in macaque motor cortex during an instructed delayed reach-to-grasp task." Scientific data 5 (2018): 180055. DOI:10.1038/sdata.2018.55
[15] https://web.gin.g-node.org/INT/multielectrode_grasp
[16] Gutzen et al. (sub.) "Reproducible neural network simulations: statistical methods for model validation on the level of network activity data"
[17] https://github.com/INM-6/NetworkUnit
[18] https://github.com/scidash/sciunit

The Human Brain Project Summit and Open Day 2018 are hosted by:

Maastricht University   provincie limburg

Co-funded by the European Union

This project/research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2).