ELIXIR-DE / de.NBI Cloud as a National Data Infrastructure for Health Data

Alexander Sczyrba

Forschungszentrum Jülich

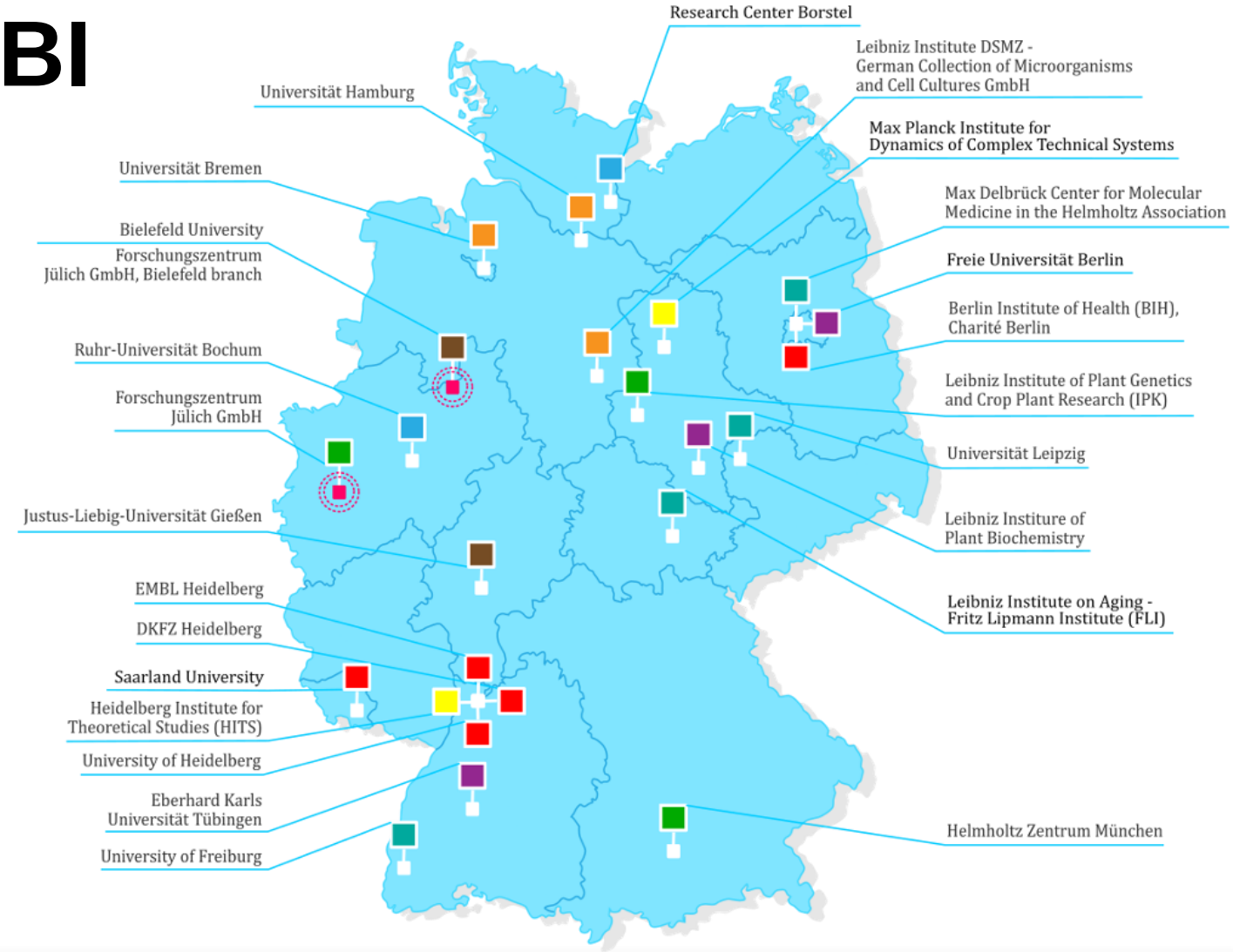
de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

elixir GERMANY

German Network for Bioinformatics Infrastructure – de.NBI

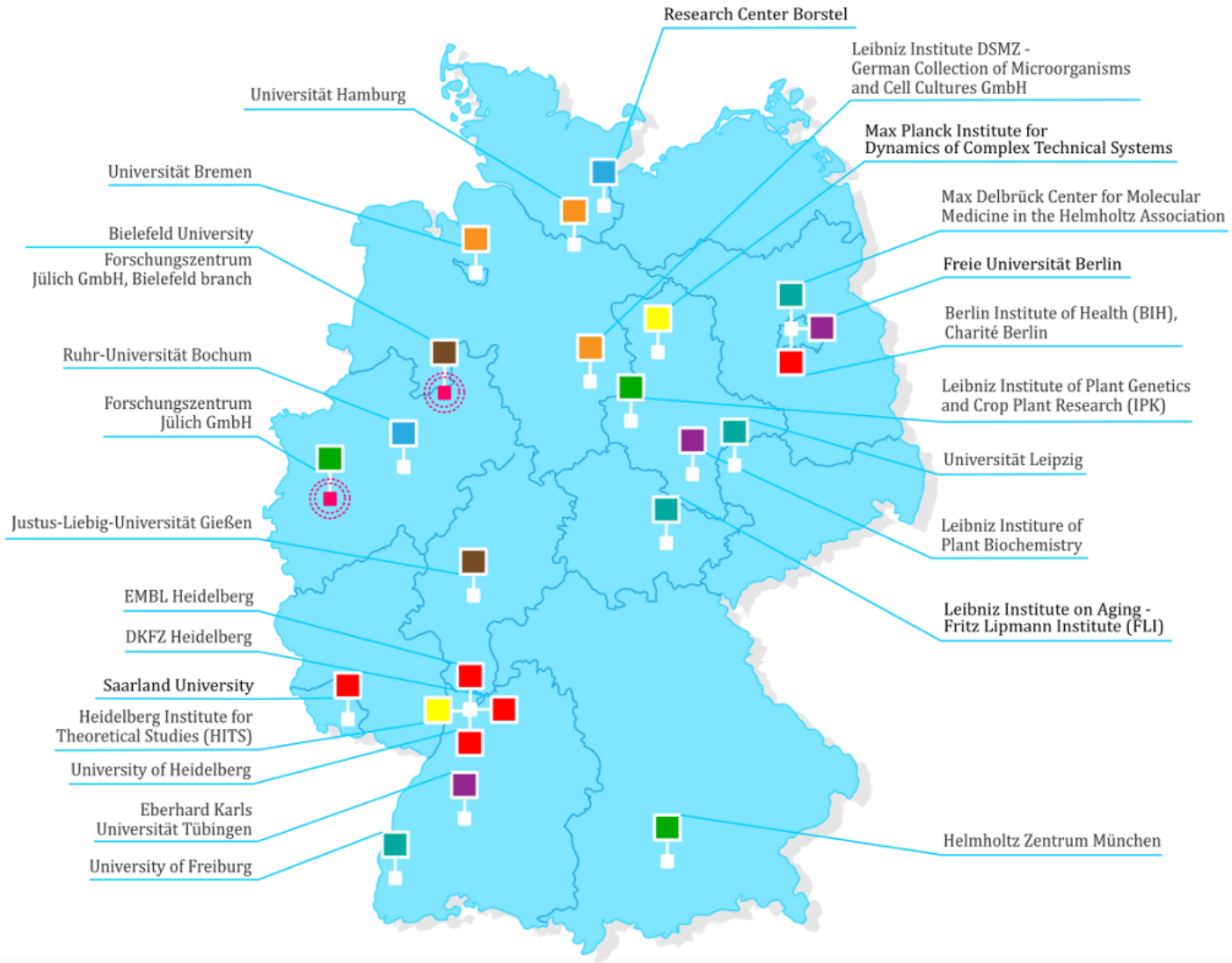
- **24 partners**
- **8 service centers**
- funded by BMFTR, federal states and partner institutions
- scientific expertise in a wide range of bioinformatics topics:

  - Human Bioinformatics
  - Microbial Bioinformatics
  - Plant Bioinformatics
  - RNA Bioinformatics
  - Bioinformatics for Proteomics
  - Integrative Bioinformatics
  - Reference Databases
  - Data Management and Systems Biology



