

# Reproducibility assessment as a service: a knowledge graph about the reproducibility of Jupyter notebooks

## Authors:

Daniel Mietchen<sup>1\*</sup> and Sheeba Samuel<sup>2</sup>

\*Lead presenter

<sup>1</sup>[daniel.mietchen@mardi4.de](mailto:daniel.mietchen@mardi4.de), FIZ Karlsruhe — Leibniz Institute for Information Infrastructure

<sup>2</sup>Technical University of Chemnitz

## Abstract:

Reproducibility has many facets, including computational ones. Assessing reproducibility at scale requires consistent workflows. In this context, the Basic Service Jupyter4NFDI offers a unique opportunity for standardization in terms of infrastructure, assessment criteria, documentation and training. In this contribution, we will describe how we used scalable workflows for automatically assessing key aspects of the computational reproducibility of Python-based Jupyter notebooks associated with research publications [1]. While we have focused on publications from the biomedical domain, the approach does not require an association with manuscripts and is transferable to other domains.

Our dataset has detailed metadata about publications, about associated code repositories and about Jupyter notebooks and their reproducibility. We have made it available in its original format as a SQLite database [2] and in addition converted it into a knowledge graph [3]. This knowledge graph can be explored in detail by way of SPARQL queries [4], of which we provide a set of examples catering to multiple use cases. These use cases include the documentation of common problems regarding the reproducibility as well as training contexts in which students or instructors might want to find real-world examples of notebooks from a particular research field and being either fully reproducible or failing in some specific way, e.g. with a given error message relevant for a training session.

Such a conversion of a rich dataset from its original format into a knowledge graph could also be an interesting use case for the Basic Service KGI4NFDI. We would be particularly interested in exploring how these two Base Services could interact to facilitate (1) routine reproducibility assessments of Jupyter notebooks relevant in NFDI contexts, (2) feeding such a knowledge graph with information about these reproducibility assessments, (3) documentation of common issues and best practices and (4) training for Jupyter, reproducibility, knowledge graphs and other facets of Base Services, like terminology or persistent identifiers.

## References:

- [1] Sheeba Samuel, Daniel Mietchen, Computational reproducibility of Jupyter notebooks from biomedical publications, GigaScience, Volume 13, 2024, giad113, <https://doi.org/10.1093/gigascience/giad113>
- [2] Sheeba Samuel, & Daniel Mietchen. (2023). Dataset of a Study of Computational reproducibility of Jupyter notebooks from biomedical publications [Data set]. In GigaScience (Vol. 13). Zenodo. <https://doi.org/10.5281/zenodo.8226725>
- [3] Sheeba Samuel, Daniel Mietchen, FAIR Jupyter: a knowledge graph approach to semantic sharing and granular exploration of a computational notebook reproducibility dataset (2024), arXiv preprint arXiv:2404.12935, <https://doi.org/10.48550/arXiv.2404.12935>
- [4] <https://w3id.org/fairjupyter>

**Keywords:** Jupyter notebooks, Python, computational reproducibility, knowledge graphs