



Contribution ID: 21 Contribution code: T-3

Type: Talk

An approach to handle provenance-tracked analysis of NEST simulations using Alpaca

Monday 17 June 2024 10:55 (20 minutes)

NEST simulations are typically executed by a script that configures and runs the simulation. Despite recent improvements in NEST 3.x, where file headers specify the detailed origin of the outputs, users still must interpret the data with respect to the simulation setup. This information is difficult to convey, especially in collaborative contexts with shared simulation results. Moreover, during the explorative process of scientific discovery, results may change without warning when details of the simulation are changed, which could lead to wrong interpretations by collaborators who are unaware of such changes. Therefore, we face two challenges: results are stored in data objects without metadata that describe their role in the simulation, and the simulation outputs are not linked to a description of their provenance with respect to the simulation building.

Here we present concepts to tackle both challenges when using the NEST Python interface. We consider a typical simulation experiment and subsequent data analysis using the Elephant (doi:10.5281/zenodo.1186602; RRID:SCR_003833) toolbox [1]. First, we show how data from a NEST simulation can be represented with data objects annotated with simulation details using the Neo library [2]. Second, we demonstrate how the software Alpaca (doi:10.5281/zenodo.10276510; RRID:SCR_023739) can capture workflow provenance when running a simulation (see Figure) [3]. The two approaches allow the semantic description of the simulation experiment that contributes to the FAIR principles [4] by improving the findability of results through detailed provenance, supporting interoperability through a standardized data model, and promoting reuse of simulation data through enhanced data description.

Acknowledgements

This work was performed as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE) and received funding from the Helmholtz Association of German Research Centres, from the European Union's Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreements No. 785907 (Human Brain Project SGA2) and No. 945539 (Human Brain Project SGA3), European Union's Horizon Europe Programme No. 101147319 (EBRAINS 2.0 Project), by the Joint Lab "Supercomputing and Modeling for the Human Brain", and the NRW network iBehave (NW21-049).

References

- [1] Denker, M., Yegenoglu, A., Grün, S., 2018. Collaborative HPC-enabled workflows on the HBP Collaboratory using the Elephant framework. *Neuroinformatics* 2018, P19. doi:10.12751/incf.ni2018.0019
- [2] Garcia, S., Guarino, D., Jaillet, F., Jennings, T., Pröpper, R., Rautenberg, P.L., Rodgers, C.C., Sobolev, A., Wachtler, T., Yger, P., Davison, A.P., 2014. Neo: an object model for handling electrophysiology data in multiple formats. *Frontiers in Neuroinformatics* 8, 10. <https://doi.org/10.3389/fninf.2014.00010>
- [3] Köhler, C.A., Ulianych, D., Grün, S., Decker, S., Denker, M., 2023. Facilitating the sharing of electrophysiology data analysis results through in-depth provenance capture. <https://doi.org/10.48550/arXiv.2311.09672>
- [4] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Preferred form of presentation

Talk (& optional poster)

Keywords

provenance, workflows, network simulations, electrophysiology data, data analysis, metadata, reproducibility, FAIR principles, Python

Topic area

Interoperability, data and infrastructure

Speaker time zone

UTC+2

I agree to the copyright and license terms

Yes

I agree to the declaration of honor

Yes

Primary authors: KÖHLER, Cristiano (Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Jülich, Germany and RWTH Aachen University, Aachen, Germany); KERN, Moritz (Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Jülich, Germany); GRÜN, Sonja (Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Jülich, Germany and Theoretical Systems Biology, RWTH Aachen University, Aachen, Germany); DENKER, Michael (Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Jülich, Germany)

Presenter: KÖHLER, Cristiano (Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Jülich, Germany and RWTH Aachen University, Aachen, Germany)

Session Classification: Talks