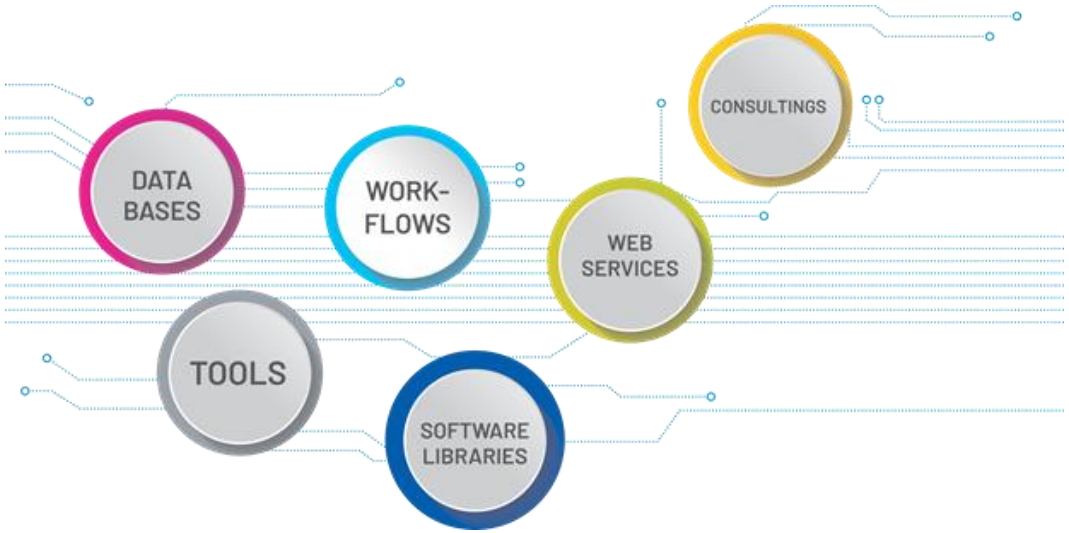
de.NBI Mission

- Provision of comprehensive first-class **bioinformatics tools & services** to users in basic and applied life sciences research
- Bioinformatics **training** in Germany and Europe through a wide range of workshops and courses
- Providing **cloud computing** resources for academia in Germany
- Transfer of expertise between academia and industry in our **Industrial Forum**
- **Cooperation** of the German bioinformatics community with **international bioinformatics networks**



de.NBI Services for Life Scientists

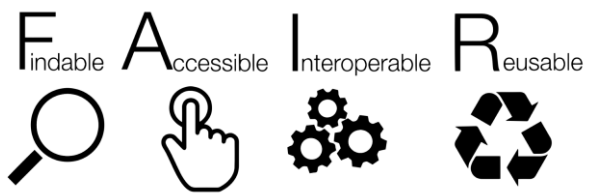
de.NBI offers **> 160 bioinformatics tools and services**
for the analysis and evaluation of biological data sets



The majority referenced, searchable
and findable in the ELIXIR **bio.tools** registry

https://bio.tools

de.NBI Services are committed to FAIR data principles



https://www.denbi.de/services

de.NBI Training for Life Scientists

- The de.NBI network provides a **state-of-the-art training program** addressing a multitude of topics
- Life scientists learn how to **handle and analyze biological big data** more effectively by applying tools, standards, and compute services
- de.NBI has evolved to be a **leading contact point for bioinformatics training** in Germany and Europe and guarantees high quality standards

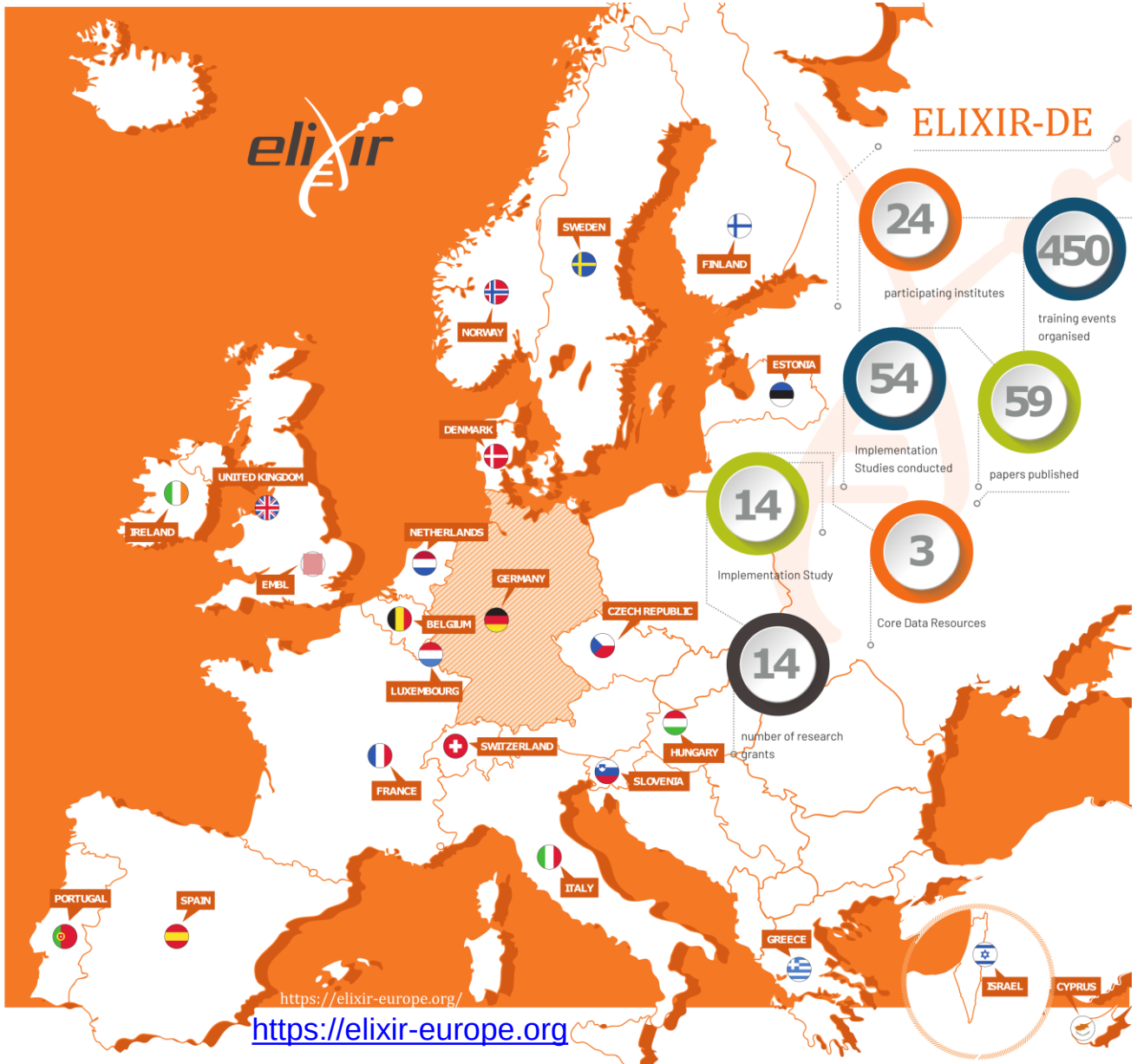
Ø 40-80 de.NBI courses per year

Upcoming events:

- Summer School on Cheminformatics
- Cloud meets RDM
- Machine learning for genomics
- Lipidomics and Bioinformatics
- Computational Pangenomics
- Galaxy training programmes
- Developer meetings
- BioHackathons

https://www.denbi.de/training

The German Node within ELIXIR Europe



https://elixir-europe.org

ELIXIR-DE

- **ELIXIR** is an intergovernmental organization and brings together life science resources from across Europe
- The ELIXIR consortium consists of **21** member states + EMBL
- **ELIXIR Germany** is the national Node of the European ELIXIR infrastructure since 2016
- The node is run by members of the German Network for Bioinformatics Infrastructure (**de.NBI**)

24
Participating German institutes

139
Tools & Services

26
Training Courses in 2025 so far

de.NBI Cloud – Cloud Computing for Life Sciences

Federated Sites - Shared Responsibility

Independent Legal Entities operate each data center

- **Cooperation based on de.NBI / ELIXIR-DE cooperation agreement**
- **Shared technical and legal requirements for operation of a de.NBI Cloud location**
- **Common (cloud portal) and local (individual cloud sites) responsibility for IT / service security and data protection**
- **Data processing agreements between users and cloud site for sensitive data processing**

Centrally organized by de.NBI Cloud Portal at Forschungszentrum Jülich, Outstation Bielefeld

- **Project resource applications and modifications**
- **Project assignment to cloud sites based on requirements**
- **Central user helpdesk - 1st and 2nd level support**

