

Infrastructure for Research Based Learning and the NFDI Basic Service Jupyter: Searching for Opportunities for Cooperation and Areas of Differentiation

Authors:

Jan F. Kraemer^{1*} and Lilian Löwenau²

*Lead presenter

¹jan.kraemer@hu-berlin.de, Humboldt-Universität zu Berlin, Germany

²Humboldt-Universität zu Berlin, Germany

Abstract:

The appeal of using hosted infrastructure is seen not only in research practice but also in many educational settings. Minimizing individual configuration effort allows focusing on the task at hand while also improving the reproducibility of results. In the educational context, centrally provisioned environments enable teaching staff to integrate the use of advanced computational methods in their curriculum, while focusing on the subject domain rather than the implementation details. Reproducible research practices are equally much easier to implement if existing environments contain the required computational tools and interfaces to data infrastructure. In both contexts, this presents an inherent tradeoff between ease-of-use and the immediacy of usability, and the concomitant explosion of the number of variations that would ideally be offered, but cannot realistically be centrally supported and maintained.

In the specific context of Jupyter based computational notebooks, a specific view heavily influenced the default tools to configure and deploy JupyterHub: a concentration on smaller numbers of similar users, i.e. individual research groups and specific classes. While this allows a local deployment under most circumstances, it also leads to a highly distributed set of “administrative users”, with a high opportunity cost regarding coordination and a multitude of “works for me” solutions to (probably) shared problems.

In the context of larger research projects with established structures, including those with regards to funding as is the case with the NFDI Basic Service Jupyter, this can be approached in a structured and organized manner. In education, requirements for the integration into a set of specific and highly idiosyncratic services for course management as well as varying computational infrastructures at universities, provide an additional layer of complications to coordination.

Inspired by our experience while planning the centrally provided jupyterhub for teaching support at Humboldt-Universität zu Berlin, we seek ways to improve this coordination bottleneck. The announcement of the Jupyter Basic Service initiative of the NFDI in April of 2024 may provide one avenue towards this goal, if sufficient overlap between the respective requirements can be found. Integrating NFDI basic services and local educational infrastructure initiatives could allow the inclusion of research based learning into earlier stages of a students educational journey and thus help to prepare future researchers for the emerging best practices in reproducible research and FAIR data practices.

AT UC4B2024 we aim to present our experiences with and requirements for a centrally managed Jupyter Environment, as well as the results of discussions at (potentially, currently in review) two workshops on the topic: one at the annual meeting of DINI (Deutsche Initiative für Netzwerkinformation) the other at the TURN Conference. We hope to initiate a constructive dialog with the NFDI community about shared needs and possible avenues for joint efforts of improvement of the jupyter stack.

Keywords: Higher Education, Research Based Learning, Infrastructure for Education in Computational (Research-)Methods, Jupyter, JupyterHub