

Challenges, Opportunities and Experiences of setting up sustainable Research Data Infrastructure

Authors:

Andreas Ganzenmüller^{1*}, Preston Rodrigues², Volodymyr Kushnarenko², Miroslav Puskaric², and Thomas Bönisch²

*Lead presenter

¹andreas.ganzenmueller@hirs.de, High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

²High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart, Germany

Abstract:

Setting up a sustainable research data infrastructure is not just about hardware and software. It is also about expertise, routines and workflows on how to handle research data on a daily basis, as well as the necessity to have the required human resources to maintain the infrastructure in the long term.

As an infrastructure provider HLRS is collaborating in several projects [1, 2, 3, 4] concerning research data management. In all these projects there is a requirement that some services (primarily research data repositories) are set up and perspective kept running for many years. To accomplish this, we at HLRS decided to develop a platform dedicated primarily to the needs of research data management. This platform is supposed to accommodate all the required services. Currently our platform comprises a small-scale virtualisation cluster running Proxmox Virtual Environment, an open-source server virtualisation management solution [5], and a dedicated storage system.

Since we began developing the platform, we have gained important insights which we would like to share with the community. Starting with hardware selection, integration into existing infrastructure, network configuration, hosting web-services and ensuring sustainable and secure maintenance. We would like to provide insight into our architecture. We will elaborate motivation behind our design decisions and demonstrate how research data management software can be deployed on the platform. Additionally, we would like to emphasize the critical role of detailed technical documentation and the dedicated team responsible for its operation, particularly in ensuring the platform's long-term operation.

Overall, we identified three issues to be major design challenges: flexibility, scalability and sustainability.

Flexibility is a key issue, because the platform must be able to run a wide variety of software and integrate different services to work together. To accomplish this, some small adaptations of software components are inevitable. But any large software customisation needs to be met with caution, because maintenance effort can increase drastically. The goal should be to generate synergy effects by having multiple projects being administrated by the same service provider and benefit from the accumulated experience and shared resources.

Scalability is critical, because the platform must be expandable to accommodate further services in the future. At the same time, it is necessary to aim for an efficient hardware utilisation to lower over-all costs for all participants. Extending the hardware base of the

platform therefore is a continuous task that requires careful planning in advance.

Concerning sustainability, the overall challenge lies in setting up organisational structures that allow to maintain the platform independent of the financial contributions of single service partners. Therefore, the platform requires a certain minimum size to serve enough projects and accommodate their services, to ensure operations in case one service partner unexpectedly drops out.

Our platform can serve as an example for sustainable research data infrastructure and our designs and workflows can be used as guidelines to develop, configure and maintain similar platforms and related services for big communities such as NFDI.

References:

- [1] SDC4Lit Homepage. URL: <https://www.sdc4lit.de/>, last access on 15.05.2024.
- [2] NFDI4Cat Homepage. URL: <https://nfdi4cat.org/>, last access on 15.05.2024.
- [3] KI Allianz Baden Wuerttemberg Homepage. URL: <https://ki-allianz.de/>, last access on 15.05.2024.
- [4] Meridional Homepage. URL: <https://meridional.eu/>, last access on 15.05.2024.
- [5] Proxmox Homepage. URL: <https://www.proxmox.com/en/>, last accessed on 15.05.2024.

Keywords: Infrastructure, Research Data, Data Management, Sustainability