- **Quarterly strategic meeting of de.NBI Cloud Working Group**
- **Monthly technical and certification coordination meetings**

- **Multiple ISMS and IT security trainings annually**
- **Mutual internal audits of ISMS state and adoption at each cloud site**
- **Primary and secondary means of communication and disaster response**
- **Directory of procedures for each participating cloud site + portal**

de.NBI GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE
elixir GERMANY

A Core de.NBI Service: The Federated de.NBI Cloud

LS LOGIN ⟷ eduGAIN

- LinkedIn
- Apple
- Google
- ORCID
- GitHub
- LifeScience Hostel

LIFE SCIENCE RI Perun

Registration and Login

- LifeScience Login account and membership in the de.NBI Virtual organisation is needed
- account usable for authentication in the Cloud

Submit an Application

- select a project type and resources of your need
- submit the application naming a responsible PI for the project

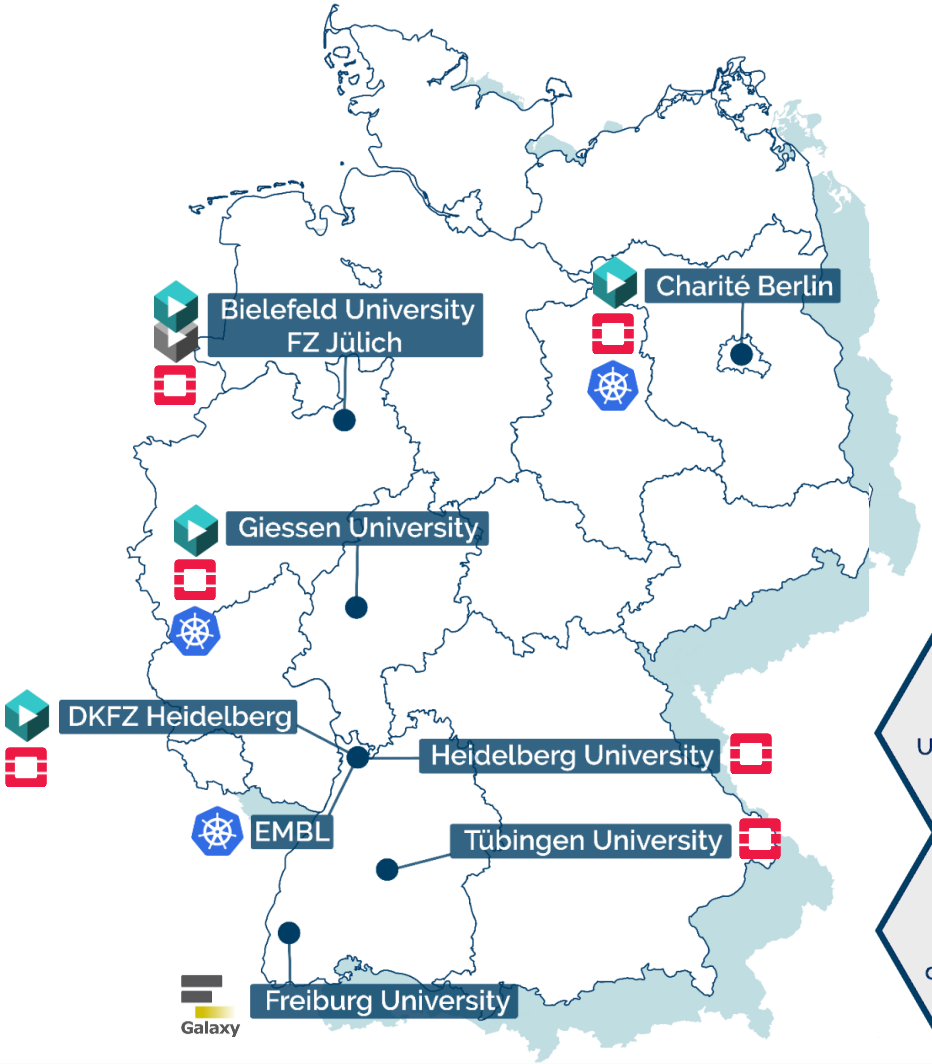
Review Process

- review by the Cloud Access Committee
- frequent updates on the review process
- resources get allocated

Use the Resources

- make use of the resources
- manage access to the project
- modify resources if needed

openstack. · kubernetes · SimpleVM



- 56.000 CPU cores
- 720TB RAM with up to 4TB per node
- 100 PB of storage capacity with 720TB high-speed SSD storage
- 520 GPUs

SimpleVM · Kubernetes · OpenStack · Galaxy Europe

- Annual User Meetings
- 3500+ Users registered in 1200+ projects
- 1000+ publications with reference to the de.NBI Cloud

de.NBI GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE · elixir GERMANY

Project Type Selection

Infrastructure-, Platform- and Workflows-as-a-Service

- High configurability, infrastructure virtualization
- API access, e.g. for use with Terraform or Ansible
- Any software of the cloud ecosystem

- High configurability, service and container orchestration
- API access, e.g. for use with kubectl
- Any containerized software / service, Helm charts

- Beginner-friendly, preconfigured Research Environments
- ELIXIR guarded interactive browser sessions
- “One-click” solution for setting up your own cluster

- Fully established bioinformatics tools and workflows, maintained by the community
- Point-and-click GUI for composition of bioinformatics workflows
- Interactive tours and comprehensive training library

SimpleVM - Simple Access to Virtual Machines

Self Service VMs and Research Environments

SimpleVM

- Beginner-friendly, preconfigured & custom research environments
- Integrated with LifeScience AAI for authentication and authorization
- Easy VM, volume and user management
- Browser and SSH access

