

Embedding the de.NBI Cloud in the National Research Data Infrastructure Activities

Nils Hoffmann¹[\[https://orcid.org/0000-0002-6540-6875\]](https://orcid.org/0000-0002-6540-6875), Irena Maus¹[\[https://orcid.org/0000-0003-3335-9514\]](https://orcid.org/0000-0003-3335-9514), Sebastian Beier²[\[https://orcid.org/0000-0002-2177-8781\]](https://orcid.org/0000-0002-2177-8781), Peter Belmann¹[\[https://orcid.org/0000-0002-1294-2869\]](https://orcid.org/0000-0002-1294-2869), Jan Krüger¹, Andreas Tauch¹, The de.NBI Cloud Consortium³, Alexander Sczyrba⁴[\[https://orcid.org/0000-0002-4405-3847\]](https://orcid.org/0000-0002-4405-3847)

¹ Institute of Bio- and Geosciences, Computational Metagenomics (IBG-5), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

² Institute of Bio- and Geosciences, Bioinformatics (IBG-4), Forschungszentrum Jülich GmbH, Jülich, 52425, Germany

³ The de.NBI Cloud Consortium in Germany: Alexander Goesmann¹[\[https://orcid.org/0000-0002-7086-2568\]](https://orcid.org/0000-0002-7086-2568), Justus-Liebig-University Giessen; Roland Eils¹[\[https://orcid.org/0000-0002-0034-4036\]](https://orcid.org/0000-0002-0034-4036), BIH-Zentrum Digitale Gesundheit, Charité - Universitätsmedizin Berlin; Peer Bork¹[\[https://orcid.org/0000-0002-2627-833X\]](https://orcid.org/0000-0002-2627-833X), European Molecular Biology Laboratory Heidelberg; Oliver Kohlbacher¹[\[https://orcid.org/0000-0003-1739-4598\]](https://orcid.org/0000-0003-1739-4598), Eberhard Karls University Tübingen; Ursula Kummer¹[\[https://orcid.org/0000-0002-6413-2382\]](https://orcid.org/0000-0002-6413-2382), BioQuant & Heidelberg University; Rolf Backofen¹[\[https://orcid.org/0000-0001-8231-3323\]](https://orcid.org/0000-0001-8231-3323), Albert-Ludwigs University Freiburg; Ivo Buchhalter¹[\[https://orcid.org/0000-0003-0764-5832\]](https://orcid.org/0000-0003-0764-5832), Deutsches Krebsforschungszentrum Heidelberg; Alexander Sczyrba, Bielefeld University

⁴ Faculty of Technology, Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany

Keywords: Cloud Computing, Federated Research Infrastructure, AAI, Virtual Research Environments, Training

Abstract

In recent years, modern life sciences research underwent a rapid development driven mainly by the technical improvements in analytical areas leading to miniaturization, parallelization, and high throughput processing of biological samples. This has led to the generation of huge amounts of experimental data. To meet these rising demands, the German Network for Bioinformatics Infrastructure (de.NBI)[1] was established in 2015 as a national bioinformatics consortium aiming to provide high quality bioinformatics services, comprehensive training[2], powerful computing capacities (de.NBI Cloud) as well as connections to the European Life Science Infrastructure ELIXIR[3], with the goal to assist researchers in exploring and exploiting data more effectively.

The establishment of the de.NBI Cloud has proven to be a flagship for the de.NBI network (see Fig. 1 for details). It consists of eight federated cloud locations that implement a common governance and use the project application and management workflow provided by the de.NBI Cloud portal. This governance facilitates secure operations of each cloud location through the centralized organization of ISO27001 Information Security Management System (ISMS) training and certification courses for the cloud staff, leading to the progressive certification of our cloud sites. Registration, project resource application and authentication are facilitated by the integration of the LifeScience AAI[4], [5] as an EduGAIN-compatible single sign-on provider, backed by institutional ID providers of universities and research institutes.

Since its foundation, de.NBI Cloud has formed the scientific and collaborative backbone for new major German initiatives like NFDI[6] or EOSC-Life in the European sector of compu-

tational biosciences. Above all, the cooperation with various NFDI consortia such as NFDI4Biodiversity, DataPLANT, GHGA, FAIRagro or NFDI4Microbiota showcases the power, range and flexibility of the de.NBI Cloud, especially for the national life science community.

The de.NBI Cloud portfolio includes several project types designed to suit different use cases and users with varying levels of knowledge in cloud computing. Two project types, OpenStack and Kubernetes, offer maximum flexibility in terms of the configuration of cloud-specific components and allow the installation of any large-scale analysis, stream processing or orchestration framework available in the cloud ecosystem. Both project types are ideal for science gateway developers to offer bioinformatics services to the national and international life sciences communities.

