

Jupyter4NFDI: a central JupyterHub providing access to various software stacks and computing resources across the NFDI consortia

Authors:

Tim Kreuzer^{1*}, Björn Hagemeyer¹, Bernd Flemisch², Arnim Bleier³, Matthias Lieber⁴, George Dogaru⁵ and Klaus Reuter⁶

*Lead presenter

¹t.kreuzer@fz-juelich.de, Forschungszentrum Jülich GmbH

²University of Stuttgart

³GESIS – Leibniz-Institute for the Social Sciences

⁴Technische Universität Dresden

⁵Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG)

⁶Max Planck Computing & Data Facility (MPCDF)

Abstract:

The widespread usage of Jupyter notebooks across scientific disciplines highlights their fundamental importance in current research practices. However, their deployment across various NFDI consortia currently occurs through individual JupyterHubs, resulting in access barriers to computational and data resources. Only some of the services are widely available, and others are barricaded within VPNs, creating a fragmented landscape that complicates access. Our base service Jupyter4NFDI aims to unify these efforts by offering a centralized service. We simplify access, significantly improve the user experience, and extend the reach of Jupyter to a broader audience within the NFDI and beyond. We integrate with IAM4NFDI to govern access to the central service and external resources, with future integrations planned with upcoming basic services such as Multi-Cloud or nfdi.software. Moreover, HPC resources from the GCS and the NHR Alliance will be accessible via the centralized Jupyter service, which lowers HPC entry barriers for a wide range of users. Importantly, the centralized system not only simplifies access but also supports the import of projects along with their necessary dependencies, fostering an ecosystem conducive to creating binder-like 1-click reproducible FAIR Digital Objects (FDOs), possibly along with notebook identifiers supported by PID4NFDI. By ensuring a consistent user experience across all consortia, we promote continued innovation. Our commitment is to a collaborative and centralized approach that brings Jupyter notebooks within everyone's reach and democratizes access to the infrastructures for reproducible computational research within the NFDI and beyond.

The project's initialization phase has started in June 2024. During this phase, we build up a central base service, initially hosted as part of the already existing JupyterHub environment at JSC, which has been providing Jupyter based access to high-performance computing resources for over seven years by now and is continuously improved and extended.

We will present an update on the current status of the project and the intended next steps, including an outlook for the next project phases.

Keywords: Jupyter, Interactive Computing, HPC, NHR