IDEs

THEIA

Visual Studio Code

Data Science Notebooks

RStudio

jupyter

Remote Desktop

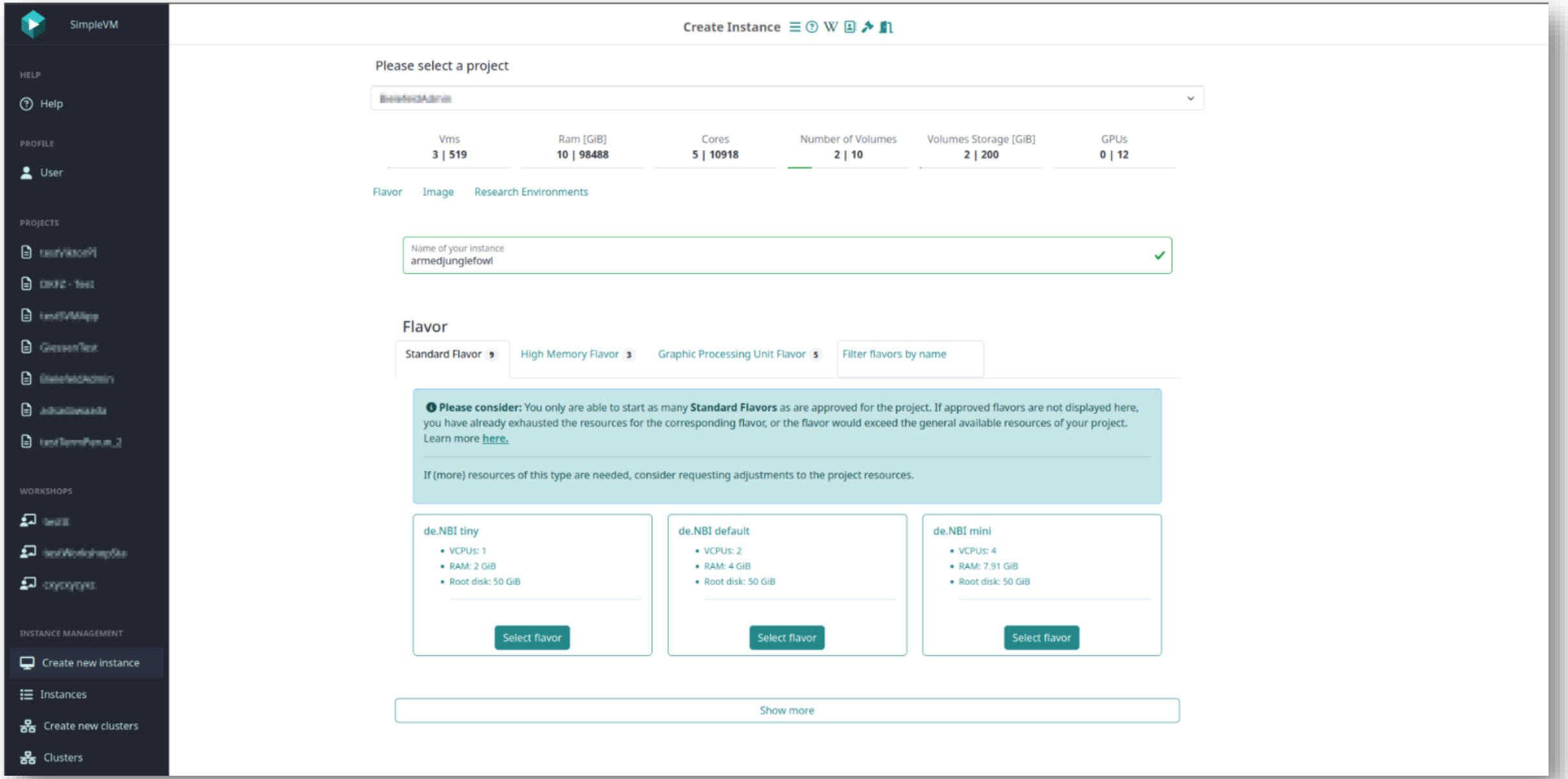
Canonical Ubuntu + XFCE

Custom VREs

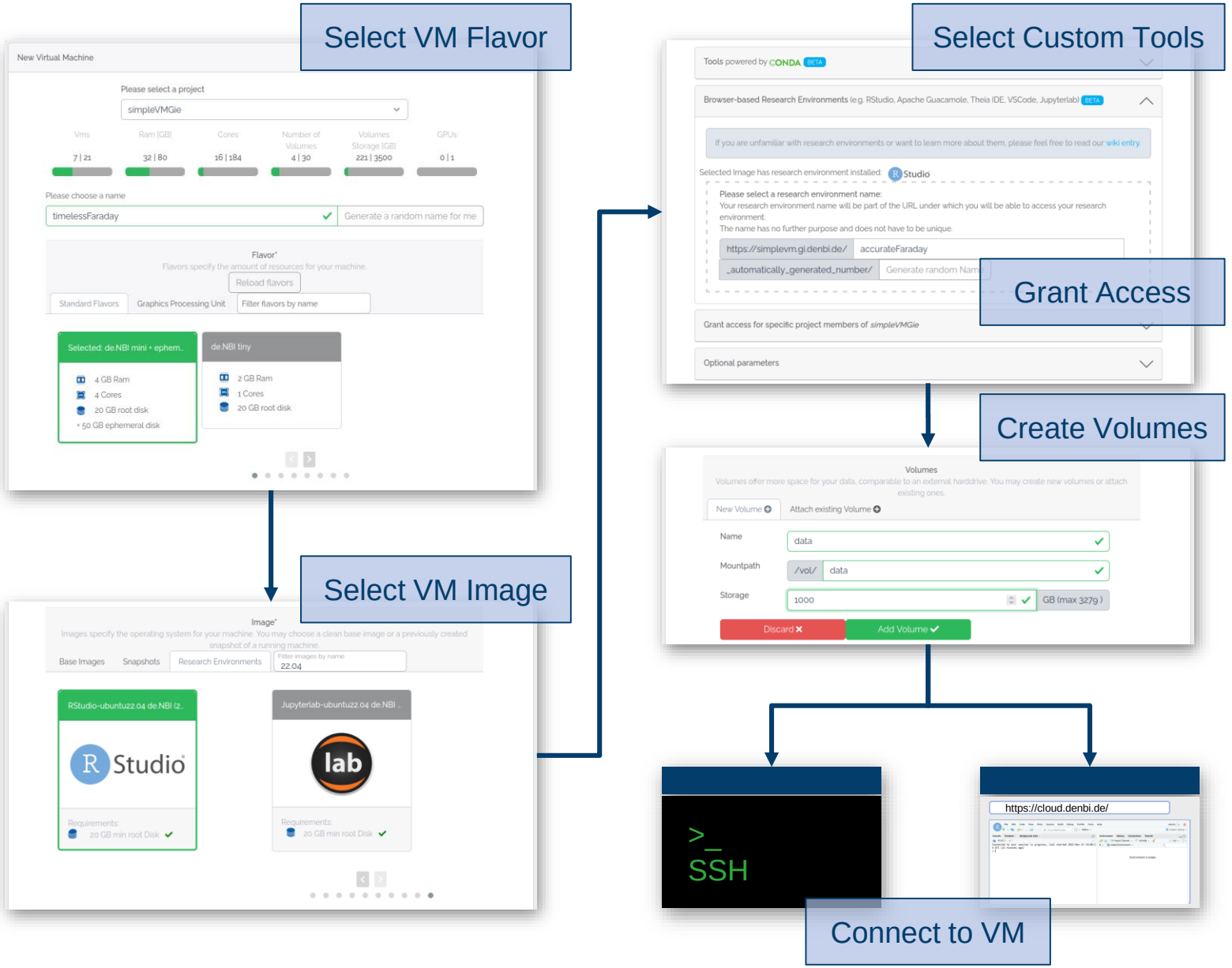
Community Curated

HashiCorp Packer

SimpleVM Portal - Create Instance - Flavor Selection



SimpleVM - Simplified Access to Virtual Machines



Steps to create a SimpleVM

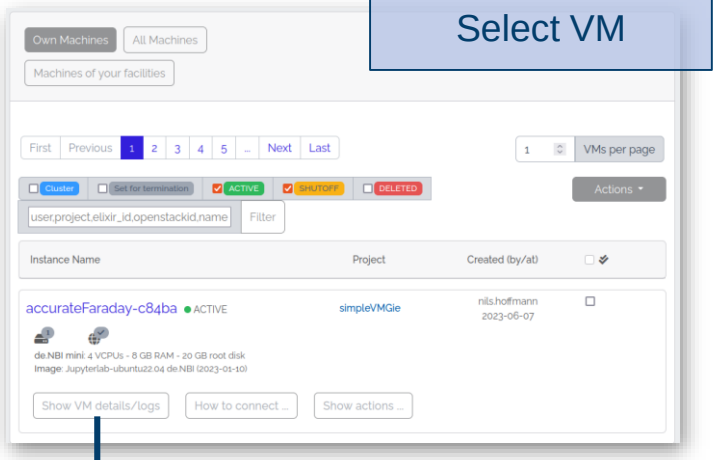
1. **Select VM Flavor**
   - # of vCPUs
   - RAM
   - base storage
   - GPU
2. **Select VM Image**
   - Base image (Ubuntu, …)
   - Research Environments:
     - RStudio
     - Visual Studio Code
     - Jupyter Lab
     - Theia
     - Remote Desktop with Guacamole
   - Snapshot
3. **Select Custom Tools**
   - Conda
   - Research Environments
4. **Grant access to VM**
5. **Create or attach volumes**
6. **Start VM**
7. **Connect via SSH or browser**

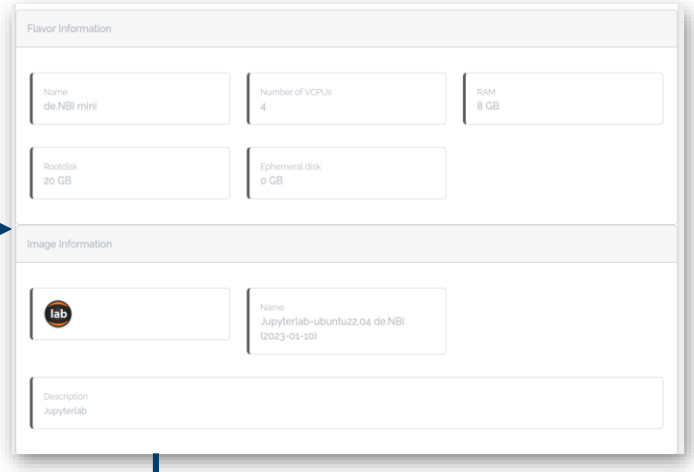
SimpleVM - VM Management

Managing a SimpleVM

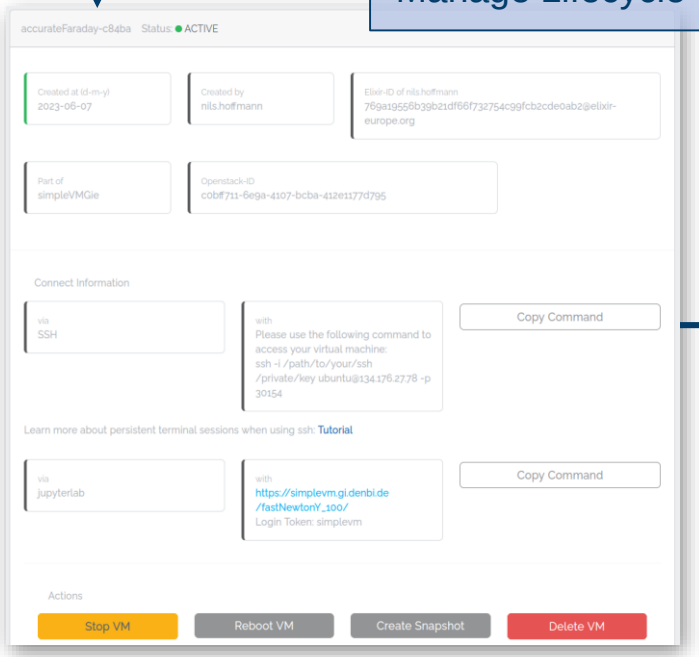
1. Select VM
2. Manage VM Lifecycle
   - Stop
   - Start
   - Reboot
   - Snapshot
   - Delete
3. Manage Volumes
   - Detach
   - Attach
4. Manage Access to VM