To be more precise, OpenStack makes controlling large pools of computing, storage, and network resources simple. Any interaction with OpenStack can be performed through its dashboard or automated through its API, using well-known cloud infrastructure management frameworks like Terraform or Ansible. While OpenStack enables the provisioning of virtual machines, networks and storage, Kubernetes (K8s) is the state-of-the-art technology for the deployment, scaling, and orchestration of highly available containerized applications. The de.NBI Cloud supports "vanilla" Kubernetes clusters on top of OpenStack using Kubermatic. Features like self-healing, automated rollback, and horizontal scaling make K8s the ideal basis for services. Confidential processing of sensitive data, e.g., pseudonymized patient-related data, is also possible at specific de.NBI Cloud locations, where data security is enforced through separated, secure processing environments. From a legal perspective, this is mediated through data processing agreements between a project's principal investigator and the hosting cloud site.

Our in-house developed project type SimpleVM enables our users to use cloud resources with little to no background knowledge in cloud computing. SimpleVM is an abstraction layer on top of OpenStack to manage single virtual machines (VMs) or clusters thereof. It was designed to support the combination of resources from independent OpenStack installations, thus operating as a multi-cloud platform which is accessible from a single web-based control panel. The entire software stack only requires access to the OpenStack API and can be deployed on any vanilla OpenStack installation. In general, SimpleVM primarily eases the creation and management of individual pre-configured virtual machines and provides web-based access to popular research and development environments such as Rstudio, Guacamole Remote Desktop, Theia IDE, JupyterLab and Visual Studio Code. On top of this functionality, a dedicated mode that simplifies the setup of cloud-based workshops for teaching purposes is also provided. Further, with SimpleVM, de.NBI Cloud users can effortlessly configure and manage their own SLURM-based BiBiGrid clusters with just a few clicks. This feature addresses the needs of researchers who want to run their tools or entire workflows across multiple machines.

For users who want to be able to define data processing workflows from tools available in BioConda and the Galaxy ToolShed with a graphical user interface, the de.NBI cloud infrastructure also hosts the Galaxy[7] service available at usegalaxy.eu. Galaxy also simplifies the discovery and adaptation of existing workflows, that were shared by other users, from multiple scientific domains and enables their execution at scale in the cloud.

In conclusion, the de.NBI Cloud provides the ability to unlock the full potential of research data and enables easier collaboration across different ecosystems and research areas, which in turn enables scientists to innovate and scale-up their data-driven research, not only in the life and computational biosciences, but across the different science domains addressed by the NFDI.