Select VM

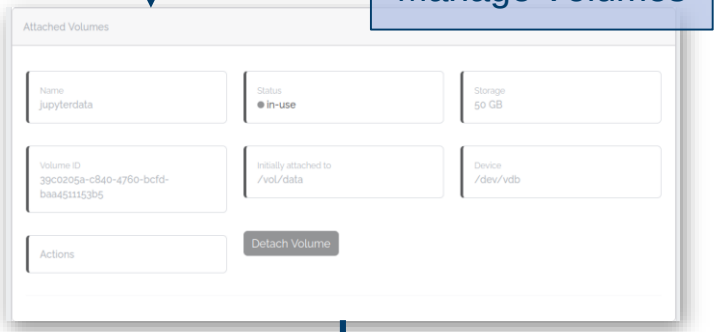




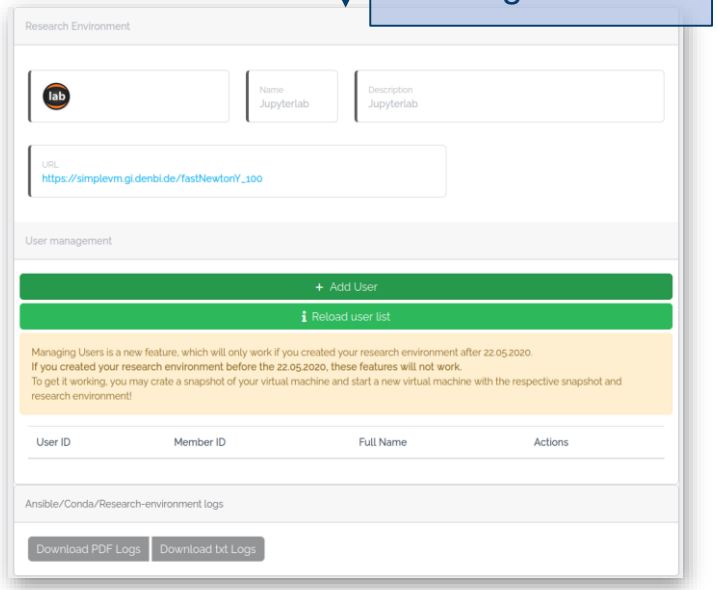
Manage Lifecycle



Manage Volumes



Manage Access



SimpleVM Features: Workshop & Cluster Module

SimpleVM

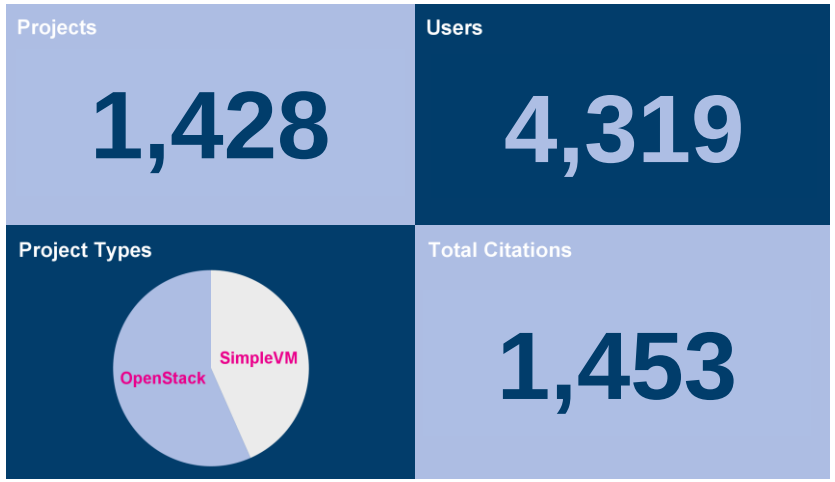
Workshop

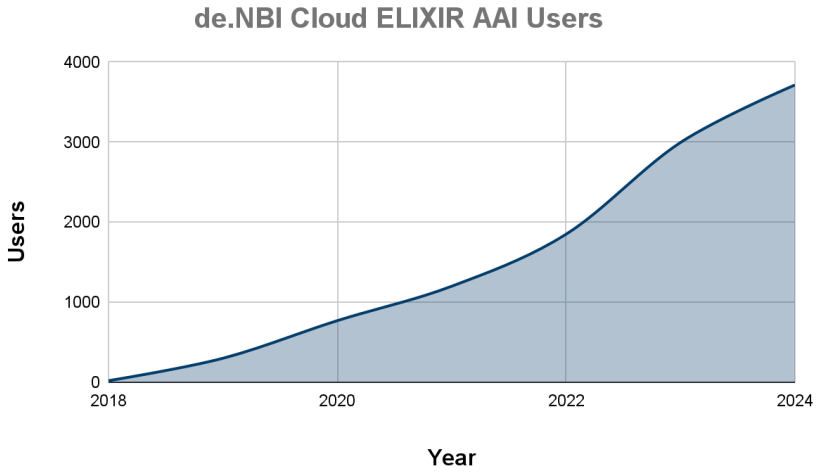
- Workshop module to define VMs using base images, virtual research environments, or snapshots
- Add participants with an existing Life Science AAI account to your project
- Send VM access details to each participant individually
- Launch VM instances for your participants

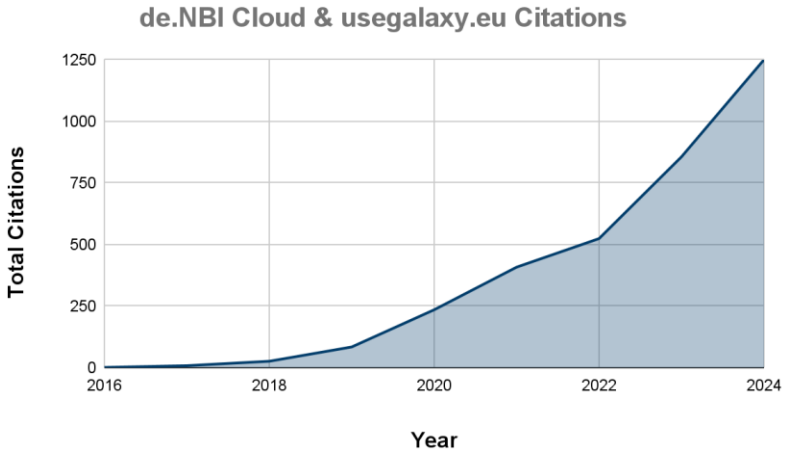
Cluster

- “One-click” solution for starting more complex cluster setups, e.g. for Nextflow
- Connect multiple machines to one SimpleVM Cluster
- Job scheduling mechanism (SLURM)
- Easy automatic scale-up and scale-down
- Currently in beta development, available on request

de.NBI Cloud in Numbers







This work was supported by the de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) and ELIXIR-DE (Forschungszentrum Jülich and W-de.NBI-001, W-de.NBI-004, W-de.NBI-008, W-de.NBI-010, W-de.NBI-013, W-de.NBI-014, W-de.NBI-016, W-de.NBI-022).

> 1000's of users of our services:

Galaxy EUROPE

EGGNOG-MAPPER genome-wide functional annotation

BIIGLE

silva ngs

& many more

numbers as of July, 10th 2025

de.NBI GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

elixir GERMANY

Training and Community Building

47 Participants

CLOUD USER MEETING BIELEFELD

75 Participants

CLOUD USER MEETING ONLINE 2020

20 Participants

5th Cloud User Meeting Jülich 2022

CLOUD USER 24 MEETING

2018 — 2019 — 2020 — 2021 — 2022 — 2023 — 2024

CLOUD USER MEETING² HEIDELBERG

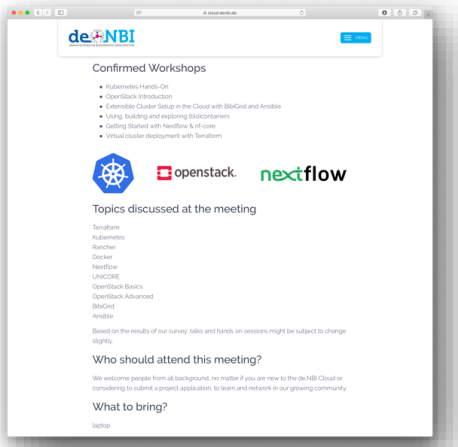
60 Participants

4th CLOUD USER MEETING ONLINE 21

67 Participants

CLOUD BI

8th de.NBI Cloud & de.KCD User Meeting 2025
30.9.-2.10.2025
Berlin



Courses:

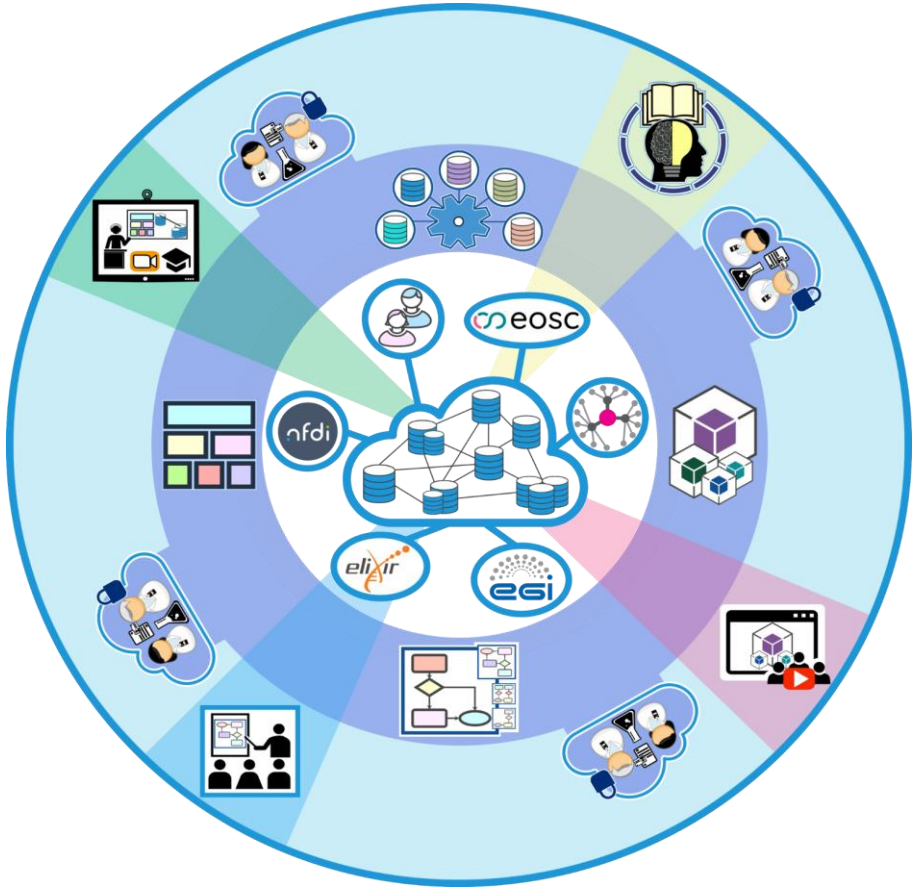
Beginners
- ➔ Introduction to VM basics: Linux, command line
- ➔ Using **SimpleVM** with Research Environments for applied bioinformatics use-cases

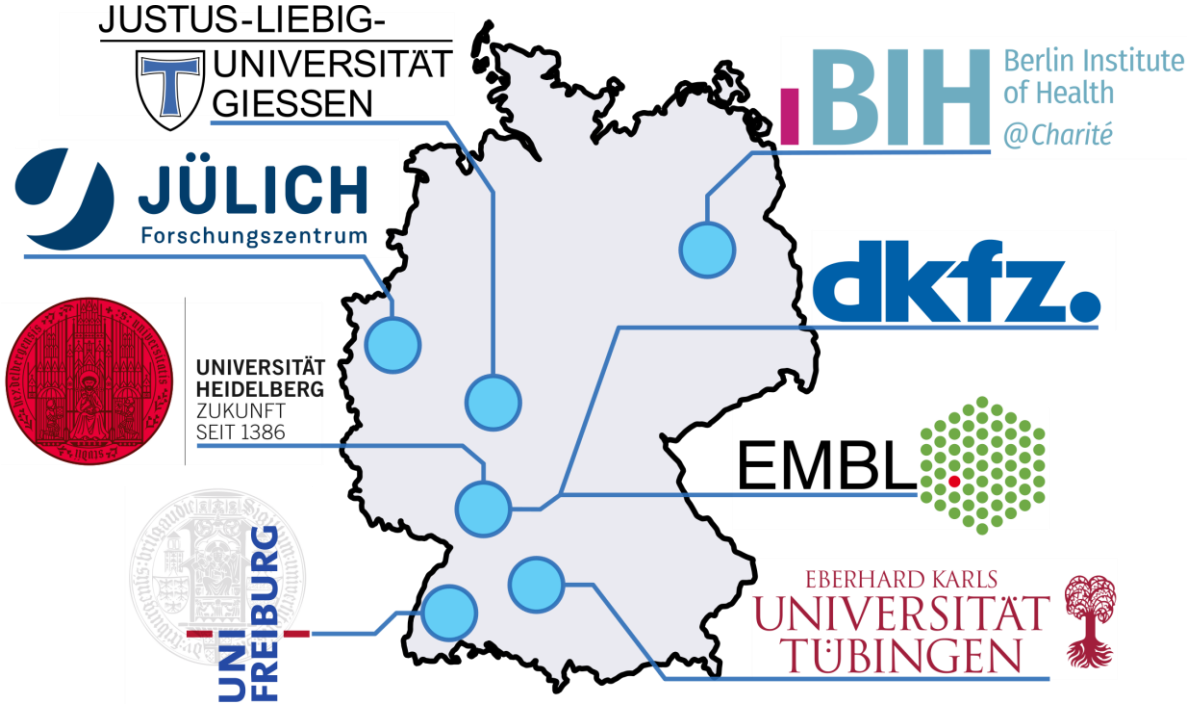
Intermediate
- ➔ Bioinformatics workflows: BioConda, BioContainers, CWL, Nextflow, Snakemake

Advanced
- ➔ IaaS / PaaS courses for: OpenStack, Kubernetes
- ➔ Infrastructure automation: Ansible, Terraform
- ➔ Grid computing in the cloud: BiBiGrid / SLURM

German Competence Center Cloud Technologies for Data Management and Processing

de.KCD





https://datenkompetenz.cloud

Collaboration with multiple RDM initiatives & NFDIs & other DCCs

Part of the Data competence centers for science, German Recovery and Resilience Plan (DARP) and BMBF's Research data action plan
Project Administration: VDI/VDE Innovation und Technik GmbH