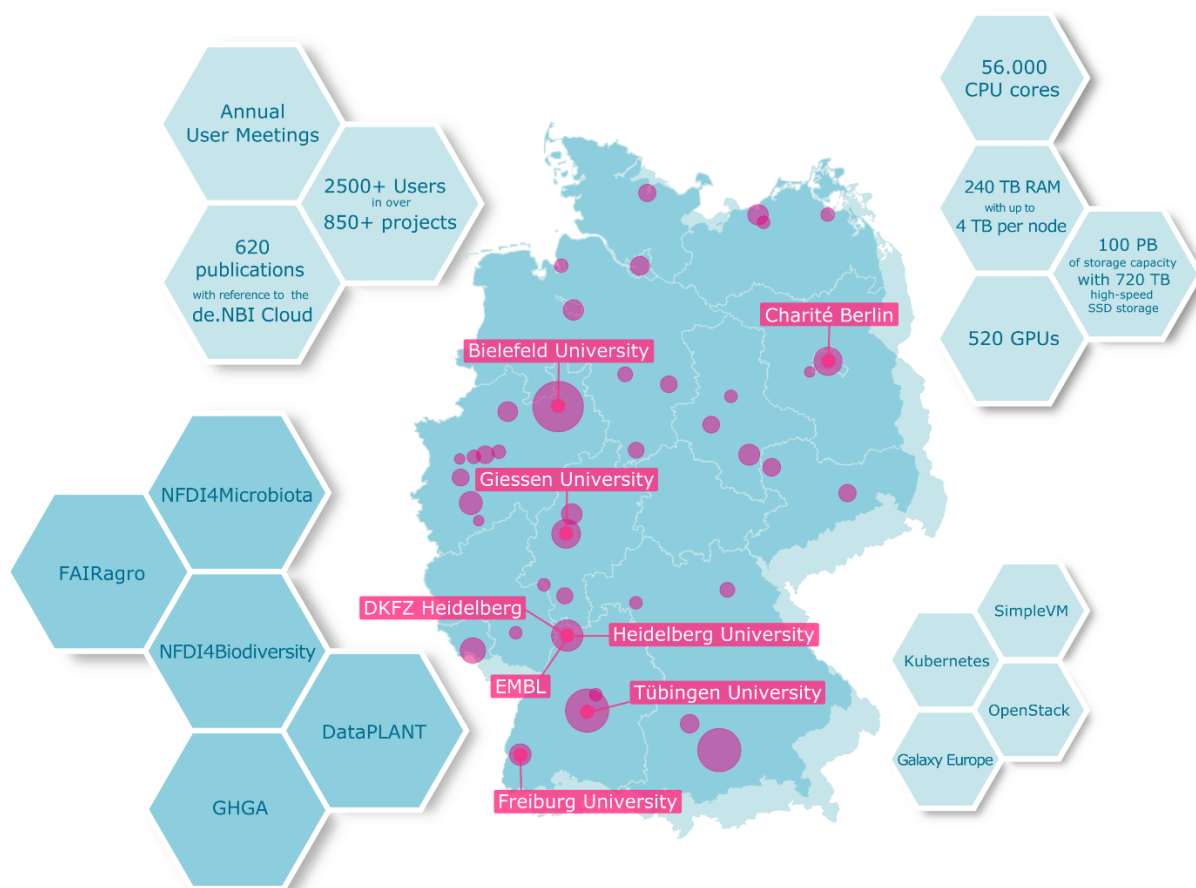


Figure 1. The de.NBI Cloud federation, maintained by the eight German cloud centers in Berlin, Bielefeld, Freiburg, Gießen, Heidelberg and Tübingen. Red circles show the distribution and approximate number of de.NBI Cloud AAI users. The upper left hexagons show data on use and visibility of the cloud, while the upper right hexagons show the total number of provided computing resources. NFDI projects that use cloud resources are listed at the bottom left. The de.NBI Cloud project types, tailored to different use cases and requirements are shown at the bottom right.

Data availability statement

This submission is not based on data.

Author contributions

NH, IM, SB, PB, JK, AT and AS drafted and wrote the original draft, all authors reviewed, edited and approved the draft manuscript. AG, RE, PB, OK, UK, RB, IB and AS acquired the funding for the de.NBI cloud and are, together with AT, responsible for project administration and supervision.

Competing interests

The authors declare that they have no competing interests.

Funding

We acknowledge funding by the German Ministry for Education and Research (BMBF) under the grant numbers (FKZ) 031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, and 031A537D.

Acknowledgement

Operating a federated infrastructure like the de.NBI Cloud is not possible with a small group of people alone. The authors would therefore like to thank all cloud site personnel and people involved with development and operations, past and present, for their dedication and support to provide an excellent service to the life sciences community in Germany.

References

- [1] A. Pühler, "Bioinformatics solutions for big data analysis in life sciences presented by the German network for bioinformatics infrastructure," *J Biotechnol*, vol. 261, p. 1, Nov. 2017, doi: 10.1016/j.jbiotec.2017.08.025.
- [2] D. Wibberg *et al.*, "The de.NBI / ELIXIR-DE training platform - Bioinformatics training in Germany and across Europe within ELIXIR," *F1000 Research*, vol. 8, p. 1877, Nov. 2019, doi: 10.12688/f1000research.20244.1.
- [3] A. Tauch and A. Al-Dilaimi, "Bioinformatics in Germany: toward a national-level infrastructure," *Brief Bioinform*, vol. 20, no. 2, pp. 370–374, Mar. 2019, doi: 10.1093/bib/bbx040.
- [4] M. Linden *et al.*, "Common ELIXIR Service for Researcher Authentication and Authorisation," *F1000Res*, vol. 7, p. ELIXIR-1199, 2018, doi: 10.12688/f1000research.15161.1.
- [5] P. Belmann *et al.*, "de.NBI Cloud federation through ELIXIR AAI," *F1000Res*, vol. 8, p. 842, 2019, doi: 10.12688/f1000research.19013.1.
- [6] S. Kraft *et al.*, "Nationale Forschungsdateninfrastruktur (NFDI) e. V.: Aufbau und Ziele," *BFDM*, no. 2, pp. 1–9, Jul. 2021, doi: 10.17192/bfdm.2021.2.8332.
- [7] The Galaxy Community, "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update," *Nucleic Acids Research*, vol. 50, no. W1, pp. W345–W351, Jul. 2022, doi: 10.1093/nar/gkac247.