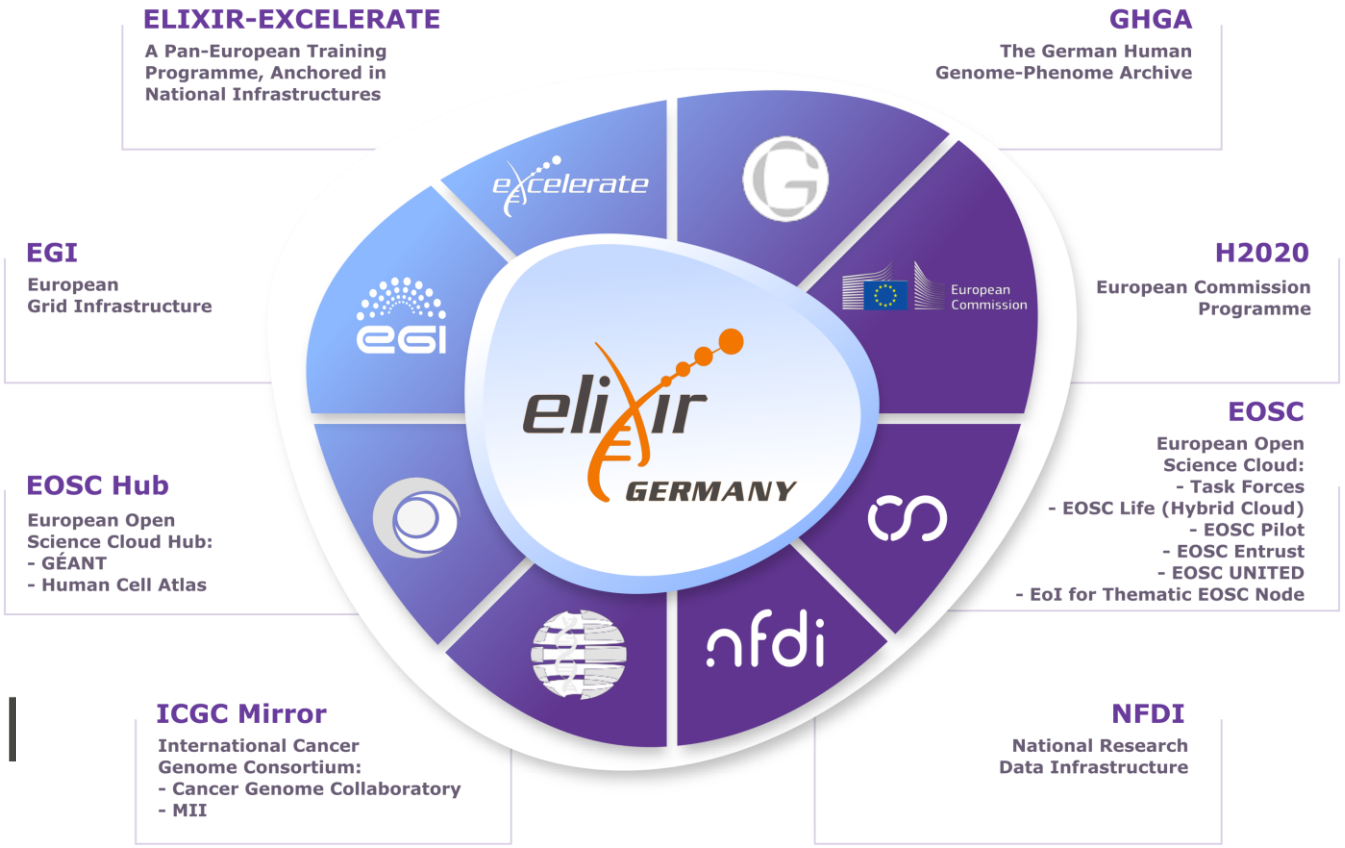
SPONSORED BY THE
Federal Ministry of Education and Research

Finanziert von der Europäischen Union
NextGenerationEU

de.NBI Cloud Collaborative Efforts and Engagement

The de.NBI Cloud engages in **national and international initiatives**, focusing on the following aspects:

- Developing, maintaining and providing a digital infrastructure that aligns with international standards
- **Technical possibilities and procedures** for pseudonymization and anonymization, as well as support for various models for potential data trusteeship solutions
- Hands-on training sessions to improve skills in data handling, cloud utilization and bioinformatics analysis



Processing of Sensitive Data @ de.NBI Cloud

LS LOGIN

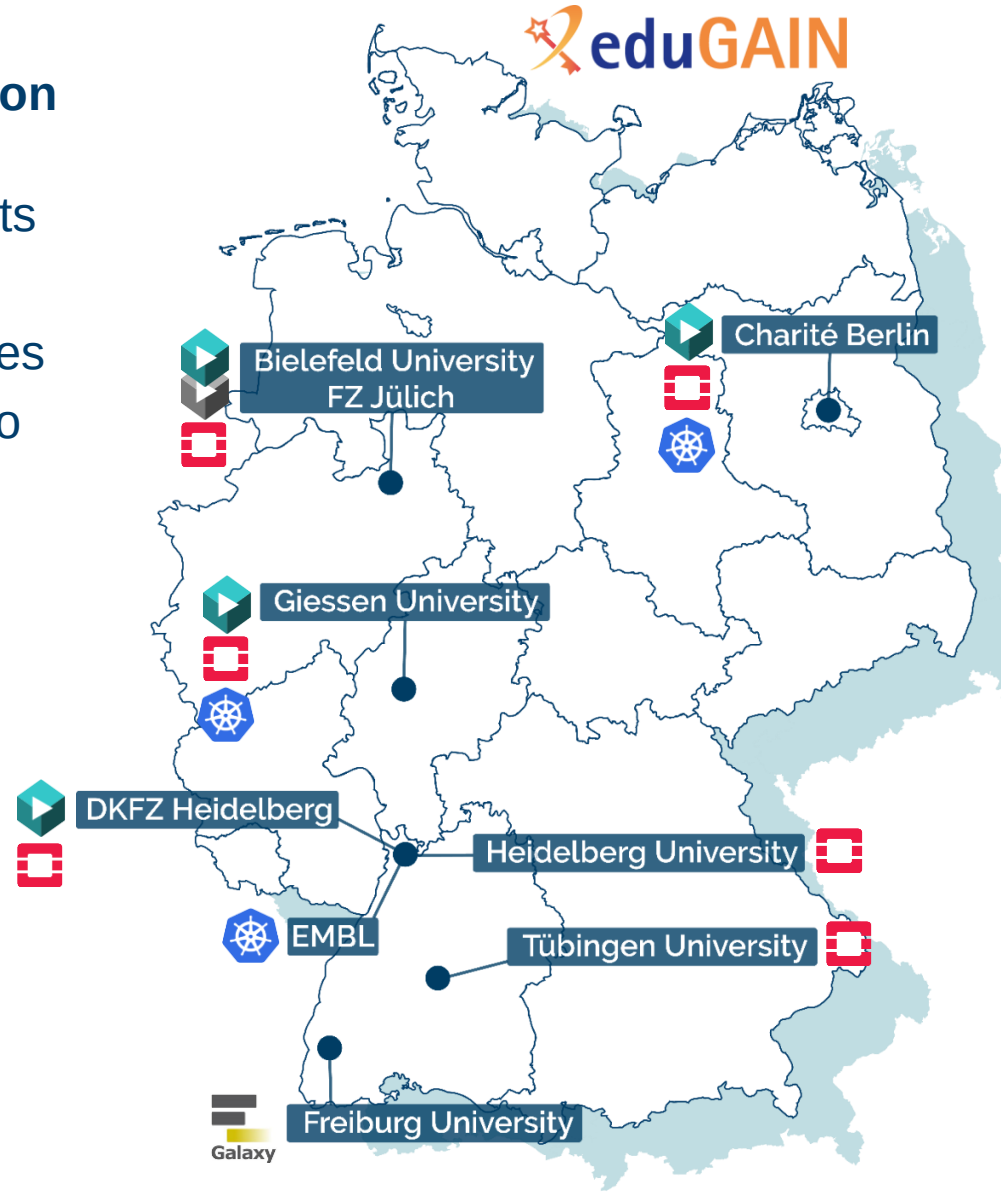
- DKFZ Heidelberg
- Charite Berlin
- Tübingen University
- FZJ / Bielefeld Univ.*

de.NBI Cloud

- Regular biosciences projects
- Projects processing
  - pseudonymized sensitive data with DP agreement
  - anonymized data

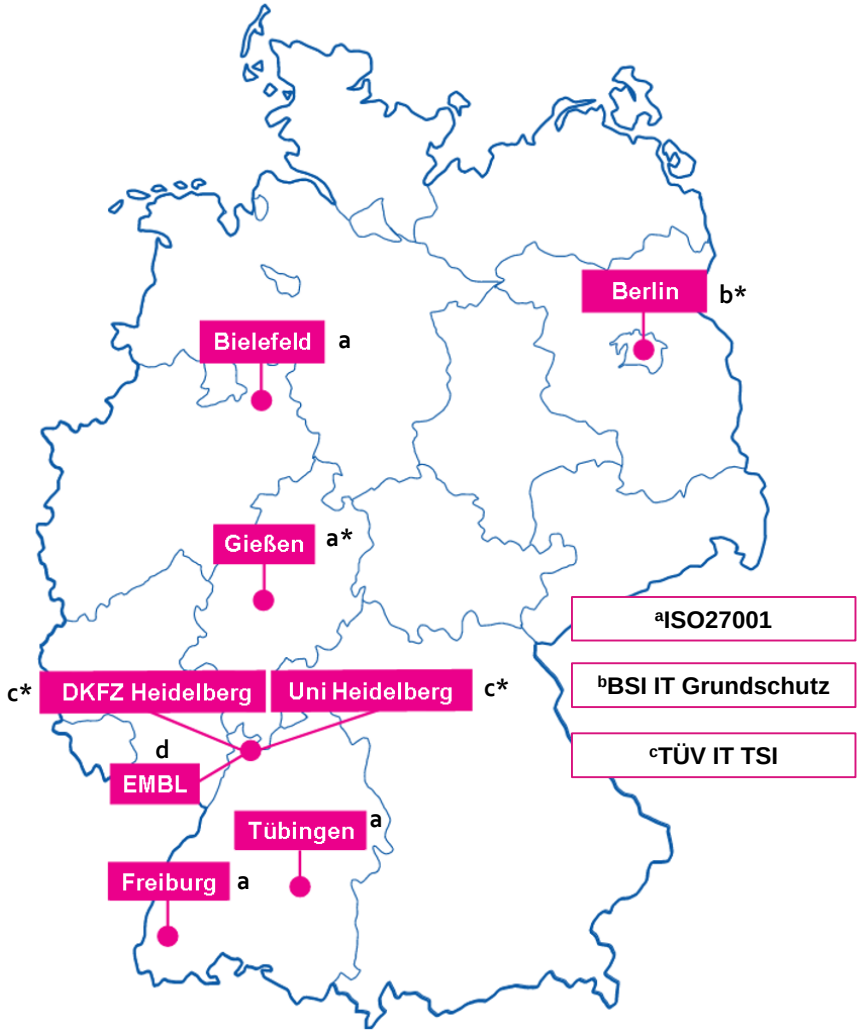
Sensitive Data Project Application Workflow

- Review of project requirements
- Assessment of data types
- Selection of suitable cloud sites
- Contact of cloud site with PI to discuss SOPs and data processing agreement
- Final allocation of project



* in development

de.NBI Cloud Federation - IT Security Certification



DIN EN ISO / IEC 27001 : 2017

Certified Information Security Management System (ISMS)

Defines technical and organisational measures and controls for risk assessment and mitigation

- de.NBI Cloud @ Bielefeld University
- de.NBI Cloud @ Freiburg University
- de.NBI Cloud @ Tübingen University
- de.NBI Cloud @ Giessen University *

BSI IT - Grundschutz

Certified ISMS and Minimum Required Implementation

Defines mandatory technical and organisational measures and controls for risk assessment and mitigation

- BIH / Charite Berlin *
- DKFZ Heidelberg (TÜV IT Trusted Site Infrastructure) *
- Heidelberg University (TÜV IT TSI) *

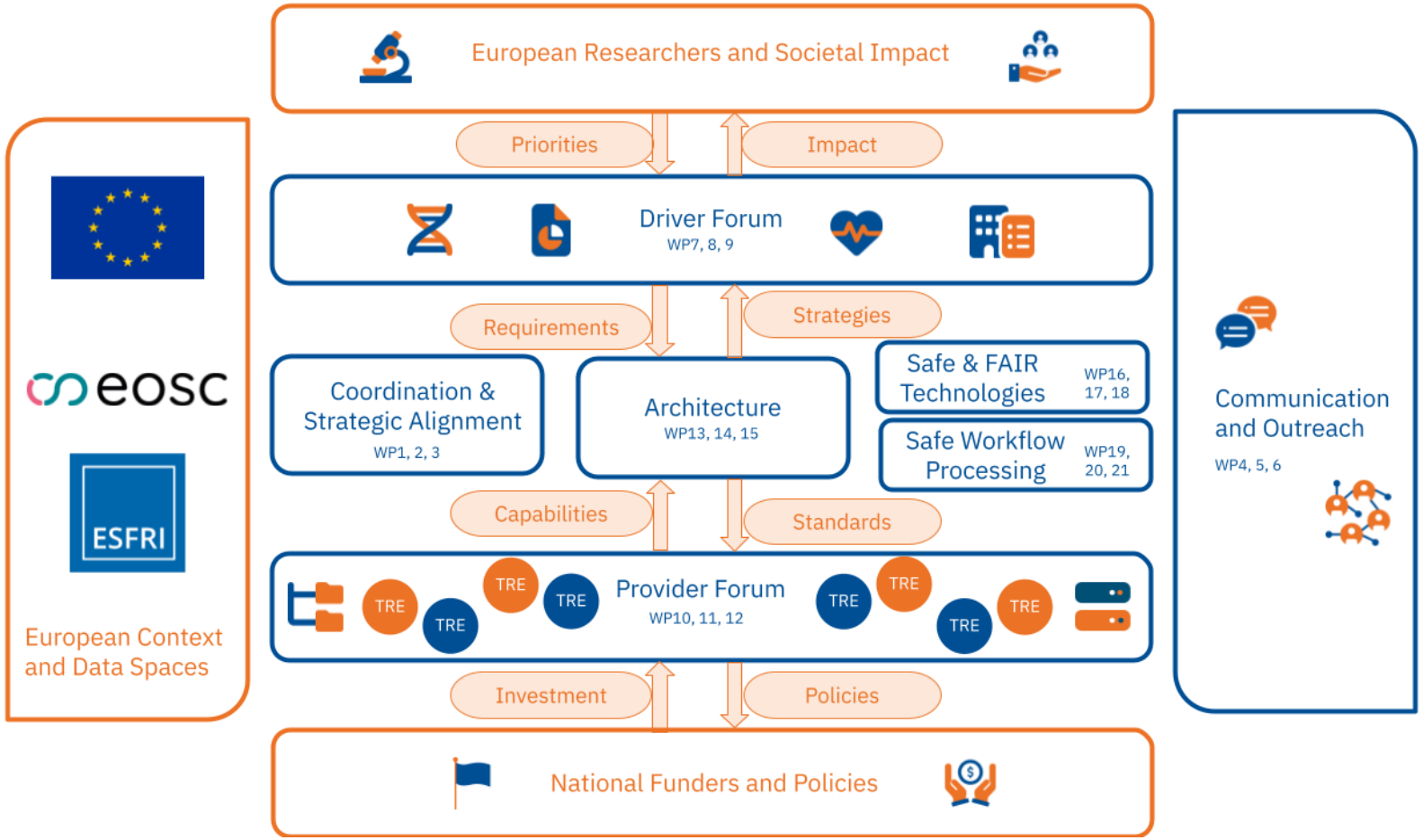
* preparation for certification ongoing

EOSC ENTRUST

2024 - 2026

EOSC | ENTRUST
European Network of Trusted Research Environments

- **Goal**:
  Create a European network of TRUSTed research environments
- Create Blueprints, Best Practice Guidelines and Software Components for sensitive data processing
- 34 European Partners
- Multiple European Infrastructures: CESSDA, ECRIN, HDR UK, ELIXIR, EUDAT
- **de.NBI Cloud / ELIXIR-DE part of TRE Provider Forum (WP 10, 11, 12)**



de.NBI GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE
elixir GERMANY

Acknowledgements

Bielefeld University
Jens Stoye
Volker Tölle
Stefan Albaum
Björn Fischer

DKFZ
Ivo Buchhalter
Martin Lang
Philip Kensche
Jordi Pujol
Bastian Beyer

Heidelberg University
Rob Russel
Nick Kepper
Max Sieckmeyer

EMBL
Peer Bork
Jan Korbel
Tobias Rausch
Alex Kelleher

Berlin Institute of Health
Roland Eils
Jürgen Eils
Martin Braun
Stefan Schneider
Valentin Schneider-Lunitz
Sven Twardziok
Harald Wagener

FZ Jülich @ Bielefeld
Peter Belmann
Christian Henke
Nils Hoffmann
Jan Krüger
Qiqi Mok
Grace Florensia
Viktor Rudko
Alexander Sczyrba
Alex Walender
David Weinholz

Gießen University
Alexander Goesmann
Burkhard Linke
Marius Dieckmann
Frank Förster
Sebasitan Beyvers

Freiburg University
Rolf Backofen
Björn Grüning
Jan Leendertse
Marc Herbstritt

Tübingen University
Oliver Kohlbacher
Jens Krüger
Johannes Werner
Mohamad Chehab
Fabian Wannenmacher
Amir Baleghi
Fabian Paz

https://cloud.denbi.de

Bundesministerium für Forschung, Technologie und Raumfahrt

JÜLICH Forschungszentrum
Shaping Change